



Universidad de Deusto  
Deustuko Unibertsitatea  
University of Deusto

## **Tercer Ciclo**

Facultad Psicología y Educación

# **Desarrollos metodológicos para la evaluación de resultados clínicos en psicología**

Estudio de la idoneidad de la propuesta de Jacobson y Truax

Tesis doctoral realizada por: Mikel Montero Matellanes

Dirigida por: Ioseba Iraurgi Castillo

El Director :

El doctorando:

Bilbao, Junio del 2015



A mi familia y a Saray.

Mis raíces, mis alas y el maná en mi senda.



## **Agradecimientos**

Aunque haya firmado la autoría de la presente tesis en solitario, sería terriblemente injusto e ingrato no reconocer las privilegiadas condiciones en las que he realizado este trabajo, debido al inestimable apoyo y a la colaboración de distintas personas e instituciones. Por ello, me gustaría mostrar mi agradecimiento a:

- La Universidad de Deusto, ya que al haber sido beneficiario de su Programa de Formación de Personal Investigador (FPI), he podido sacar un gran rendimiento del tiempo que he disfrutado de esta beca. Lo que me ha permitido realizar mi tesis doctoral en estos tres años, y ha posibilitado un desarrollo de mis competencias profesionales (tanto de investigación, como docentes), que de otra forma me hubiese sido imposible alcanzar.
- El Dr. Ioseba Iraurgi Castillo, mi director de tesis, que me descubrió su apasionante línea de investigación, y me contagió de su entusiasmo por reducir la distancia entre el ámbito de la práctica clínica y el de la investigación. Durante estos años, siempre se ha mostrado accesible y dispuesto a facilitarme el proceso de elaboración de la tesis, con sus amplios conocimientos, sabiduría, afabilidad y pragmatismo.
- Los profesionales clínicos que me han permitido analizar los resultados de sus intervenciones, tanto en el ámbito público (Osakidetza), como privado (Psiquialia y Deusto Psych). Y por supuesto, a los participantes que han cedido sus datos para configurar las muestras (clínicas o no clínicas) de los estudios, sin los cuales no existiría la presente tesis.
- Todos los compañeros del equipo de investigación “Evaluación, Clínica y Salud” de la Universidad de Deusto, así como al resto de los profesionales que desarrollan su labor en la Unidad I+D+i “Deusto Psych”. Todos ellos, han hecho de mi día a día en la Universidad, una experiencia muy enriquecedora, no solo a nivel profesional sino también humano.
- Dra. Begoña Matellanes Dr, Jose Manuel Montero y Dña. Saray Lorenzo por acometer la tediosa labor de revisión externa del presente manuscrito. Y así asegurar que el resultado final sea un documento legible y comprensible para alguien más que el propio autor.



# Índice temático

---

<b>Índice de tablas .....</b>	<b>VII</b>
<b>Índice de figuras .....</b>	<b>XI</b>
<b>Glosario de Acrónimos .....</b>	<b>XIII</b>
<b>Capítulo 1. Introducción.....</b>	<b>XIX</b>
<b>PRIMERA PARTE. REVISIÓN TEÓRICA.....</b>	<b>25</b>
<b>Capítulo 2. Consideraciones generales sobre la evaluación psicológica.....</b>	<b>27</b>
2.1. Conceptualización de la evaluación psicológica.....	27
2.1.1. Antecedentes históricos y términos relacionados .....	27
2.1.2. Paradigmas teóricos y evaluación psicológica .....	30
2.1.3. La evaluación psicológica contemporánea entendida como proceso .....	32
2.2. Instrumentos de Evaluación .....	35
2.2.1. La observación .....	36
2.2.2. El autoinforme .....	37
2.2.3. La entrevista .....	38
2.2.4. Las técnicas proyectivas .....	38
2.2.5. Las técnicas subjetivas .....	39
2.2.6. Técnicas psicométricas (Test) .....	40
2.2.7. Las técnicas objetivas .....	41
2.3. Garantía científica de las técnicas de evaluación: fiabilidad y validez .....	42
2.3.1. La fiabilidad .....	43
2.3.2. La validez .....	45
2.4. Bondad ética de la evaluación psicológica.....	50
2.4.1. Código Deontológico del psicólogo .....	50

2.4.2. Ética en la evaluación psicológica.....	52
2.4.3. Principios bioéticos del ámbito sanitario.....	54
<b>Capítulo 3. Evaluación de Resultados en salud general .....</b>	<b>55</b>
3.1. Antecedentes conceptuales y contextuales.....	55
3.1.1. Gestión de la calidad en servicios sanitarios .....	55
3.1.2. Medicina Basada en la Evidencia (MBE) .....	60
3.2. Conceptos relevantes en la evaluación de resultados en salud.....	66
3.2.1. Eficacia, efectividad y eficiencia.....	66
3.2.2. Calidad de la información: Validez, precisión, errores y sesgos .....	71
3.2.3. Apecepciones de la evaluación en entornos sanitarios.....	75
3.2.4. Principales diseños evaluativos utilizados en intervenciones en salud .....	77
3.3. Investigación de Resultados en Salud (IRS) .....	82
3.3.1. Desarrollo de la IRS y aclaraciones conceptuales. ....	82
3.3.2. Tipos de Resultados en IRS.....	85
<b>Capítulo 4. Evaluación de Resultados en Intervenciones psicológicas.....</b>	<b>89</b>
4.1. Desarrollo histórico de la evaluación de resultados en psicología.....	89
4.2. Tratamientos Psicológicos Basados en la Evidencia (TPBE). ....	94
4.2.1. Establecimiento de la eficacia en TPBE.....	96
4.2.2. Utilidad clínica de los TPBE.....	99
4.2.3. Limitaciones de la TPBE.....	102
4.2.4. TPBE en psicología de la salud. ....	106
4.3. Practica Psicológica Basada en la Evidencia (PPBE).....	109
4.3.1. Factores comunes en intervenciones psicológicas.....	110
4.3.2. Investigación en la práctica clínica.....	114
4.3.3. Medir resultados en la práctica clínica habitual .....	118
<b>Capítulo 5. Paradigmas estadísticos en la evaluación de resultados clínicos.....</b>	<b>121</b>
5.1. De lo estadísticamente significativo a lo clínicamente relevante.....	121
5.1.1. Significación estadística .....	121
5.1.2. Magnitud y relevancia de los efectos .....	127
5.2. Análisis estadísticos habituales en la evaluación del cambio terapéutico.....	135

5.2.1. Pruebas de significación .....	135
5.2.2. Índices de tamaño del efecto .....	141
5.3. Metodologías de análisis de cambio individual y de su relevancia clínica .....	147
5.3.1. Método de Índice de Cambio Fiable de Jacobson y Truax (ICF <sub>JT</sub> ) .....	148
5.3.2. Metodologías alternativas al ICF <sub>JT</sub> .....	154
<b>Capítulo 6. Problemáticas de salud implicadas en el desarrollo empírico.....</b>	<b>167</b>
6.1. Definición de Salud y modelo biopsicosocial .....	167
6.2. Ansiedad.....	169
6.2.1. Aclaraciones conceptuales sobre la ansiedad .....	170
6.2.2. Categorización y epidemiología de la ansiedad .....	173
6.2.3. Modelos explicativos de los trastornos de ansiedad .....	177
6.3. Depresión .....	180
6.3.1. Aclaraciones conceptuales sobre la depresión.....	180
6.3.2. Categorización y epidemiología de la depresión.....	183
6.3.3. Modelos explicativos de la depresión.....	189
6.4. Asma .....	192
6.4.1. Epidemiología, clínica , diagnostico y control del asma .....	193
6.4.2. Medidas de la función pulmonar y espirometría .....	198
6.4.3 Factores psicológicos en el asma.....	200
<b>SEGUNDA PARTE. DESARROLLO EMPÍRICO.....</b>	<b>205</b>
<b>Capítulo 7. Introducción a estudios empíricos .....</b>	<b>207</b>
7.1. Objetivo general .....	207
7.2. Hipótesis generales.....	207
7.3. Hipótesis específicas referidas a cada estudio .....	208
7.4. Justificación y presentación de estudios empíricos.....	209

<b>Capítulo 8. Estudio 1. Evaluación de resultados en intervención sobre trastornos de ansiedad en ámbito privado .....</b>	<b>215</b>
8.1. Objetivo.....	215
8.2. Hipótesis.....	215
8.3. Metodología .....	215
8.3.1. Participantes .....	215
8.3.2. Diseño.....	216
8.3.3. Variables dependientes.....	216
8.3.4. Variable Independiente.....	217
8.3.5. Instrumento.....	218
8.3.6. Procedimiento.....	220
8.3.7. Análisis de datos.....	220
8.4. Resultados .....	222
8.4.1. Análisis mediante contrastes de significación clásicos de promedios.....	222
8.4.2. Análisis de cambio individual mediante Índice de Cambio Fiable .....	224
8.5. Discusión y conclusiones .....	234
<b>Capítulo 9. Estudio 2. Evaluación de resultados en un Centro de Salud Mental del ámbito publico .....</b>	<b>239</b>
9.1. Objetivo.....	239
9.2. Hipótesis.....	239
9.3. Metodología .....	239
9.3.1 Participantes .....	239
9.3.2. Diseño.....	241
9.3.3. Variables dependientes.....	241
9.3.4. Variable Independiente.....	242
9.3.5. Instrumento.....	243
9.3.6. Procedimiento.....	245
9.3.7. Análisis de datos.....	246
9.4. Resultados .....	247
9.4.1. Análisis mediante contrastes de significación clásicos de promedios.....	247

9.4.2. Análisis de cambio individual mediante Índice de Cambio Fiable .....	249
9.5. Discusión y conclusiones .....	254
<b>Capítulo 10. Estudio 3. Evaluación de resultados en una intervención sobre trastornos depresivos en el ámbito público. Comparación de diferentes estrategias de análisis de cambio individual.....</b>	<b>260</b>
10.1. Objetivos .....	260
10.2. Hipótesis.....	260
10.3. Metodología .....	261
10.3.1 Participantes .....	261
10.3.2. Diseño.....	262
10.3.3. Variable dependiente .....	262
10.3.4. Variable Independiente.....	263
10.3.5. Instrumento.....	263
10.3.6. Procedimiento.....	264
10.3.7. Análisis de datos.....	266
10.4. Resultados .....	269
10.4.1. Análisis mediante contrastes de significación clásicos grupales.....	269
10.4.2 Análisis de cambio individual mediante metodología ICF <sub>IT</sub> .....	271
10.4.3. Análisis y contraste de cambio individual entre diferentes metodologías.....	274
10.5. Discusión y conclusiones .....	282
<b>Capítulo 11. Estudio 4: Uso del índice de cambio fiable en la evaluación de la efectividad de intervenciones clínicas. Aplicación en un programa formativo de adhesión al tratamiento en asma.....</b>	<b>289</b>
11.1. Objetivo.....	289
11.2. Hipótesis.....	289
11.3. Metodología .....	290
11.3.1. Procedimiento y participantes .....	290
11.3.2. Diseño.....	293
11.3.3. Análisis de datos.....	293
11.4. Resultados .....	294

11.4.1. Análisis mediante contrastes de significación clásicos de promedios.....	294
11.4.2. Análisis de cambio individual mediante Índice de Cambio Fiable .....	295
11.5. Discusión y conclusiones .....	298
<b>Capítulo 12. Discusión y conclusiones generales .....</b>	<b>303</b>
12.1. Contraste de hipótesis y discusión .....	303
12.2. Conclusiones generales .....	318
12.3. Limitaciones .....	320
12.4. Implicaciones, aplicaciones y futuras líneas de investigación .....	321
 <b>BIBLIGRAFÍA Y ANEXOS .....</b>	 <b>325</b>
 <b>Bibliografía .....</b>	 <b>327</b>
<b>Anexos .....</b>	<b>401</b>
Anexo 1. Sintaxis SPSS para el cálculos de los Índices de cambio Fiable .....	401
Anexo 2. Instrumento de Evaluación de Resultados OQ-45.....	407
Anexo 3. Tablas de Resultados de Cambio Individual .....	408

# Índice de tablas

---

## **Tabla 1**

Beneficios y Limitaciones de la MBA en la práctica clínica ..... 66

## **Tabla 2**

Caracterización Extremos del continuo pragmático-explicativo en investigación clínica ..... 69

## **Tabla 3**

Estrategias para controlar o minimizar sesgos en investigación clínica ..... 74

## **Tabla 4**

Principales concepciones de evaluación en el ámbito sanitario ..... 75

## **Tabla 5**

Criterios y niveles de evidencia en los "Tratamientos Psicológicos Basados en la evidencia" ... 97

## **Tabla 6**

Revisión de la evidencia acerca de la eficacia de intervenciones psicológicas en salud..... 108

## **Tabla 7**

Ventajas y limitaciones de los diseños de caso único en investigación en psicoterapia..... 116

## **Tabla 8**

Errores y riesgos asociados a la prueba de hipótesis de Neyman y Pearson ..... 124

## **Tabla 9**

Diferencias entre prueba de significación de Fisher y prueba de hipótesis Neyman-Pearson ... 125

## **Tabla 10**

Medios para establecer significación clínica del paciente en psicología..... 134

<b>Tabla 11</b>	
Modelos de ANOVA en investigación en psicología clínica .....	137
<b>Tabla 12</b>	
Formulación de los principales indicadores de tamaño del efecto grupal comentados .....	146
<b>Tabla 13</b>	
Formulación de principales alternativas a ICF <sub>JT</sub> para el análisis del cambio individual .....	159
<b>Tabla 14</b>	
Clasificación categorial de los trastornos de ansiedad.....	175
<b>Tabla 15</b>	
Clasificación categorial de los trastornos depresivos .....	187
<b>Tabla 16</b>	
Diagnostico diferencial del asma .....	196
<b>Tabla 17</b>	
Representación gráfica del diseño del estudio 1 .....	216
<b>Tabla 18</b>	
Datos normativos nacionales y fiabilidad de escalas del SCL-90 utilizadas en el estudio 1.....	219
<b>Tabla 19</b>	
Estadísticos descriptivos de la muestra y pruebas de normalidad del estudio 1 .....	222
<b>Tabla 20</b>	
Resultados de contraste de significación de diferencias de medias y magnitudes del efecto 1	223
<b>Tabla 21</b>	
Resultados individuales en variable Ansiedad según método ICF <sub>JT</sub> en estudio 1 .....	227

<b>Tabla 22</b>	
Resultados individuales en variable Ansiedad Fobica según método ICF <sub>JT</sub> en estudio 1 .....	228
<b>Tabla 23</b>	
Resultados individuales en variable Somatización según método ICF <sub>JT</sub> en estudio 1.....	229
<b>Tabla 24</b>	
Resultados individuales en variable Obsesión-compulsión según método ICF <sub>JT</sub> en estudio 1 .	230
<b>Tabla 25</b>	
Resultados individuales en Índice Global de Gravedad según metodología ICF <sub>JT</sub> en estudio1 .	231
<b>Tabla 26</b>	
Caracterización y comparación grupos estudio y grupo de casos abandono del estudio 2.....	240
<b>Tabla 27</b>	
Representación gráfica del estudio 2 .....	241
<b>Tabla 28</b>	
Datos normativos de OQ-45 en diferentes estudios de validación .....	244
<b>Tabla 29</b>	
Datos normativos ponderados de OQ-45 utilizados en el estudio 2 .....	245
<b>Tabla 30</b>	
Estadísticos descriptivos de la muestra y pruebas de normalidad del estudio 2 .....	248
<b>Tabla 31</b>	
Resultados de contraste de significación de diferencia y magnitudes del efecto en estudio 2 ...	249
<b>Tabla 32</b>	
Representación gráfica del diseño del estudio 3 .....	262

<b>Tabla 33</b>	
Recuentos de clasificación de cambio individual según los diferente métodos en estudio 3 .....	275
<b>Tabla 34</b>	
Porcentajes de acuerdo entre métodos de clasificación de cambio individual en estudio 3 .....	276
<b>Tabla 35</b>	
Comparación por pares entre clasificaciones de cambio en grupo clinico del estudio 3.....	278
<b>Tabla 36</b>	
Comparación por pares entre clasifiaciones de cambio en grupo no clínico del estudio 3 .....	279
<b>Tabla 37</b>	
Datos de caracterización muestral del estudio 4 .....	291

# Índice de figuras

---

## **Figura 1**

Relación de los tipos de resultados en salud según modelo IRS ..... 86

## **Figura 2**

Uso Intervalos de Confianza para valorar relevancia práctica de resultados..... 130

## **Figura 3**

Puntos de corte de significación clínica en el modelo de Jacobson y Truax ..... 150

## **Figura 4**

Resultados  $ICF_{JT}$  en variable OQ-45 Total en el estudio 2.....252

## **Figura 5**

Resultados  $ICF_{JT}$  en variable OQ-45 Síntomas en el estudio 2 .....252

## **Figura 6**

Resultados  $ICF_{JT}$  en variable OQ-45 Interpersonal en el estudio 2 .....253

## **Figura 7**

Resultados  $ICF_{JT}$  en variable OQ-45 Rol Social en el estudio 2 .....253

## **Figura 8**

ANOVA mixto con medidas repetidas en grupos del estudio 3 .....271

## **Figura 9**

Clasificación de cambio individual  $ICF_{JT}$  en puntuación BDI del estudio 3.....272

**Figura 10**

Esquema general del estudio 4 en diagrama de flujo.....292

**Figura 11.**

ANOVA medidas repetidas para las dos condiciones experimentales del estudio 4.....295

**Figura 12.**

Resultados según Índice Cambio Fiable, separados por sexo del estudio 4 .....296

## Glosario de Acrónimos

---

ACh- Acentilcolina.

AENOR- Asociación Española de Normalización y Certificación.

AERA- American Educational Research Association.

AHCPR- Agency for Health Care Policy and Research.

ANOVA- Analysis of Variance.

APA- American Psychological Association.

APA<sub>2</sub>- American Psychiatric Association.

ARIMA- Autoregressive Integrated Moving Average.

ATS- American Thoracic Society.

ATM- Agència d'Avaluació Tecnologia Mèdica.

AVAC- Año de Vida Ajustado por Calidad.

BDI- Beck's Depression Inventory.

BOE- Boletín Oficial del Estado.

BTS- British Thoracic Society.

CCS- Cambio Clínicamente Significativo.

CDP- Código Deontológico del Psicólogo.

CEBM- Centre for Evidence-Based Medicine de Oxford.

CGCOP- Consejo General de Colegios Oficiales de Psicólogos.

CIE- Clasificación Internacional de Enfermedades.

CLES- Common Language Effect Size.

COP- Colegio Oficial de Psicólogos.

CORE- Clinical Outcomes in Routine Evaluation.

CPM 50%- Criterio de Mejoría Sintomática del 50%.

CTFHE- Canadian Task Force on the periodic Health Examination.

CVRS- Calidad de Vida Relacionada con la Salud.

DE- Desviación Estándar.

DIT- Diferencia Individual Tipificada.

DOUE- Diario Oficial de la Unión Europea.

DSM- Diagnostic and Statistical Manual of Mental Disorders.

DTW- Diferencia Tipificada de Wyrwich.

ECRHS- European Community Respiratory Health Survey.

EED-Error Estándar de la Diferencia.

EEM- Error Estándar de la Media.

EEUU – Estados Unidos.

EFPA- European Federation of Psychologists' Associations.

EN- Edward-Nunnally (índice).

EPOC- Enfermedad Pulmonar Obstructiva Crónica.

ESEMeD- European Study of the Epidemiology of Mental Disorders.

ESG- Estadístico de Sensibilidad de Guyatt.

ETE- Error Típico de Estimación.

ETM- Error Típico de Medida.

FEV<sub>1</sub>-Forced Expiratory Volume first second (índice).

FVC- Forced Vital Capacity (índice).

GABA- Gamma-Aminobutírico (Ácido).

GAP- Guidelines for the Assessment Process.

GEMA- Guía Española Manejo del Asma.

GINA- Global Initiative for Asthma.

GLN- Gullisken-Lord Novick (índice).

GMA- Growth-Modeling Analysis.

GRADE- Grading Of Recommendations Assessment.

HA- Hagerman Arrindal (método).

HLM- Hierarchical Linear Modeling.

HLL- Hsu-Linn-Lord (índice).

IAPT- Improving Access to Psychological Therapies.

IC- Intervalo de Confianza.

ICD- International Classification of Diseases.

ICF<sub>HA</sub>- Índice de Cambio Fiable de Hagemany Arrindal.

ICF<sub>JT</sub>- Índice de Cambio Fiable de Jacobson y Truax.

ICS<sub>HA</sub>- Índice de Cambio Significativo de Hagerman y Arrindal.

IGG-Índice Global de Gravedad.

IPI- Intervalo para Pronóstico Individual.

IPP- Intervalo para Pronóstico Promedio.

IRS- Investigación de Resultados de Salud.

ISAAC-International Study of Asthma and Allergies Childhood.

ISRA- Inventario de Situaciones y Respuestas de Ansiedad.

ISRS- Inhibidores Selectivos de la Recaptación de la Serotonina.

M- Media.

MBE- Medicina Basada en la Evidencia.

MC<sub>e</sub>- Media Cuadrática del error.

NA- Noradrenalina.

NCME- National Council on Measurement in Education.

NHLBI-National Health, Lung and Blood Institute.

NHST- Null Hypothesis Significance Testing.

NICE- National Institute for Clinical Excellence.

NIMH- National Institute of Mental Health.

NK-Nunnally-Kotsch (índice).

OMS/WHO- Organización Mundial de la Salud/ World Health Organization.

OQ-45-Outpatient Questionnaire 45.

PCF- Punto de Corte con la Funcionalidad.

PEF- Peak Flow (índice).

PORT-Patient Outcomes Research Teams.

PPBE- Practica Psicológica Basada en la Evidencia.

PRN- Practice Research Networks.

PRO- Patient Reported Outcomes.

RCI- Reliable Change Index.

SCL-90R-Symptom Checklist 90 Revised.

SEPAR-Sociedad Española de Neumología y Cirugía Torácica.

SIGN-Scottish Intercollegiate Guidelines Network.

SPSS- Statistical Package for Social Sciences .

SS- Sum of Squares.

STAI- State Trait Anxiety Inventory.

TCT- Teoría Clásica de los Test.

TEI- Tamaño del Efecto Individual.

TFPDP -Task Force on Promotion and Dissemination of psychological Procedures

TPBE- Tratamientos Psicológicos Basados en la Evidencia.

TRI- Teoría de Respuesta al Ítem.

UK- United Kingdom.

USPSTF- United States Preventive Services Task Force.

VD- Variable Dependiente.

VI- Variable Independiente.

VPC- Verdadero Punto de Corte con funcionalidad.

5-HT- Serotonina.



# Capítulo 1

## Introducción

---

La evaluación de resultados en psicología implica valorar y constatar el cambio que ha podido acontecer en una persona o grupo, en determinadas variables, durante el curso de una intervención. Sin este análisis no podríamos saber los beneficios atribuibles a nuestra labor profesional. Por lo que se trata de un procedimiento indispensable para la determinación de la efectividad de nuestro proceder (Rivera y Bernal, 2008).

Sin embargo, la forma de determinar y evaluar los verdaderos efectos generados por las intervenciones psicológicas ha sido objeto de un intenso debate académico en las últimas décadas. A pesar de ello, en el momento actual, la necesidad de integrar el ámbito de la investigación científica con el contexto clínico de la práctica real, es señalado de forma consensuada como una prioridad para el avance de la psicología aplicada. De esta forma, se ha observado como insuficiente para validar científicamente intervenciones, realizarlo únicamente a través de los resultados obtenidos en las condiciones artificiales de entornos de control experimental (dificultad para generalizar conclusiones), resultando necesario, también, observar los efectos de las mismas en contextos de desempeño habitual. Y por otro lado, se ha advertido que la práctica clínica y su evaluación de resultados ven drásticamente reducida su efectividad y fiabilidad si no atienden a los avances y herramientas generadas desde el ámbito científico (Kadzin, 2008; Botella, 2009; Echeburúa, Salaberría, Corral y Polo-López, 2010; Estupiña, 2012; Lambert, 2013).

A pesar de la reconocida idoneidad de esta integración, es un hecho persistente la gran distancia que existe entre ambas realidades, no beneficiándose todo lo que podrían una de la otra si no fuese así. Para remediar esto, han surgido en los últimos años varias iniciativas estables de colaboración entre psicoterapeutas no pertenecientes al contexto académico e investigadores universitarios, que han tratado de alinear los objetivos de ambos ámbitos (Castoguy, 2011; Henton, 2012). Además se ha ampliado el foco de la investigación de resultados para tener en

cuenta no solo el tipo de intervención realizada sino otras condiciones de aplicación (como las características de terapeuta o el paciente). Se ha llegado incluso a constatar empíricamente que la evaluación sistemática de resultados y su comunicación, es beneficiosa para el proceso terapéutico. (Kazdin, 2011; Lambert, 2010; Wampold, 2013).

En coherencia con lo anterior, la propia consideración de los resultados terapéuticos ha ido evolucionando hacia una focalización cada vez mayor en la importancia real de los efectos conseguidos. Esta circunstancia no solo atañe a la inclusión de más variedad de indicadores de resultados, sino a la misma forma de analizarlos e interpretarlos. El objetivo de la evaluación desde esta perspectiva no se limita a constatar que ha existido algún tipo de cambio en el estado de la persona, si no que implica determinar si ese cambio es de hecho clínicamente relevante para ella. En este contexto, los medios habituales de análisis de resultados (basados en significación estadística de diferencias grupales) se muestran muy limitados en su utilidad tanto para el profesional clínico como para el investigador, al no reflejar esta realidad (Iraurgi, 2010; Gómez y Roussos, 2012).

También es relevante destacar que tanto la integración de la investigación científica con la práctica clínica, como la consideración de la relevancia clínica de los resultados, no son inquietudes particulares de la ciencia psicológica. Si no que en la actualidad lo son de todo el sector sanitario, el cual ha seguido una evolución análoga al de la psicología en su reflexión acerca de estos dos elementos. De hecho, el protagonismo contemporáneo de la perspectiva de investigación en condiciones de desempeño asistencial habitual y de los indicadores de resultados basados en medidas autoinformadas por el paciente en patologías médicas (sobre elementos como funcionalidad general y calidad de vida), responden a estas inquietudes (Badia y Lizan, 2007; Emilfork, 2009).

Pero este tipo de orientación de la evaluación de resultados no solo tiene su eco sobre el ámbito científico y clínico. En la coyuntura actual de la gestión de los recursos limitados por parte de los centros asistenciales y administraciones públicas, cobra gran protagonismo analizar y detectar los efectos reales de las intervenciones que se están realizando (y su relevancia para el cliente que representa cada paciente atendido). La importancia de la evaluación de resultados se ve reflejada en las reflexiones sobre gestión de la calidad y la optimización de la eficiencia de la asistencia (Badia y Bigorras, 2000).

Todo lo anterior, nos lleva a subrayar lo actual y pertinente que resulta en la situación contemporánea, la propuesta que realizaron Jacobson y Truax (1991) para evaluar resultados clínicos de forma individual (y que representa el objeto central de esta tesis). Este método desarrollado originalmente para el análisis del cambio acontecido en psicoterapia, establece un doble criterio para clasificar la relevancia clínica de este, en cada paciente. Para ello, combina un criterio estadístico preciso (centrado en el error estándar de medida) con una pauta de interpretación de la importancia clínica (basada en la comparación normativa del indicador utilizado). De esta forma, se genera una evaluación de resultados útil tanto para la toma de decisiones inmediata del terapeuta (interpretación del efecto conseguido en cada paciente), como para las conclusiones grupales de la efectividad de la intervención del investigador (porcentajes de clasificación de tipo de cambio). Y que dadas sus características es de fácil implementación y uso en entornos clínicos reales.

A pesar de la aceptación académica que tuvo esta propuesta, su uso y el de otros procedimientos de la misma naturaleza aún sigue siendo minoritario en la investigación de resultados terapéuticos (Hiller, W., Schindler, y Lambert, 2011). El objetivo de la presente tesis, es estudiar y reivindicar la idoneidad del uso del método en diferentes contextos terapéuticos reales, y ante distintas medidas de resultado. Este trabajo pretende responder a la demanda realizada desde esta línea de investigación de estudiar su aplicación ante diferentes condiciones, e incluso fuera del ámbito de los resultados de psicoterapia donde se desarrolló (Jacobson, Follette, y Revenstorf, 1999; Lambert y Ogles, 2009). Y además se planea contrastar su comportamiento y aportación práctica con los del procedimiento habitual de pruebas de significación estadística de diferencias promedio, y cálculo del tamaño del efecto grupal. Esta comparativa también se realizará con otras alternativas de análisis individual de la relevancia clínica de los resultados que se han planteado en los últimos años en la psicología.

El contenido del presente trabajo se estructura en dos partes. Una primera parte dedicada a la **revisión teórica** realizada para contextualizar y conceptualizar la evaluación psicológica y la evaluación de resultados terapéuticos, identificando los retos y necesidades actuales en este sentido, y describiendo las principales vías metodológicas propuestas para afrontarlos. Y una segunda parte, destinada al **desarrollo empírico** de la tesis, que incluirá cuatro estudios

independientes de evaluación de resultados de intervenciones psicológicas, donde se analizará el comportamiento del método propuesto por Jacobson y Truax.

La primera parte **de revisión teórica**, se inicia en el *capítulo 2* dedicado a exponer el área general de la evaluación psicológica por ser el marco general de la evaluación de resultados. En este apartado se presenta: su evolución histórica como disciplina y concepto, los principales tipos de herramientas evaluativos, y se termina explicitando las garantías científicas y éticas a las que se debe atender en su realización

El *capítulo 3* se centra en la cuestión de la evaluación de resultados terapéuticos en el ámbito sanitario general: se identifica los dos hitos históricos que han antecedido a su consideración actual, se revisan los conceptos básicos necesarios para comprenderla y se concluye con la perspectiva de investigación médica basada en la práctica clínica habitual y la evaluación de resultados relevantes para el paciente.

En el *capítulo 4* se presenta de forma específica las implicaciones de la evaluación de resultados en psicología: la evolución histórica de su consideración, las distintas posiciones teóricas planteadas acerca de su afrontamiento y finalmente se plantea la necesidad actual de integrar diferentes fuentes de evidencia y estudiar los efectos de las intervenciones en el contexto clínico natural.

El *capítulo 5* está dedicado a exponer los principales medios estadísticos disponibles para el estudio del cambio terapéutico, conceptualizando además la significación estadística, la magnitud del efecto y la significación clínica. Es en este apartado donde se presenta la propuesta metodológica objeto del presente trabajo, detallando su lógica y computación, así como la de otros métodos alternativos a este.

En el *capítulo 6* se describen las problemáticas de salud implicadas en los estudios de la parte empírica (ansiedad, depresión y asma), así como el modelo conceptual de salud asumido en el desarrollo de la tesis.

En cuanto a la segunda parte, dedicada al **desarrollo empírico**, esta se inicia en el *capítulo 7* donde se explicitan el objetivo e hipótesis generales para la investigación en su conjunto, así como las hipótesis específicas referidas a cada estudio. También en este apartado se realiza la presentación y justificación de cada uno de los cuatro estudios que se desarrollan en los siguientes capítulos.

En los *capítulos 8 , 9, 10 y 11* se describen los cuatro estudios realizados ajustándose al formato de informe científico (a excepción de la introducción ya realizada en el *capítulo 7*), detallando la metodología empleada, los resultados obtenidos y realizando una discusión de resultados específica para cada uno.

El *capítulo 12* concluye la parte empírica del trabajo, realizando un discusión de resultados global de los resultados de todos los estudios, para así contrastar las hipótesis generales planteadas en el *capítulo 7*. A partir de lo anterior, se concretan una serie de conclusiones generales para la tesis y se explicitan las limitaciones encontradas en su realización. Finalmente, se aborda las implicaciones y aplicaciones prácticas que se desprenden del trabajo, y se proponen posibles líneas de investigación y desarrollo.

En los *Anexos* se han adjuntado: la sintaxis SPSS desarrolladas para el cálculo de los diferentes índices de cambio individual empleados en el tercer estudio, la traducción del cuestionario OQ-45 utilizada en el segundo estudio y varias tablas complementarias de resultados individuales de los estudios 2, 3 y 4.



## **PRIMERA PARTE**

### **REVISIÓN TEÓRICA**

---



## Capítulo 2

### Consideraciones generales sobre la evaluación psicológica

---

El presente capítulo pretende establecer una serie coordinadas históricas y conceptuales de la evaluación psicológica. Se pretende establecer así un marco general para la evaluación de resultados, objeto de esta investigación. Se incide sobre la evolución del concepto de evaluación psicológica hasta su noción contemporánea. También se hace referencia a la naturaleza de las principales herramientas de las que se hace uso para realizarla. Y finalmente se explicitan las garantías científicas y éticas que se deben tener en cuenta en este proceder.

#### 2.1. Conceptualización de la evaluación psicológica

##### 2.1.1. Antecedentes históricos y términos relacionados

Según Bowman (1989) se pueden señalar antecedentes históricos de la evaluación psicológica tan lejanos que se remontan al antiguo Imperio Chino, en el cual en el año 2200 a.C. se empezaron a realizar pruebas a los funcionarios para determinar si eran aptos para cubrir el puesto que ocupaban. Sin embargo, existe consenso a la hora de señalar a los verdaderos precursores de la evaluación en la psicología contemporánea, ubicándoles a finales del siglo XIX con la emancipación de la psicología como disciplina del conocimiento. Fueron los desarrollos positivistas de Galton, Binet y Cattell, particularmente interesados en medir y cuantificar las diferencias entre los seres humanos en su dimensión psicológica (percepción, inteligencia y personalidad), los que establecieron las bases de esta área (Carpintero, 2003). De estas aportaciones resultó de especial relevancia la concepción de los “test mentales” (Cattel, 1890) como dispositivos o procedimientos que pretendían determinar la magnitud y naturaleza de las facultades mentales a través de muestras de comportamiento, campo que se desarrolló con rapidez desde la década de los años 20 y cuyo uso se generalizó a nivel internacional a partir de la Segunda Guerra Mundial (Aiken, 1996).

Al observar la evolución histórica de la disciplina, se comprueba que la mayor parte de sus desarrollos hasta mediados del siglo XX se centraron en elaborar una tecnología evaluativa dirigida al análisis y constatación de atributos psicológicos y entidades nosológicas (Barrios y Hartmann, 1986). Por ello, no fue extraño que hasta después de la Segunda Guerra Mundial, no existiese una distinción semántica entre la aplicación de pruebas y evaluación como actividad global, incluso en la jerga profesional, equiparando la labor de evaluación al posteriormente denominado “Testing” (Anastasi y Urbina, 1998). Durante la Segunda Guerra Mundial, la Office Of Strategic Services (OSS, 1948) de Estados Unidos utilizó una serie de procedimientos y técnicas de medición, entre ellas pruebas psicológicas, con el objetivo de seleccionar personal militar para puestos especializados. Los datos que se generaban de esos procedimientos eran sometidos a una integración y evaluación minuciosa por personal muy capacitado. Fue entonces la primera vez que se definió Evaluación (Assessment) como una disciplina científica en la que se integraban los tests como una parte del proceso pero no como la totalidad (Cordero, 1997). Posteriormente, Maloney y Ward (1976) subrayaron las diferencias entre prueba psicológica (test) y evaluación psicológica caracterizando esta última como un proceso de solución de problemas que se valía de los test pero que no se limitaba a ellos como única fuente de información dando peso tanto a la integración de datos como a otros métodos de observación (como la ejecutada en ambientes naturales).

El modelo médico también ha influido notablemente sobre la concepción y el desarrollo de la evaluación psicológica. De hecho el término “psicodiagnóstico”, que en el contexto europeo se utilizó durante años para denominar esta disciplina, posee un evidente origen médico (Pelechano, 1988). A pesar de su origen clínico, este término se utilizó como designación general de la materia en los planes de estudios de nuestro país durante los años ochenta (Ávila et al.,1994). Debido a la influencia de la psiquiatría y posteriormente de la neurología, se desarrollaron instrumentos para confirmar la presencia en el ámbito clínico de entidades nosológicas coherentes con los sistemas de clasificación categoriales. El término psicodiagnóstico ha perdido actualmente la vigencia que tenía en el pasado como sinónimo de evaluación. Actualmente se ha limitado su aplicabilidad a la evaluación psicológica que tiene como objeto establecer la clasificación del sujeto con fines clínicos (Fernández-Ballesteros, 2011).

Según demarca Garaigordobil (1998) a mediados de los años sesenta se consolida la propuesta conductista en el campo de la evaluación a partir de lo estipulado por Skinner (1957) con el Análisis Funcional de la Conducta. Se denominó “Evaluación Conductual” para presentar un claro contraste a la hasta entonces, tendencia predominante de centrarse en las variables endógenas del sujeto (Hersen y Bellack, 1978). Precisamente los primeros modelos de Evaluación conductual (Kanfer y Saslow, 1965) pusieron énfasis en las condiciones ambientales o de la situación. Preferían utilizar bajos niveles de inferencia. Se basa en la observación directa y utiliza pruebas psicológicas referidas a criterios (comparación del sujeto consigo mismo y objetivos preestablecidos) en vez de a normas (comparación del sujeto con grupos de referencia poblacional). Aunque su planteamiento inicial fue radicalmente situacionista (la conducta en función de los estímulos externos) ha terminado decantándose por un modelo interaccionista en el que las variables de la persona y del ambiente interaccionan recíprocamente (Haynes y Heiby, 2004)

Otro termino también relacionado históricamente con la evaluación psicológica ha sido el de “Valoración” (“Evaluation” en su acepción anglosajona original). Las reformas sociales propuestas desde los gobiernos occidentales a partir de mediados de los años 70, respondiendo al modelo del estado del bienestar, se trataron de materializar a través de programas de intervención sobre la comunidad de diferente índole cuyo objetivo era el de resolver algún problema que incumbía a los ciudadanos (Fernández-Ballesteros, 1987). Se hizo necesario valorar la efectividad y bondad de esos programas creándose así un área de evaluación de intervenciones concreta en la realidad social, fue cuando se empezó hablar de “Valoración” (Rossi y Freeman, 1989). Según el análisis de Fernández-Ballesteros (1985) esa connotación de juicio de bondad sobre el objeto del estudio es lo que diferencia “Valoración” (Evaluation) del término más global de “Evaluación” (Assessment) el cual implica una exploración pero no es necesario un juicio de idoneidad o eficacia.

De forma general se ha constatado una evolución histórica integrativa del foco de la evaluación psicológica. Partiendo de las variables internas inferidas del sujeto hasta llegar al análisis de sus contextos, y entendiéndose su marco de actuación contemporáneo a la consideración de múltiples variables en interacción, incluyendo el entorno en el que se desenvuelven (Forns, Kirchner y Torres, 1998). A pesar de que como se expondrá en el siguiente

apartado, los diferentes modelos teóricos relevantes en el área se han presentado históricamente en claro enfrentamiento entre sí (Yela, 1996), en el momento actual existe una significativa tendencia a la integración de posturas antagónicas (Hersen, 2004).

### **2.1.2. Paradigmas teóricos y evaluación psicológica**

Los distintos desarrollos en evaluación psicológica se han visto determinados desde su inicio por la evolución de los paradigmas en la psicología en general. Por ello, en cada época la evaluación ha sido subsidiariamente acompañada por unas determinadas concepciones y usos prácticos (Barrios y Hartmann, 1986). Basándose en el estudio sobre las diferentes aportaciones teóricas en psicología de Coan (1968) tanto Fernández-Ballesteros (1992) como Forns, Kirchner y Torres (1991) vertebran en cuatro ejes los principales modelos que han servido de base al desarrollo de esta disciplina: los focalizados en variables internas, los basados en el conductismo, los generados a partir de la concepción computacional de la mente y los que tratan de aunar los elementos de los distintos niveles sistémicos de análisis.

En primer lugar, los modelos centrados en la persona, que entienden la conducta en función de las variables internas del sujeto. Configurándose estos como los patrones primigenios de la psicología (Gregory, 2001):

- Centrado en el atributo: entendido este como una inferencia de la magnitud de variables intrapsíquicas diferenciales, a través de manifestaciones fenotípicas y mediante estrategias correlacionales.
- Dinámico: que interpreta todo comportamiento externo como reflejo de una estructura teórica de personalidad estable y una dinámica de fuerzas inconscientes.
- Médico: que se basa en confirmar entidades nosológicas en el sujeto bajo el supuesto de que su problemática posee un etiología lineal respecto a unos condicionantes biológicos endógenos.
- Humanista: una visión fenomenológica del sujeto que parte del supuesto de que no es posible tener acceso directo a la realidad, por lo que es necesario comprender e interpretar a cada individuo particular, y evaluar sus construcciones ideográficas.

En segundo lugar, los modelos que surgen en contraposición a los anteriores y ponen el énfasis en los determinantes externos o periféricos al sujeto o en el protagonismo de la interacción sujeto ambiente (Day, 1980):

- Conductual Radical: considera la conducta totalmente determinada por el estímulo externo por lo que no se contemplan variables internas. Exclusivamente interesado en el análisis de conducta manifiesta y en establecer relaciones funcionales.
- Mediacional: que a pesar de seguir considerando las variables ambientales determinantes para la conducta, señala que su influencia en la conducta está mediatizada por factores intermedios internos que deben ser analizados.
- Interactivo: entiende el comportamiento en base a transacciones o interacciones entre las variables internas personales y las externas situacionales, siendo necesario evaluar la interacción entre ambas facetas.

En tercer lugar, los modelos de procesamiento de la información que se focalizan en la percepción que tiene el sujeto del mundo y en la disposición de su estructura interna de este procesamiento (Feuerstein, 1979):

- Cognitivo: centrado en evaluar los procesos cognitivos en sí mismos más que sus productos tales como la solución de problemas, la percepción, las distorsiones de procesamiento o la formación de representaciones mentales.
- Neuropsicológico: interesado también en analizar los estilos de procesamiento que utilizan los sujetos pero buscando la relación entre estos y las estructuras y funciones cerebrales.

Y por último, los modelos integradores que están en coherencia con los enfoques biopsicosociales de salud. Señalan necesario evaluar todas las facetas del individuo (personal, orgánica, interpersonal, contextual) desde una posición integradora y de colaboración interdisciplinar (Engel, 1977; Blanco, 1987; Bronfenbrenner y Morris, 2006):

- Ecológico: el análisis del sujeto requiere de la evaluación de los distintos niveles del ecosistema en el que se encuentra inmerso, teniendo en cuenta las dimensiones individual, familiar y social

- Comunitario: pone el énfasis en la influencia del contexto sobre el sujeto y busca de elementos preventivos en el entorno natural desde una colaboración multiprofesional.

Para Coan (1979) la diversidad de modelos en psicología no deja de representar los diferentes posicionamientos respecto a una serie de factores bipolares que intervienen directamente en ella, y que determina la conceptualización de las disciplinas derivadas (como la evaluación psicológica). Los identificó a través de análisis factorial en su famoso estudio sobre aspectos definitorios de la teoría psicológica (Coan, 1968) y a continuación se reproducen con la traducción realizada por Tortosa y Civera (2006, p.21):

- Subjetivismo (procesos internos conscientes o inconscientes, informes introspectivos) frente a Objetivismo (conducta observable y determinismo biológico).
- Holismo (organización global del individuo, observación naturalista y determinantes sociales) frente a Elementalismo (organización molecular y mecanicista del sujeto).
- Transpersonal (determinantes inmediatos externos, control experimental, orientación nomotética) vs Personal (patrones y rasgos personales, consideración de lo individual, orientación ideográfica).
- Orientación Dinámica (influencia de la experiencia, procesos de motivación y aprendizaje) frente a Orientación Estática (estabilidad de conducta, predecir conducta a través de disposiciones).
- Cualitativo (libre observación, especulación o inferencia sobre procesos y patrones) frente a Cuantitativo (uso estadístico para descripción, formulación numérica y generalización normativa).
- Endógeno (determinantes internos biológicos o emocionales-observación naturalista) frente a Exógeno (determinantes ambientales, aprendizaje, definiciones operacionales).

### **2.1.3. La evaluación psicológica contemporánea entendida como proceso**

Desde los años ochenta existe el consenso de conceptualizar la evaluación como un proceso de diferentes momentos orientado hacia unos objetivos específicos (Fernández-Ballesteros, 1980; Blanco, 1986; O'Leary y Wilson, 1987; Marín y Buisán, 1986; Pelechano, 1988; Martorell, 1988).

Para Silva (1988) entender la evaluación psicológica como proceso nos lleva necesariamente a identificarla como una disciplina aplicada de la psicología científica cuyo objetivo esencial es proveer información en tareas de: clasificación, modificación y evaluación de cambios. Siendo esto aplicable tanto a personas como objetos o entidades siempre que esté presente la acción humana.

Pawlik (1979) subrayó que la naturaleza de los objetivos en evaluación son los que van a definir el proceso evaluativo, y estos se sitúan dentro de un continuo entre dos focos:

- Selectivo-Clasificadorio: Evaluar características que permitan predecir rendimientos futuros bajo supuestos de estabilidad. Por ejemplo en procesos de Selección de Personal o de Diagnóstico Clínico cuando el objeto de estudio sean las personas o de Asesoramiento Vocacional cuando el objeto sean las condiciones.
- Modificativos: Evaluar e identificar las variables que mantienen o generan determinados comportamientos y analizar sus variaciones con el fin de realizar un cambio en su patrón. Por ejemplo en procesos de Modificación de conducta o Facilitación del Aprendizaje cuando el objeto es la persona o los Cambios en entornos de aprendizaje o laborales cuando el objeto sean las condiciones.

Alonso (2004) añade a la tipificación de Pawlik, un tercer foco de objetivos: el valorativo, que busca recoger información para determinar si los resultados de una determinada intervención se adecuan a lo deseado y a qué es debido ese resultado (aplicable por ejemplo en procesos de mejora de programas). Sin embargo, el mismo autor también advierte que los objetivos no son el único condicionante de la forma de proceder al evaluar y que el profesional actúa desde sus conocimientos y supuestos sobre el problema y los factores relevantes, es decir, desde su modelo.

No existe un modelo claro de evaluación universalmente admitido por la psicología contemporánea (Alonso, 2004), pero sí existen propuestas sólidas que integran las conclusiones de numerosas investigaciones aportadas desde acercamientos conceptuales diferentes, como son: las Guidelines for the Assesment Process (GAP) de la European Association of Psychological Assesment. Desde su propuesta inicial se resaltan las tres principales acciones que cualquier

proceso de evaluación implica (Fernández-Ballesteros et al., 2003): una necesaria generación y contrastación de hipótesis, un proceso de solución de problemas y una toma de decisiones.

A pesar de que, desde sistemas de computación se ha señalado la posibilidad e idoneidad de estandarizar el proceso evaluativo en una secuencia ordenada y lógica de etapas (Westmeyer y Hageböck, 1992), las diferentes propuestas no alcanzan a fijar un acuerdo con respecto a esas etapas (Godoy, 1991; De Bruyn, 1992; Aderraga y Zaccagnini, 1992; García-Riaño, 1999). Como respuesta a esto, en el desarrollo de las GAP (siguiendo la tipología de Pawlik anteriormente mencionada) se han propuesto dos variantes del proceso de evaluación que se ajustan a las fases del método hipotético-deductivo. Seleccionando una u otra variante en base al tipo de demanda evaluativa (Fernández-Ballesteros, Márquez, Vizcarro y Zamarrón, 2011):

- La vía descriptivo-predictiva o correlacional: en procesos donde se clasifica-diagnostica y pronostica pero no se interviene. Con cuatro fases secuenciadas de la siguiente forma:
  - 1- Primera recogida de Información (recoger demanda, fijar objetivos y recoger condiciones históricas relevantes).
  - 2- Formulación de hipótesis (plantear hipótesis y operativización de variables para contraste con instrumentos).
  - 3- Contratación de hipótesis de evaluación (preparación ejecución de técnicas de evaluación, y análisis de resultados).
  - 4- Comunicación de resultados (integración de resultados y elaboración de informe para el demandante).
  
- La vía interventivo-valorativa o experimental: en procesos donde se pretende modificar las variables evaluadas y constatar en qué medida ha existido variación después de la intervención. Añade tres etapas e incluye las etapas del momento descriptivo-predictivo como el inicio de su proceso:
  - 5- Plan de intervención (selección de variables dependiente e independiente en coherencia con la evaluación inicial, selección de técnicas de intervención y diseño de su posterior evaluación).

- 6- Intervención (administración de la intervención y evaluación continua cuando sea necesaria).
- 7- Valoración y seguimiento (valorar efectos de intervención, comunicar resultados a demandante y planificar seguimiento futuro).

## **2.2. Instrumentos de Evaluación**

Aunque ya se ha señalado que la labor de evaluar implica un proceso más amplio que la mera aplicación de pruebas, también es necesario destacar que entre las competencias esenciales del evaluador experto debe figurar precisamente la correcta selección y aplicación de las diferentes técnicas de recogida de información (Matarazzo, 1990). No es extraño por tanto que Cohen y Swerdlik (2000, p.7) centren su definición de evaluación psicológica en la: “recopilación e integración de datos lograda mediante el uso de herramientas como pruebas, entrevistas, observación conductual, aparatos y procedimientos de medición diseñados de forma especial”.

Según García-Montalvo (1997) un instrumento de evaluación psicológica es todo aquello que un evaluador puede utilizar como legítima fuente de datos acerca de una persona. A pesar de que a menudo se han usado como sinónimos los términos “Test” y “Técnica de Evaluación” no significan lo mismo aún refiriéndose ambos a instrumentos de evaluación (Pelechano, 1976).

El “Test” es un dispositivo o procedimiento de medición tipificado, mediante el cual se obtiene una muestra de conducta del examinado en un determinado dominio que posteriormente se puntúa y se valora de modo normalizado (American Psychological Association APA, 1999). Deben cumplir siempre tres características: presenta escalas numéricas, a todos los evaluados se les pregunta u observa de la misma manera (estandarización) y permite comparaciones interindividuales (Cronbach, 1990). Mientras que la “Técnica de evaluación” supone también un dispositivo de recogida de información pero sin que necesariamente requiera tipificación de su material o que con ella se permita comparaciones intrasujetos (Fernández-Ballesteros, 1985).

Una de las primeras clasificaciones sobre técnicas de evaluación la establece Campbell (1957) mediante tres factores bipolares:

- Estructurado-No estructurado: existen límites al número y extensión de las respuestas que se pueden recoger o no existen esos límites.
- Objetivo-Voluntario: solo se pueden dar respuestas correctas o incorrectas, o se pueden dar respuestas voluntarias a los reactivos de la técnica.
- Directo-Indirecto: el sujeto evaluado conoce el propósito y sentido de la prueba o técnica, o por el contrario lo desconoce.

Recogiendo lo establecido por Campbell, Pervin (1980) divide las técnicas de evaluación en cuatro categorías (las denomina test al no considerar la matización contemporánea que hemos señalado sobre las técnicas de evaluación): Proyectivas (no estructuradas, indirectas y voluntarias), Psicométricas (estructuradas y directas), Subjetivas (no estructuradas, directas y voluntarias) y Objetivas (estructuradas y objetivas).

La clasificación de Pervin ha sido la más comúnmente aceptada en los manuales de evaluación psicológica en las últimas décadas, sin embargo, la evolución de la disciplina ha dado lugar a un importante desarrollo de técnicas específicas, incluidas hasta ahora en categorías más amplias y de las cuales es necesaria la consideración de sus características particulares (Buela-Casal y Sierra, 1997). Por ello, se reproduce a continuación una taxonomía ampliada con las mencionadas técnicas (Forns, Abad, Amador, Kirchner y Roig, 2002): técnicas de observación, técnicas de autoinforme, técnicas de entrevista, técnicas subjetivas y técnicas proyectivas, técnicas psicométricas y técnicas objetivas.

### **2.2.1. La observación**

La observación consiste en un procedimiento que pone de manifiesto la ocurrencia de conductas perceptibles, para proceder a su registro organizado y su análisis (tanto cualitativo como cuantitativo) mediante un instrumento adecuado (registro, codificación, muestreo focal...) y parámetros convenientes, posibilitando la detección de las relaciones de diverso orden existentes entre ellas y evaluándolas en un contexto natural o equivalente (Anguera, 2003).

En numerosas situaciones la metodología observacional es la mejor estrategia, o incluso la única posible como por ejemplo en: la evaluación de programas de baja intervención, el estudio de la interacción social, las discusiones en una pareja, el repertorio conductual del bebé, la

comunicación kinésica no verbal análisis del movimiento en múltiples actividades, ocupación de espacios, o análisis de pautas de socialización y desocialización (Angera, 2010).

Es necesaria una cuidadosa planificación para evitar los diversos sesgos y dificultades de la técnica, que suelen provenir principalmente de los siguientes grupos: la reactancia de la persona observada al saberse observada, las expectativas del observador que pueden sesgar su registro y las propias dificultades técnicas posibles al registrar la observación (Anguera, Blanco, Losada y Sánchez-Algarra, 1999). En las últimas décadas, el gran avance tecnológico que se ha producido ha facilitado en gran medida el rigor y la precisión del registro como se constata a través la extendida práctica de la grabación mecánica de las sesiones de observación (Losada, 1999; Felce y Symons, 2000).

### **2.2.2. El autoinforme**

Un autoinforme supone un mensaje verbal que una persona emite sobre cualquier tipo de manifestación propia (conductas, pensamientos, sentimientos, percepción de una situación, experiencias...). Sin embargo, no solo se puede dar cuenta de la experiencia subjetiva sino de un gran abanico de eventos tanto internos como externos informados por el sujeto. Lo que hace de esta categoría la más diversa, compleja y difícil de valorar desde el punto de vista científico (Fernández-Ballesteros, 1991). Los autoinformes implican el método de recogida más utilizado en evaluación psicológica, siendo la base de la mayor parte de los cuestionarios, inventarios o listados de adjetivos (Haynes y Heiby, 2004).

Debido a su naturaleza en los autoinformes es necesario controlar una serie de fuentes de error desde lo evaluado, habituales en esta técnica (Hernández, Santacreu y Rubio, 1999):

- Simulación: falseamiento conscientemente de las respuestas por parte de la persona para obtener un fin concreto.
- Deseabilidad social: tratar de dar una imagen congruente con lo que se espera de él y se considera adecuado socialmente, no tiene por qué ser voluntario.
- Tendencias de respuesta: la posible influencia distorsionante del tipo de alternativa de respuesta que ofrece la prueba sobre la respuesta del individuo).

- Otras condiciones particulares de la persona como la falta de motivación para la complementación de la prueba y/o dificultades para la introspección.

Desde la psicometría se han ofrecido herramientas para controlar en parte algunas de estas fuentes de error como es el caso de las escalas de control (escalas de sinceridad y escalas de deseabilidad social) para detectar incongruencias o anomalías en las respuestas del sujeto (Echeburúa, Amor y De Corral, 2003).

### **2.2.3. La entrevista**

La entrevista es la técnica más utilizada dentro del ámbito aplicado y en él se han ido incrementando gradualmente el uso de entrevistas estructuradas (Perpiña, 2012). Pope (1979, p. 3) define entrevista como: “aquel encuentro basado en una diferencia de roles, en el que se lleva a cabo una interacción verbal y no verbal, y donde la conducción de la misma y sus objetivos son responsabilidad del entrevistador”. Aunque se ha llegado a considerar una técnica subjetiva según la clasificación de Pervin (1980), los desarrollos de estructuración de la entrevista problematizan esta consideración (Hersen, 2004).

Para Silva (1986) tiene las ventajas de ser una herramienta flexible, que permite el registro de información variada y que más allá de la verbal, evaluando conductas o acontecimientos que no se podrían registrar de otra forma, en base a una interacción interpersonal con potencial de recibimiento empático y comprensivo para el evaluado. Aunque también señala dos serias limitaciones: el coste relativamente elevado en esfuerzo y tiempo para el evaluador, y las muchas fuentes potenciales de sesgo de la información recogida.

### **2.2.4. Las técnicas proyectivas**

Las técnicas proyectivas son planteadas como un método de estudio de la personalidad, que pone al sujeto ante una situación en la cual contestará libremente según el sentido que para él tenga esa situación, siendo el carácter esencial de esta técnica la evocación del mundo interno del sujeto y sus procesos de personalidad (Pichot, 1979) .

Estas pruebas se apoyan en la denominada “Hipótesis Proyectiva Básica” (Abt y Bellak, 1967) que postula que ante una prueba de respuesta libre, el sujeto no solo proporcionará la

información básica para satisfacer los requerimientos de la tarea, sino que también al hacerlo ofrecerá información a partir de la cual se pueden hacer deducciones relativas a su organización única de personalidad.

Aunque existente múltiples acercamientos teóricos para la construcción de este tipo de técnicas e incluso casos en los cuales una misma técnica es interpretada de diferente manera según diversos enfoques (Peña, Cañoto y Santalla de Banderali, 2006) comparten ciertos supuestos básicos heredados del modelo psicodinámico (Álvarez, Arce y De Diuk, 1994): carácter estructural de la personalidad, rasgos organizados en estratos de diferente profundidad, estabilidad de esos rasgos y la relación dinámica entre estratos y rasgos.

Las técnicas proyectivas tienen la ventaja de que son muy difíciles de falsear debido a su naturaleza enmascarada pero tienen la desventaja de ser instrumentos muy dispersos, heterogéneos y que normalmente no cuentan con controles rigurosos de fiabilidad y de validez (Clemente, 1998). Propuestas como la estandarización de la prueba Rorschach propuesta por Exner (2001) o la búsqueda del rigor metodológico con el uso de instrumentos estadísticos de Meyer y Archer (2001) tienen como objetivo superar estas limitaciones y dotar de fiabilidad y validez a este tipo de técnicas aunque como estos mismos autores destacan, aún es necesario mucho trabajo en esta área.

### **2.2.5. Las técnicas subjetivas**

Las técnicas subjetivas se constituyen en procedimientos de evaluación escasamente estandarizados y no enmascarados, donde los datos son obtenidos de lo que el sujeto describa de sí mismo (o de la información aportada por personas allegadas). Al no existir apenas estandarización, el evaluador puede diseñar el instrumento de forma muy flexible aunque requiera de un conocimiento profundo del individuo evaluado para realizar una correcta interpretación (Cano, Rodríguez, Martínez y Antuña, 2005).

Desde la concepción fenomenológica de la persona, se ha venido proponiendo este modo de análisis diferenciado de los modelos psicométricos tradicionales sin planteamientos nomotéticos, ni esquemas previos donde lo importante sea el individuo y no a la comparación entre sujetos (Pervin, 1988). Para Peñate y Matud (1996) los diferentes desarrollos de la tecnología evaluativa

han problematizado la entidad de esta categoría, pudiendo integrarse sus características constituyentes dentro de las técnicas de auto informe o de observación. Sin embargo, se definen como tales una serie de técnicas basadas en la elección y clasificación de elementos desarrolladas desde de la tradición fenomenológica (Casado, 2006) como son las siguientes: Listado de Adjetivos (Lubin y Zuckerman,1999), Diferencial Semántico (Osgood, Suci y Tannenbaum, 1957) y la Técnica de Rejilla de la Teoría de Constructos Personales (Kelly, 1955).

#### **2.2.6. Técnicas psicométricas (Test)**

Una técnica psicométrica es la medida de una muestra, objetiva y estandarizada, del comportamiento de una persona y además es sinónimo de test psicológico (Anastasi y Urbina, 1998) con las connotaciones explicitadas al principio del apartado. Los test psicológicos se desarrollan bajo un modelo teórico que entiende la conducta determinada por atributos intrapsíquicos estables (rasgos), que no son evaluables directamente si es a través de sus manifestaciones externas (las respuestas en los tests) y se basan en el principio de medir la ejecución de las personas en la propia prueba y comparar los resultados con los de otros sujetos pertenecientes a un grupo de referencia (Gonzalez, 2007).

Para Cronbach (1985) existen dos tipos de test: los de “Rendimiento máximo” como los de inteligencia o de aptitudes (donde se busca recoger los niveles más altos de rendimiento de una persona, con ítems de una única respuesta correcta siendo relevante la dificultad y la rapidez de ejecución que exige) y los de “Rendimiento típico” como los de personalidad o de actitudes (que pretenden recoger información del modo habitual de comportarse de un individuo y existen diferentes opciones de respuesta en cada ítem que implican diferente significado).

Dentro de la psicometría se identifica dos grandes modelos teóricos de construcción y análisis de test (Barbero, 1999a):

- Teoría Clásica de los Test (TCT): donde son fundamentales la aleatorización y la estandarización, y los parámetros de los ítems depende de la muestra utilizada en su construcción.
- Teoría de Respuesta al Item (TRI): que a través de sus modelos probabilísticos permiten establecer la relación funcional entre cada ítem con cada sujeto.

Gulliksen (1950) sintetizó la óptica clásica (TCT): subrayando que su fundamento teórico se basa en el Modelo Lineal Clásico (Spearman, 1913) que asume que la puntuación obtenida en un test tiene dos componentes la verdadera respuesta del sujeto (la media de la puntuación que una persona obtendría si se pasara infinitas veces) y un error que hace variar la puntuación de una medida a otra por diversos factores circunstanciales. Muñiz (2010) señala las dos importantes limitaciones que históricamente se le han achacado a la TCT: las medidas no resultan invariantes respecto al instrumento utilizado (la medición de una misma variable varía dependiendo que instrumento se use) y la ausencia de la invariancia de las propiedades de los test respecto a las persona que utilicemos para calcularlas (las propiedades de los instrumentos de medida depende del tipo de sujetos que se hayan utilizado).

Tratando de superar las limitaciones del enfoque clásico surge la perspectiva TRI que utiliza modelos matemáticos logísticos para establecer una relación funcional entre los valores de la variable que miden los ítems y la probabilidad de acertar en estos, generando una función denominada “Curva Característica del Item”; que hace posible separar lo que se mide de quien se mide (Lord, 1980).

### **2.2.7. Las técnicas objetivas**

Las técnicas objetivas son aquellos procedimientos de recogida de información de eventos psicológicos observables o amplificables; que en gran parte de los casos no son controlables por el sujeto, y que utilizan sofisticados aparatos que permiten una administración, registro, puntuación y análisis objetivo sin necesitar de la intervención directa del evaluador (Fernández-Ballesteros y Caldero, 1992).

Dentro de las técnicas objetivas se pueden distinguir tres grupos de pruebas que se diferencian por la naturaleza de las variables que pretenden registrar (Moreno, 2003):

- Para variables cognitivas: que se centran en evaluar los procesos de percepción, memoria, y atención con instrumentos como el Estesiómetro o el Taquistoscopio.
- Para variables motoras: que comprenden toda aquella instrumentación utilizada para medir la coordinación motora con instrumentos como el Estabilímetro o el Rotor de Persecución.

- Para variables psicofisiológicas que se focalizan en evaluar los correlatos fisiológicos del comportamiento a través de la evaluación de factores fisiológicos y su comprensión en su contexto con instrumentos como el Polígrafo o el Electromiograma.

Cleary (1982) señala las ventajas del uso de este tipo de instrumentación: el control preciso de estímulos, la detección y medición de respuestas y el almacenamiento de datos se pueden llevar a cabo por la instrumentación de forma automática y mecánica, aumentando considerablemente la objetividad del procedimiento, al reducir el contacto directo entre el observador y lo observado (y sus correspondientes sesgos).

Sin embargo, este tipo de medidas plantea una serie de problemas y limitaciones que es necesario tener en cuenta y controlar (Carrobbles, 1981): el aparataje y la situación de exámen generan una reactividad fisiológica en el sujeto que pueden ofrecer indicadores erróneos (son necesarias sesiones de habituación); a la hora de sacar conclusiones de la medida hay que tener en cuenta la Ley de Valores Iniciales (Wilder, 1957) que demarca que cuanto más bajo sea el nivel basal de la respuesta psicofisiológica mayor incremento potencial de la misma tendrá ante el estímulo. Además de lo anterior cada individuo puede tener un patrón específico de respuesta estable en el tiempo que complica la comparación y la generalización de las respuestas; y por último las condiciones tan artificiales de evaluación dificultan enormemente la generalización de respuestas a la vida real (problematiza la validez ecológica).

### **2.3. Garantía científica de las técnicas de evaluación: fiabilidad y validez**

Aunque en el apartado anterior hemos expuesto las particularidades de cada tipo de instrumento, hay dos propiedades que en todo instrumento de medida se debe tener en cuenta para garantizar su calidad científica, y estas son: fiabilidad y la validez (Martínez, 1996). Es necesario matizar que en su consideración actual no son directamente características de los instrumentos sino propiedades de los usos específicos de las medidas que esos instrumentos proporcionan (American Psychological Association, American Educational Research Association y National Council on Measurement in Education, 1999). De forma genérica se puede tener presente que la fiabilidad es una cuestión relativa a la calidad de los datos, mientras que la validez se refiere a la calidad de la inferencia (Zumbo, 2007).

### 2.3.1. La fiabilidad

El concepto de fiabilidad o confiabilidad hace referencia a la precisión de la medida, observando la consistencia de diferentes mediciones sobre un mismo sujeto con el mismo instrumento; si existe una variabilidad grande, se considerará que los valores son imprecisos y en consecuencia, poco fiables (Martínez-Arias, Hernández y Hernández-Lloreda, 2006).

En psicología, debido a la naturaleza del objeto epistémico y de los instrumentos utilizados, la variabilidad entre medidas es mucho mayor que en otras disciplinas científicas, y al no ser perfectamente consistentes se supone que las medidas contienen cierta cantidad de error (Cortada de Kohan, 1999). Se trata del, ya mencionado anteriormente, “Error de Medida” enunciado en el “Modelo Lineal General” (Spearman, 1913). La cantidad de error en cada caso sería la diferencia entre la puntuación observada directamente (que incluye cierto margen de error aleatorio no controlable) y la puntuación verdadera (el hipotético valor verdadero estable en el tiempo y no observable directamente). La variabilidad de una persona en diversas ejecuciones alrededor de su puntuación verdadera, se denomina Error Típico de Medida (ETM), siendo este la desviación típica de los Errores de Medida aleatorios (Brennan, 2006). Sin embargo, la magnitud del ETM se genera en las mismas unidades de medida que el instrumento utilizado, limitando su interpretabilidad como índice general de consistencia. Por ello, se ha propuesto un índice estandarizado de precisión denominado coeficiente de fiabilidad, que puede oscilar entre 1 y 0 (siendo 1 la ausencia de error y el 0 la no existencia de consistencia alguna), y que se obtiene mediante un coeficiente de correlación, entre las puntuaciones obtenidas en las diferentes estrategias de replicación (Prieto y Delgado, 2010).

En relación a los diferentes diseños de replicación, para el cálculo de la consistencia, Anastasi y Urbina (1998) realizan una taxonomía de los tipos de fiabilidad según el número mínimo de momentos de medida que requiera: dos sesiones o una única sesión.

Sobre el cálculo de la consistencia a partir de dos sesiones, el método propuesto más directo ha sido el denominado “Test-Retest”, que consiste en repetir la aplicación con los mismos sujetos (separado por un periodo significativo de tiempo) y obtener el índice de correlación entre las dos puntuaciones de forma que se obtiene un índice de estabilidad temporal (Nunnally y Bernstein, 1994). Sin embargo, esta técnica puede presentar problemas en muchos tipos de

pruebas donde el efecto de la práctica puede sesgar los resultados de la segunda aplicación (como por ejemplo en los tests de rendimiento), para estos casos se plantea una forma alternativa que reduce este efecto denominada de “Forma Equivalente” o de “Test Paralelos” donde al mismo grupo de sujetos se les aplica una prueba y posteriormente otra prueba totalmente equivalente en contenido, longitud y dificultad pero en cuya presentación los elementos sean diferentes (Traub, 1994)

Por otro lado, las vías de cálculo de una sola sesión no se focalizan en la consistencia temporal, sino en determinar la homogeneidad de la prueba. Esto es conocido como “Consistencia Interna”, y sirve para comprobar si las respuestas a los elementos integrantes en una sola aplicación son consistentes entre sí (Mehrens y Lehmann, 1984). Uno de los medios para lograr esto es un método denominado “División de las Dos Mitades” que propone que al analizar los resultados de la aplicación se dividan en dos grupos comparables los elementos de la prueba y se realice un coeficiente de correlación entre ellos (Feldt, 1975). La principal dificultad de este sistema consiste precisamente en conseguir dividir los elementos en dos mitades paralelas comparables (Streiner, 2003). Bajo este mismo diseño se proponen diferentes fórmulas para calcular el índice de consistencia interna, la más habitual es la fórmula Spearman-Brown (Spearman, 1927) que trata de predecir el aumento de la fiabilidad si se aumenta la longitud de la prueba (Cureton, 1965). Kuder y Richarsdson (1937) desarrollaron una técnica alternativa a la división de las dos mitades para el cálculo de la consistencia interna, que se basa en el análisis de la ejecución en cada elemento de la prueba. Posteriormente, de esta propuesta se derivaron dos fórmulas: una para elementos dicotómicos “Kuder-Richarsdson-20” y otra para ítems continuos “Alfa de Cronbach” (Cronbach, 1951). Estos dos últimos índices son ampliamente utilizados porque equivalen a la fiabilidad media que obtendríamos dividiendo un test en todas sus posibles dos mitades (Schmitt, 1996). Finalmente, el otro índice de consistencia para el cual no es necesaria más de una aplicación es el conocido como fiabilidad “Interjueces o del Puntador”. Este tipo de fiabilidad se calcula en las pruebas donde el evaluador asigna las puntuaciones de forma subjetiva (como es el caso de numerosos instrumentos clínicos), y consiste en correlacionar las puntuaciones de dos examinadores independientes sobre la misma aplicación para controlar la varianza de la medida del examinador (Stemler, 2004).

Aunque las estrategias planteadas (basadas en el paradigma de la TCT) son el afrontamiento clásico y más frecuente del problema de la fiabilidad de la medida (Martínez-Arias, Hernández-Lloreda y Hernández-Lloreda, 2006), no se pueden desdeñar desarrollos posteriores que intentan superar las limitaciones de estas. La “Teoría de la Generalización” (Brennan, 2001) intenta llevar más allá el análisis de la varianza de la medida observada (no limitándose solo una parte de varianza verdadera y otra de error como se ha expuesto antes) incluyendo un mayor número de facetas y estimándolas mediante análisis de la varianza para poder realizar un control mucho más preciso de las fuentes de error. “La Teoría de Respuesta al Ítem” sustituye el error típico de medida de la TCT (como ya se ha explicado, un único valor constante que representa la variabilidad en la estimación del parámetro verdadero) por el Error Típico de Estimación (ETE) de la persona o del ítem (una medida individualizada de la precisión), generando una función que permite diferenciar entre las personas en los distintos niveles del constructo medido (De Ayala, 2009).

### **2.3.2. La validez**

La validez hace referencia a que lo medido por la prueba es de hecho lo que se pretende medir o detectar (en el diseño del instrumento) y no a otra cosa (Silva, 1989). El concepto de validez ha ido evolucionando a la vez que lo hacían los usos de las pruebas psicológicas (Kane, 2006). Esta evolución ha pasado por diferentes etapas a través de las cuales, de forma no excluyente, se han ido añadiendo diferentes fuentes de evidencia acerca de la validez, de manera que se pueden identificar tres etapas que coinciden con las diferentes revisiones de los Standards for educational and psychological Testing (Elosua, 2003): Operacional (centrada en la validez externa), Teórica (que pone énfasis en el análisis de la validez de constructo como concepto integrador) y Contextual (que añade a lo anterior el matiz del uso propuesto para la prueba).

Los primeros desarrollos acerca de la validez (etapa Operacional) tienen un cariz pragmático que dan protagonismo a las fuentes externas de validación; donde la validez de una prueba viene dada por aquello con lo que correlaciona (Guilford, 1946). Se diferencia en ese momento dos tipos de validez: de contenido y referida a criterio (American Psychological Association APA, American Educational Research Association AERA y National Council on Measurement in Education NCME, 1974).

La validez de contenido se refiere al grado en que los diferentes elementos de la prueba muestrean el constructo que se está evaluando, lo que supone un examen sistemático del contenido de la prueba para determinar si es una muestra representativa de lo que se pretende medir (Buela-Casal, Caballo y Sierra, 1996). Lo habitual es recurrir a jueces externos, expertos en el área que se está evaluando, para que valoren los elementos de la prueba. Esto implica analizar sistemáticamente el área del contenido para tener seguridad de que todos los aspectos principales están adecuadamente representados por los elementos de la prueba y en proporciones correctas (Messick, 1995). Para instrumentos que traten de medir rendimiento en capacidades concretas este tipo de validación es muy relevante, pero en otras áreas, como pueden ser las pruebas de personalidad, no suele ser suficiente y puede conducir a error si no se consideran otros indicadores de validez (Sireci, 1998).

La validez de criterio o empírica indica la eficacia del instrumento en la predicción de la conducta del sujeto en situaciones específicas. Para ello se compara el resultado del instrumento con lo que se conoce como criterio: una medida independiente y directa de la variable que nos interesa evaluar (Martínez-Arias, 2007). La relación temporal entre la medida del instrumento y el criterio genera dos tipos de validación empírica: predicativa y concurrente (Adkins, 1994).

Si el criterio es recogido con posteridad a la medida, se tratará de una validación predictiva que busca anticipar rendimientos o comportamientos futuros a largo plazo, algo relevante en entornos empresariales o forenses donde esta predicción es fundamental para los usos de la evaluación. Si el criterio se recoge antes o al mismo tiempo que al realizar la medida con el instrumento, se estará realizando una validación concurrente, ya que se busca confirmar un estado actual más que anticipar uno futuro, algo adecuado para usos donde lo que se busque es diagnosticar sobre el estado presente como suele ser habitual en entornos clínicos (Santisteban, 1990). La correlación entre la puntuación de la prueba y la medida criterio se denomina “Coeficiente de Validez”, y es el índice que se utiliza para informar de la validación de los instrumentos, siendo deseable detallar su obtención y calcularlo (con su correspondiente criterio) para cada rendimiento o comportamiento que pretendamos medir con el mismo instrumento (Barbero, 1999b). Para lograr una generalización de la validación empírica, se recomienda el uso de estrategias de meta análisis que controlen la variabilidad de los resultados de diferentes grupos (Sánchez-Meca y Botella, 2010).

En segundo momento (Etapa Teórica) se impone el constructo de validez de constructo, como concepto unificador que integra los diferentes tipos de validez descritas anteriormente (APA, AERA y NCME, 1985). La validez estructural o de constructo es el grado en el que puede afirmarse que el instrumento mide un determinado constructo o rasgo teórico (Bornsboom, 2005). Al fijarse en un elemento más permanente y de una clase más abstracta que los tipos de validez anteriormente expuestos, requiere de una acumulación gradual de información procedente de diferentes fuentes (Anastasi y Urbina, 1998). Según Messick (1980) la validez de constructo es un concepto global que integra las consideraciones de validez de contenido y de criterio en un marco común. Idea que es aceptada por otros autores que otorgan a este concepto un papel central en la validación de los instrumentos (Tenopyr, 1977; Guion, 1980; Moss, 1992). Los procedimientos metodológicos más utilizados para la obtención de datos referentes a la validez son: la “Matriz Multirasgo-Multimétodo” y el “Análisis Factorial” (Tomás, Oliver y Hontangas, 2000).

Campbell (1960) señaló que la validación estructural debería implicar dos condiciones: que los resultados de la prueba correlacionen positivamente con las variables que teóricamente deben estar correlacionadas (validación convergente) y que estos resultados no correlacionen significativamente con las variables que deberían diferir (validación discriminante). Para tener en cuenta este enfoque dual, Campbell y Fiske (1959) propusieron un plan experimental sistemático para la validación de constructo denominado “Matriz Multirasgo-Multimétodo”, que consiste en una tabla que cruza dos o más rasgos (constructos) con dos o más métodos (instrumentos), señalando sus respectivas correlaciones (coeficientes de validez). Para que la validez estructural fuera satisfactoria, los coeficientes de validez deberían ser más altos que las correlaciones entre rasgos diferentes, medida por métodos diferentes y también ser más altos que las correlaciones entre rasgos diferentes, medidas por el mismo método. Con posterioridad se han planteado críticas a este modelo que problematizan el uso del análisis de correlación bivariado ya que las diferencias entre las fiabilidades de las distintas medidas pueden distorsionar las relaciones entre las diferentes variables (Schmidt y Stults, 1986).

Como respuesta a estas limitaciones se plantearon otras metodologías de cálculo de la validez entre las que destaca el uso del “Análisis Factorial” (Kenny y Kashy, 1992). El “Análisis Factorial” es una técnica estadística de análisis multivariado que permite realizar una

estimación de los factores que agrupan una serie de variables analizadas, pudiendo reducir los datos a grupos homogéneos según su nivel de covarianza (Abraira y Pérez de Vargas, 1996). En un instrumento, esto implica correlacionar todos sus elementos constituyentes con el resto de elementos (Anastasi y Urbina, 1998). El Análisis factorial se puede plantear desde dos perspectivas: exploratoria o confirmatoria. La aproximación exploratoria o inductiva supone analizar los elementos para buscar patrones de relación entre ellos y a posteriori definir un constructo a partir de esas agrupaciones. La aproximación exploratoria o confirmatoria trata de contrastar una hipótesis teórica planteada con anterioridad (modelo preestablecido) viendo si efectivamente los elementos que se prevén relacionados se agrupan en el constructo de la teoría, representando este último una aproximación fuerte a la validación de constructo de los instrumentos (Pérez-Gil, Chacón y Moreno, 2000).

En el momento actual (Etapa contextual) se presenta un concepto unitario de la validez como proceso, donde ya no se habla de diferentes tipos de validez, sino de distintas fuentes de evidencia (todo lo descrito anteriormente) en un único proceso y donde se añade el matiz de que es necesario delimitar la validez en el contexto de su uso real (APA, AERA y NCME, 1999). Esta faceta del uso propuesto condiciona su validez al contexto de uso y hace necesario que se especifique en la prueba las condiciones de situación de medida, teniendo en cuenta la relevancia y utilidad de los resultados para los fines propuestos (Messick, 1989). Por ello, el objetivo de los estudios de validez será recoger suficientes evidencias que puedan prestar base científica a la interpretación de los resultados en un uso concreto por parte del evaluador, siendo éste el que deba valorar la adecuación del contexto a los requerimientos de la validez (Elosua, 2003). Los últimos estándares (APA, AERA y NCME, 1999) resaltan cinco fuentes de evidencia de validez que se encuentran en el análisis: del contenido, del proceso de respuesta, de la estructura interna de la prueba, de las relaciones con otras variables y de las consecuencias de la prueba.

Respecto al análisis del contenido se incluyen las reflexiones anteriores acerca de validez de contenido pero añadiendo el estudio de los factores contextuales tales como las propias situaciones de evaluación: instrucciones, relación examinador-examinado, formato de administración, diferencias de motivación... (Holmbeck y Devine, 2009).

El análisis de procesos de respuesta es un elemento relativamente reciente como fuente de validez provocado por las demandas de la psicología cognitiva, especialmente ante las pruebas de rendimiento máximo, que consiste en analizar los procesos mentales que usan los sujetos para responder a cada elemento de la prueba (analizado a través de diferentes estrategias como entrevistar a los sujetos acerca de su procedimiento, analizar tiempos de respuesta o registrar los movimientos oculares). Si se constata que los procesos utilizados se ajustan a los que postulan las teorías relativas al constructo medido, se puede considerar válido ese elemento de la prueba (Goodwin y Leech, 2003).

En el análisis de la estructura interna de la prueba se integran los diferentes desarrollos precedentes referidos a la validez de constructo y se añaden nuevas perspectivas. Se diferencian dos estrategias dentro de este análisis: la dimensionalidad de la prueba (análisis de los factores subyacentes a la prueba en su conjunto, a través de estrategias como el ya mencionado análisis factorial) y el funcionamiento diferencial de los ítems (analiza la interacción de cada elemento de la prueba con las características del sujeto que la realiza, con técnicas derivadas de los modelos de respuesta al ítem y factorial) (Gómez-Benito, Hidalgo y Guilera, 2010).

El análisis de la relación de la prueba con otras variables, responde a la perspectiva clásica sobre la validez externa de relacionar las medidas de la prueba con alguna otra medida, criterio que se espera que pronostique (Loevinger, 1957). Sus estrategias son las que se han planteado anteriormente en este mismo apartado al hablar de la primera etapa conceptual de la validez.

Por último, el análisis de las consecuencias de la aplicación de la prueba o técnica de evaluación es el elemento que más controversias teóricas despierta como fuente de validez, ya que no existe acuerdo si la consideración de estas, debe formar parte del proceso de validación o no (Messick, 1980; Meherens, 1997; Popham, 1997). La postura de los actuales estándares (APA, AERANC y NCME, 1999) delimita esta faceta a componentes concretos tales como: el análisis de la posible infrarrepresentación del constructo evaluado o a la existencia en la prueba de partes irrelevantes para ese constructo; dejando fuera del análisis de la validez otros elementos muy relevantes pero que tal vez se corresponden con otro orden de reflexión como puede ser la ética profesional.

## **2.4. Bondad ética de la evaluación psicológica**

Para Lindsay (2009) la Psicología no solo encuentra su fundamento en el conjunto de conocimientos y habilidades que ha ido adquiriendo a través de la investigación y la práctica, sino que también se fundamenta en la ética de las acciones que realiza en su entorno. Y aunque en el pasado hubo voces discordantes frente a la necesidad de generar códigos éticos profesionales que diesen cuenta de una reflexión ética colectiva, en la actualidad no hay duda de la relevancia de este componente en cualquier área de actuación de la psicología, incluyendo a la Evaluación (Lindsay, Koene, Ovreide y Lang, 2008).

### **2.4.1. Código Deontológico del psicólogo**

El término deontología utilizado por primera vez por Bentham (1836) se entiende como una rama de la ética ocupada de regular los fundamentos del deber y las normas exigibles para un colectivo profesional concreto. En la actualidad se ha llegado a asumir que una de las características definitorias de una profesión es precisamente la existencia de un código ético específico (Prywansky y Wendt, 1999). Llegándose a asegurar que la deontología es la esencia y razón de ser de los colegios profesionales (Colegio Oficial Psicólogos COP Madrid, 2011). De hecho, en España, la Ley de Colegios Profesionales de 13 de febrero de 1974 (Boletín Oficial del Estado BOE, 1974) encargó a estos, la función de ordenar la actividad profesional de sus miembros velando por la ética, dignidad profesional y respeto a los derechos, manteniendo, defendiendo y promoviendo la deontología profesional. En psicología se empezaron a generar códigos deontológicos en las asociaciones profesionales en la década de los cincuenta y desde entonces se han desarrollado numerosas propuestas en las diferentes partes del mundo (Kimmel, 1996).

En el contexto Europeo, se ha establecido un meta-código entre las diferentes asociaciones de psicólogos que enumera cuatro principios éticos fundamentales con los que los códigos propios de cada zona deben estar en coherencia (European Federation of Psychologists' Associations EFPA, 2005):

- Respeto: por los derechos de la persona y su dignidad. Respetar la autodeterminación de la persona, su libertad de consentimiento, su derecho de confidencialidad y privacidad.

- Competencia profesional: obligación de conocer el alcance de técnicas y procedimientos que se están utilizando, así como una correcta formación en estos. Reconociendo las propias limitaciones.
- Responsabilidad: frente a sus clientes y la comunidad en la que trabajan. Evitando producir daños en estos o el uso inapropiado de los conocimientos o prácticas psicológicas.
- Integridad: Honestidad y sinceridad frente a clientes. No incurriendo en engaños, ni aprovechando la situación privilegiada como profesional para sacar beneficios personales de los clientes.

En España, la elaboración de su primera versión del Código Deontológico del Psicólogo (CDP) tuvo un carácter inminentemente práctico alejado de grandes principios éticos, una puesta en orden racional de ciertas actividades y modos de realizar las actividades (Fiero, 1987). Por ello, desde la inclusión del Colegio Oficial de Psicólogos de España en la Federación Europea de Asociaciones de Psicólogos (EFPA) se ha promovido una modificación del Código para destacar los principios enunciados (Rodríguez-Sutil, 1996).

A parte de la inclusión de principios éticos generales, Bermejo (2009) destaca un elemento muy relevante a tener en cuenta en el necesario cambio paradigmático del Código Deontológico: la legislación sobre datos de carácter personal en el contexto europeo. La Directiva de la Comunidad Europea (Diario Oficial de la Unión Europea DOUE, 1995), otorga al titular de los datos personales derecho de control sobre los mismos y unas garantías de protección y amparo institucional ante el uso de terceros de esos datos, lo que significa un control del uso de los propios datos personales que supera la consideración de “derecho a la intimidad” del que se tenía en esos momentos en España (BOE, 1982). Este reconocimiento de la autonomía de la persona respecto de sus datos en cualquier contexto se ha materializado en la legislación española con la aprobación de la Ley Orgánica de Protección de datos 15/1999 (BOE, 1999) y posteriormente mediante la creación de un Reglamento específico que desarrolla esta ley orgánica (BOE, 2008). Obligando a cualquier profesional a establecer garantías de acuerdo a las normas de protección de datos y dando gran relevancia al consentimiento del interesado del uso de sus datos. Se crea además la Agencia de Protección de Datos con potestad sancionadora y capacidad de actuación en instituciones públicas y privadas que vele por el correcto cumplimiento de la normativa. En

este mismo sentido en el contexto sanitario se aprueba la Ley 41/2002 (BOE, 2002) que regula los derechos y obligaciones en materia de información y documentación clínica, que concede al paciente un papel más activo respecto al uso de sus propios datos clínicos, además de regular los derechos y deberes que conlleva el acceso, uso y conservación de esos datos.

#### **2.4.2. Ética en la evaluación psicológica**

Por todo lo anterior, no es extraño que la última modificación CDP en España (Consejo General de Colegios Oficiales de Psicólogos CGCOP, 2009), no solo recoja de forma implícita los principios enunciados por la EFPA sino que además dedique el 5º apartado de forma específicamente a la obtención y uso de la información atendiendo a la legislación vigente.

De forma específica en el área de la evaluación psicológica para Del Barrio (2003) los posibles conflictos éticos que se le pueden presentar al profesional se concretan en dos problemas básicos:

- Invasión de la intimidad: la percepción del cliente o usuario de que se viola su intimidad, recogiendo información innecesaria o que va más allá del interés de su demanda explicitada.
- Comunicación de resultados (o uso de la información): el conflicto entre la obligación de informar del profesional a partes legítimamente interesadas y el derecho a la confidencialidad de la persona. A quién y cómo informar.

Para afrontar el primer problema Anastasi y Urbina (1998) sugieren atender a dos principios:

- Adecuación de la evaluación: implica que lo evaluado sea apropiado, conveniente y necesario para los propósitos de la evaluación. Pudiendo ser fácilmente justificable por el profesional ya que recabará la información estrictamente necesaria para el desempeño de su labor. Tal y como se explicita en el Artículo 39 de la última versión del CDP España (CGCOP, 2009).
- Consentimiento informado (o con conocimiento): implica solicitar autorización del usuario o cliente o de sus representantes legales si es pertinente, informándole previamente sobre el propósito de la evaluación, qué información se va a recabar, la

utilidad que va tener esa evaluación para su persona y el empleo que se va a hacer con la información así obtenida. Un uso adecuado de este principio presupone participación voluntaria sin coacción del evaluado, que se otorgará suficiente información para poder tomar una decisión y que la persona disponga de la capacidad mental suficiente para otorgar ese consentimiento de forma consciente (Gregory, 2001).

Con respecto al segundo problema, Pelechano (1988) señala que se ha de informar siguiendo el principio de “primum non nocere”. Lo que significa que se ha de informar al interesado, y si procede a los terceros legítimamente interesados (y a estos en la medida de lo posible sólo de lo que sea estrictamente necesario para la finalidad de la evaluación), pero procurando no generar perjuicio para el propio evaluado. Adicionalmente Gregory (2001), vuelve a señalar el uso previo del conocimiento informado para detallar al usuario las limitaciones legales de la confidencialidad y las particularidades de cada situación de evaluación, en especial las condiciones que impliquen información a terceros. Un ejemplo claro de este tipo de situación potencialmente conflictiva es la evaluación de adolescentes (cuya condición de menor identifica como terceros legítimamente interesados a los progenitores o tutores legales), donde el consentimiento informado del propio menor así como su implicación en el proceso de transmisión de información a terceros puede evitar desavenencias con el menor (Silva, 1995). Esta circunstancia se recoge de forma específica en los artículos 40º, 41º, 42º y 43º de la última versión del CDP España (CGCOP, 2009) incidiendo en el consentimiento del usuario del uso de la información, limitar la información a terceros a lo indispensable y la generación de informes de calidad (claros y útiles).

La evaluación en contextos científicos o experimentales requiere de la misma consideración ética con los sujetos de estudio que con los demás tipos de usuarios, lo que implicará necesariamente un consentimiento informado, el derecho a abandonar la investigación en cualquier momento, la protección de la intimidad y confidencialidad con respecto a los datos recogidos, la no generación de daños al evaluado y el desengaño al finalizar el estudio si han existido encubrimientos o engaños necesarios para que se diesen las condiciones experimentales (Wadeley y Blasco, 1995). Sobre este contexto se expresan los artículos 34º, 35º 36º y 37º de la última versión del CDP España (CGCOP, 2009).

### **2.4.3. Principios bioéticos del ámbito sanitario**

No existe actualmente acuerdo entre las diferentes asociaciones de psicología de asumir de forma explícita los postulados bioéticos de las profesiones sanitarias en sus códigos deontológicos, de hecho, el Meta-código Europeo no los ha incluido hasta ahora por falta de consenso (Bermejo, 2009). Pese a ello, se ha considerado relevante incluir una mención a estos para concluir el apartado, ya que la ubicación del presente trabajo es precisamente la evaluación en contextos sanitarios o clínicos. Beauchamp y Childrees (1999) enuncian los cuatro grandes principios de bioética que deben de tenerse en los procesos de toma de decisiones que afecten a la vida humana en los entornos sanitarios:

- **Autonomía:** Todos los individuos deben ser tratados como seres autónomos y con libertad para tomar decisiones sobre su propia persona. La materialización práctica de este principio en el entorno clínico es la existencia del consentimiento informado.
- **Beneficencia:** La obligación moral del profesional de actuar en beneficio del usuario. Optando por la alternativa más beneficiosa para resolver el padecimiento del usuario por encima de cualquier otro criterio.
- **No Maleficencia:** Que el profesional evite dañar o perjudicar intencionadamente al usuario. Debido a que en el entorno sanitario es relativamente común generar algún tipo de daño para generar un bien, se puede matizar este principio como no perjudicar innecesariamente a otros.
- **Justicia:** Supone el reparto equitativo tanto de las cargas como de los beneficios. Es un rechazo a la discriminación de cualquier tipo. En un entorno de recursos limitados es el respeto al reparto igualitario de estos, frente a otras consideraciones.

Estos principios sirven como guía general y no se pueden tomar como normas precisas de acción, puesto que tienen un carácter orientador, además es frecuente que entren en conflicto entre sí de forma que sea necesario ponderarlos según el contexto de referencia (Beauchamp, 2003).

## Capítulo 3

### Evaluación de Resultados en salud general

---

La evaluación de resultados se ha ido convirtiendo en un aspecto cada vez más relevante para cualquier ámbito de la asistencia sanitaria (Sinha, Poloniecki, Thompson y Holt, 2012). Antes de focalizarse sobre su implicación concreta en la evaluación de intervenciones psicológicas, en el presente capítulo se pretende realizar una síntesis de los conceptos y desarrollos relevantes para la comprensión del fenómeno en salud general. Se presente inicialmente los dos precedentes históricos que han generado que se le preste tanta atención en la actualidad (tanto desde el punto de vista de la practica clínica, como el de la gestión de centros). Posteriormente se realizan una serie de aclaraciones conceptuales sobre términos vinculados a esta área. Y se finaliza presentando una perspectiva contemporánea de investigación en el ámbito sanitario, que se centre en la relevancia clínica de las intervenciones, y su estudio en contextos reales: Investigación de Resultados en Salud (IRS).

#### 3.1. Antecedentes conceptuales y contextuales

Se pueden reconocer principalmente dos hitos dentro del desarrollo del área sanitaria que perfilaron la consideración actual de la evaluación de resultados (Emilfork, 2009): la Gestión de la Calidad de la asistencia sanitaria y la Medicina Basada en la Evidencia (MBE).

##### 3.1.1. Gestión de la calidad en servicios sanitarios

El concepto de calidad tiene su origen en la evolución industrial y la comercialización de sus productos y ha ido pasando por diversos estadios conceptuales en un sentido cada vez más amplio: desde su naturaleza inicial de evaluación de productos (contraste de las características finales de un producto con un estándar preestablecido); a la perspectiva contemporánea de calidad total (donde se integran todas las áreas de las organización y del proceso productivo para satisfacer las necesidades de todos los agentes implicados) (Berlinches, 2004).

De forma general la calidad se entiende como el grado en el que un conjunto de características de un producto o servicio cumple con una serie de requisitos. Entendiendo como requisitos las necesidades o expectativas de las diferentes partes interesadas (Asociación Española de Normalización y Certificación AENOR, 2005; 2008). Para Ishikawa (1986) la calidad se traduce a nivel empresarial en diseñar, producir y servir un bien o servicio que sea útil, lo más económico posible y siempre satisfactorio para el cliente, orientando la estrategia global de la empresa a este fin.

En la segunda mitad del siglo XX se evidenció la superioridad de los modelos de gestión centrados en la calidad frente a los que respondían únicamente al paradigma clásico (de súper-especialización de la producción), tanto en términos de rentabilidad como de satisfacción del cliente (Womack, 1992). Por ello, no es de extrañar que el ideal actual de excelencia organizacional occidental se acomode a los parámetros de gestión de la calidad total (Fundación Vasca para la Excelencia Euskalit, 2014). En estos modelos se presupone una continuada evaluación de resultados y de impactos de la producción y distribución (a través de herramientas fiables en todos los niveles) para ir perfeccionando cada vez más el proceso empresarial (Gómez, Vilar y Tejero 2003; García, 2007).

Sin embargo, como señala Rosander (1992) este paradigma industrial de la calidad necesita de una serie de matizaciones para su aplicación a las organizaciones que ofrezcan servicios en vez de productos totalmente tangibles. Ya que en las empresas de servicios:

- La calidad percibida por el cliente está determinada por las actitudes, comportamiento y la competencia profesional del que les atiende.
- Algunas características exigen el máximo nivel de perfección, ya que en determinadas actividades no se pueden aceptar errores.
- La calidad que aprecia el cliente es diferente de la percepción de la organización. Por lo que es necesario tenerla en cuenta.
- La planificación de la mejora tiene que estar basada en datos para lo cual se deben recoger en tiempo real diferentes indicadores del servicio.
- La calidad implica incidir sobre factores psicológicos tales como elementos relacionales, comunicacionales y actitudinales.

- El factor clave en la calidad es la presencia de una formación adecuada que asegure la capacidad operativa del profesional.

Para exponer la repercusión de la gestión de la calidad en los servicios sanitarios se va a recurrir al ejemplo específico que representa la figura del hospital, ya que como señala Jiménez (2002) es la unidad clave y paradigmática de las organizaciones de asistencia sanitaria, y la mayoría de los estudios y modelos propuestos en asistencia sanitaria se han desarrollado en el seno de la compleja gestión de los hospitales.

Relman (1998) identifica tres etapas en la cuestión de la calidad en la asistencia sanitaria, vinculándola a la evolución de la gestión hospitalaria. Aunque señala que ya hay antecedentes particulares de inquietud por evaluar la calidad de los resultados, concretamente en el área de cirugía, no es hasta después de la Segunda Guerra Mundial cuando se empiezan a realizar cambios paradigmáticos dentro de los hospitales:

- Etapa de la Expansión (en los años cincuenta y sesenta): se crearon muchos hospitales y otros servicios. Los hospitales aumentaron de tamaño y gracias al desarrollo científico también lo hicieron el número de especialidades sanitarias. El objetivo era hacer accesible a toda la población los servicios de salud profesionales.
- Etapa de la Contención de Costes (en los años setenta y ochenta): se hizo evidente que se tenía que lidiar con una realidad donde los recursos eran limitados. Fue cuando cobró protagonismo la regulación de presupuestos, la profesionalización de los gestores, la aplicación de sofisticados sistemas de contabilidad y con la aparición de la informática el control de numerosos indicadores logísticos y económicos.
- Etapa de la Evaluación (década de los 90 en adelante): se hace patente la necesidad de encontrar un equilibrio entre la cantidad y el coste de los servicios. Es cuando realmente se generaliza el control de la calidad en los hospitales, se quiere mejorar la calidad y la efectividad de la atención sanitaria y ofrecer información como base suficiente para la toma de decisiones clínicas y de gestión, por ello son necesarios mejores sistemas de información y una recogida más meticulosa de indicadores. Se asumió también en el ámbito público la orientación al cliente, entendiendo que se tenía que dar respuestas a las necesidades y expectativas de los pacientes.

Para Cuesta, Moreno y Gutiérrez (1986) la atención hospitalaria tiene mayor o menor grado de calidad en la medida en que sus atributos se aproximan a los del prototipo, es decir a la práctica ideal de ese proceso que ha sido diseñado para cubrir la necesidad del paciente. Pero el problema, como señala Varo (1995), es identificar qué atributos son los componentes de la asistencia sanitaria de calidad, es decir la dificultad está en elaborar el prototipo y posteriormente poder medirlo.

Donabedian (1984) hace una propuesta sistémica sobre la composición de la asistencia sanitaria, habla de tres grupos de factores que configurarían los atributos a tener en cuenta para juzgar la calidad de la asistencia. En primer lugar, los aspectos técnicos y científicos, referidos exclusivamente al uso de las técnicas científicas más adecuadas para la resolución de la patología del paciente, su valoración estará en función de los avances científicos de esa área de la salud y a la cualificación de estos del profesional. El siguiente grupo hace referencia a los factores interpersonales que se refieren a la relación entre profesional y paciente y a las expectativas y necesidades del paciente, que también influirán en el correcto desempeño del primer grupo, dado que son necesarias unas vías de comunicación sin obstáculos para elaborar un diagnóstico y contar con el paciente para establecer un tratamiento lo más satisfactorio posible según las condiciones particulares del paciente. Y el tercer grupo, el que él denomina “amenidades” hace referencia al entorno integrando de factores ambientales o derivados de estos como la accesibilidad al servicio, comodidad o la intimidad.

La Organización Mundial de la Salud (OMS/WHO, 1985, p.5) recogió esta visión sistémica en su propia definición de la calidad de la asistencia sanitaria que dice así: “Asegurar que cada paciente reciba el conjunto de servicios diagnósticos y terapéuticos más adecuados para conseguir una atención sanitaria óptima, teniendo en cuenta todos los factores y los conocimientos del paciente y del servicio médico, y lograr el mejor resultado con el mínimo riesgo de efectos iatrogénicos y la máxima satisfacción del paciente en el proceso”

Donabedian (1991) posteriormente subraya que también existe una perspectiva social de la calidad asistencial, que entra en conflicto con los factores más individuales que se reflejan en los tres grupos que enunció en el pasado, y es la consideración del beneficio o utilidad neto de toda la población, es decir ponderar el coste social de las intervenciones. Esta matización es solo aplicable a los servicios de salud públicos.

En las conceptualizaciones anteriores, no se hace referencia al componente económico pero como antes hemos mencionado a partir de la década de los 70, se empieza a hacer patente la necesidad de gestionar los hospitales teniendo en cuenta variables económicas además de las de salud debido a que los recursos tanto en la sanidad pública como en la privada son limitados (Del Llano-Señarís, Ortún, Martín, Millán, Gené y Campella, 2003). Los costes de la asistencia sanitaria se deben contemplar en la definición de su calidad y tenerse en cuenta al evaluarla. De los métodos de evaluación económica de resultados, los análisis coste-efectividad, coste-eficacia y coste-eficiencia resultan los más interesantes para facilitar las decisiones sanitarias ya que se ponderan los costes de oportunidad pudiendo identificar entre las opciones posibles la más adecuada (Drummond, Stoddart y Torrance, 1991). Nos extenderemos más en estos métodos, en este mismo capítulo al definir la eficiencia en el ámbito sanitario.

Siguiendo la lógica de incluir los costes en la consideración de la calidad Donabedian (1991), diferencia calidad eficaz de calidad óptima. La primera representa a toda la atención necesaria para mejorar la salud lo más posible, independientemente de sus costes. La segunda pretende lograr la máxima utilidad de la asistencia a un coste razonable. La calidad óptima vendría dada por el punto en el cual la relación directa entre el coste y el beneficio de la asistencia tiene una eficacia máxima, considerando que se puedan sufragar con los recursos disponibles, tanto para el cliente como para la organización.

Todo lo anterior muestra la complejidad del fenómeno de la calidad asistencial y por qué es necesario definirlo de una manera multidimensional descomponiéndolo en componentes de cara a poder medirlos y controlarlos (Herrera, Pino, Mora, Doblas y Castillo, 2001).

Varo (1995) realiza una síntesis de los principales componentes determinantes de la calidad asistencial que coinciden con los destacados actualmente por la Joint Commission on Accreditation of Healthcare Organizations (2005): efectividad, eficacia, seguridad del paciente, competencia del profesional, calidad científica-técnica, satisfacción del paciente, satisfacción del profesional y disponibilidad, accesibilidad y continuidad del servicio (se incidirá sobre algunos de estos elementos más adelante).

Tempranamente, en la reflexión sobre la estimación de la calidad se acuñó el término "outcomes" -resultados, desenlaces o consecuencias- (Donabedian, 1966), como validantes

fundamentales de la calidad del cuidado sanitario. Y la búsqueda de la evaluación de estos "outcomes" en las diferentes facetas que se han señalado a lo largo de este apartado, representó la génesis de la Investigación de Resultados en Salud (Emilfork, 2009).

### **3.1.2. Medicina Basada en la Evidencia (MBE)**

La MBE tiene su origen en la Escuela de Medicina de la Universidad de McMaster en Ontario, donde un grupo de clínicos y epidemiólogos, iniciaron este movimiento que ha tenido una enorme repercusión en la asistencia sanitaria a nivel mundial (Adefna, Leal, Izquierdo, Castellanos y Arocha, 2003). Uno de sus fundadores Guyatt (Guyyat y Rennie, 2002) define la MBE como un proceso cuyo objetivo es el de obtener y aplicar la mejor evidencia científica disponible en el ejercicio de la práctica médica cotidiana. Esta perspectiva deriva de los desarrollos de la Epidemiología Clínica (Jiménez, 2012) que se centraron en: evaluar la eficacia terapéutica de las diferentes técnicas, la utilidad de los medios diagnósticos y la comprobación de hipótesis de factores de pronóstico. Todo ello con el objeto de mejorar la toma de decisiones clínica mediante estudios científicamente rigurosos, tomando un gran protagonismo el Ensayo Clínico Controlado (Feinstein, 1985).

Los percursores de la MBE se percataron de que en la práctica clínica no se asumían los nuevos descubrimientos y avances terapéuticos y diagnósticos generados en la rigurosa investigación de por ejemplo la Epidemiología Clínica. Se perduraban medios menos eficaces que los señalados por las nuevas evidencias. Parecía existir pues una disonancia entre el avance científico académico y la práctica clínica real (Sackett y Rosenberg, 1995). Se hizo patente en diferentes estudios (Covell, Uman y Manning, 1985; Ramsey et al., 1991) que los profesionales sanitarios tendían a quedarse obsoletos según pasaba el tiempo desde que se habían graduado y apenas dedicaban tiempo en su día a día para actualizar sus conocimientos clínicos en consonancia con los últimos avances.

Para cambiar esta situación se propuso un nuevo enfoque de la práctica médica al que llamaron Medicina Basada en la Evidencia. Se planteó un proceso concreto en el cual el profesional transformará cualquier problema clínico en preguntas, y después, de forma sistemática, localizase, recuperase y utilizase los hallazgos de investigación actuales como base para la toma de decisiones clínicas. Se señaló que la generación de preguntas y búsqueda de

respuestas iba más allá del esquema clásico de proceso diagnóstico-tratamiento (basado en el conocimiento y experiencia del clínico), ya que se deberían incluir búsquedas eficaces de evidencia adecuada (sin sesgos) en la literatura científica, primero para diagnosticar y posteriormente para tratar (Rosenberg y Donald, 1995). Para ello, el clínico debería tener la capacidad y los medios para poder buscar y evaluar críticamente la evidencia según su validez y utilidad, y así incorporar los medios más eficaces a su práctica (Oxman, Cook, Guyatt y Evidence-Based Medicine Working Group, 1994).

Para facilitar esta tarea a los clínicos, se señaló la necesidad de generar revisiones sistemáticas en la literatura científica en forma de búsquedas exhaustivas de artículos con respuesta potencial a una cuestión clínica determinada y una correcta síntesis de lo encontrado (Cook, Mulrow y Haynes, 1997). En este mismo sentido se subrayó la utilidad de las revistas especializadas en publicar resúmenes filtrados de resultados de las investigaciones en diversas áreas (Davidoff, Haynes, Sackett, y Smith, 1995). También se promovió la confección de “guías de práctica clínica basadas en la evidencia”, donde se detallase cada paso en la atención de una determinada problemática de salud, respaldándolo por cierto “nivel de evidencia” en consonancia con la forma en que se ha llegado a obtener cada conocimiento (Browman, Gómez, Haynes, Jadad y Gabriel, 2001). Otro punto sobre el que se incidió especialmente fue la necesidad de mejorar la accesibilidad de estas herramientas de evidencia a los profesionales de salud para poder hacer factible su incorporación a la práctica diaria: mejor acceso a fuentes electrónicas de información, generación de “carros” de evidencia en los centros (repositorios de evidencia sobre el área de interés rápidamente accesibles) y posibilidad real de formación continuada mediante cursos y congresos útiles (Sackett y Strauss, 1998; Grol y Grimshaw, 2003).

Un herramienta singularmente destacada de la propuesta de la MBE, es el reporte de los denominados “niveles de evidencia” de las investigaciones, según los cuales cada tipo de resultado evaluado genera más o menos evidencia, de acuerdo con el tipo de diseño empleado para su obtención. Consiste en sistematizar la evidencia externa sobre la base de unas escalas que la clasifican y valoran su calidad, y así hacer posible emitir recomendaciones sobre la adopción técnicas y procedimientos asistenciales según el rigor científico del tipo de diseño en el cual se ha generado sus evidencias (Brito, González y Cruz, 2010). En este sentido existen

diferentes propuestas para clasificarla, pero todas son bastante similares entre sí, estableciendo un “nivel 1” que representa la mejor evidencia posible y posteriormente desplegando los siguientes niveles una jerarquía, en orden decreciente de calidad (Primo, 2003). El primer intento de crear una clasificación así, la realizó la Canadian Task Force on the periodic Health Examination CTFHE (Morgan y CTFHE, 1979) que posteriormente fue adaptada y actualizada en Estados Unidos por la U.S. Preventive Services Task Force USPSTF (Harris et al., 2001). Como ejemplo paradigmático de estas clasificaciones a continuación se reproduce la gradación de la Agència d’Avaluació Tecnologia Mèdica ATTM de la Generalitat de Catalunya (Jovell y Navarro-Rubio, 1995):

- *Nivel 1.* Evidencia Adecuada. Tipo de Diseño: Meta-análisis de Ensayos Clínicos Aleatorizados y Controlados. Condiciones de rigurosidad: análisis de datos individuales, diferentes técnicas de análisis, sin heterogeneidad, meta-regresión, meta-análisis y calidad de los estudios
- *Nivel 2.* Evidencia Adecuada. Tipo de diseño: Ensayos Clínicos Aleatorizados y Controlados con una muestra grande. Condiciones de rigurosidad: evaluación del poder estadístico, multicéntrico y calidad del estudio
- *Nivel 3.* Evidencia Buena a Regular. Tipo de diseño: Ensayos Clínicos Aleatorizados y Controlados de muestra pequeña. Condiciones de rigurosidad: evaluación del poder estadístico y calidad del estudio.
- *Nivel 4.* Evidencia Buena a Regular. Tipo de diseño: Ensayo prospectivo controlado no aleatorizado. Condiciones de rigurosidad: Controles coincidentes en el tiempo, multicéntrico
- *Nivel 5.* Evidencia Regular. Tipo de diseño: Ensayo retrospectivo controlado no aleatorizado. Condiciones de rigurosidad: controles históricos y calidad del estudio.
- *Nivel 6.* Evidencia Regular. Tipo de diseño: Estudios de cohorte. Condiciones de rigurosidad: multicéntrico y calidad del estudio.
- *Nivel 7.* Evidencia Regular. Tipo de diseño: Estudios de casos y controles. Condiciones de rigurosidad: multicéntrico, apareamiento y calidad del estudio.
- *Nivel 8.* Evidencia Pobre. Tipos de diseño: Estudios de descriptivos (vigilancia epidemiológica, encuestas, registros y bases de datos). Series clínicas no controladas,

comités de expertos y conferencias de expertos Condiciones de rigurosidad: Multicéntrico.

- *Nivel 9*. Evidencia Pobre. Tipo de diseño: Anécdotas de la experiencia y casos únicos.

Las citadas clasificaciones, junto a la mayoría de las propuestas realizadas en este sentido en la literatura, se centraron en jerarquizar la evidencia de las investigaciones en resultados de intervenciones de prevención o tratamiento. Dando un gran protagonismo a los ensayos clínicos controlados con asignación aleatoria y a las revisiones sistemática de estos (Manterola y Zavado, 2009). Sin embargo, Sackett (1997) uno de los principales precursores del MBE, matizó que al jerarquizar los diseños había que tener en cuenta también que tipo de pregunta se buscaba responder, es decir, sobre qué escenario de la práctica clínica se estaba buscando evidencia. Por ejemplo, la calidad de la evidencia del uso de un diseño de cohorte para evidenciar factores pronósticos en una patología será mucho mayor que su uso para evaluar resultados de intervenciones (Doménech, 2013a). La clasificación del Centre for Evidence-Based Medicine de Oxford- CEBM (2011) trata de reflejar esta realidad incluyendo jerarquías específicas para tres áreas clínicas a parte de las referidas a la prevención y tratamiento como son el diagnóstico, el pronóstico o los análisis económicos.

Como consecuencia de la jerarquización de los niveles de evidencia, en estas clasificaciones se ofrece también un código de graduación, que señala lo recomendable de la adopción de lo evidenciado en la práctica clínica, donde identifica el grado de recomendación máximo con una A y los sucesivos por las siguientes letras del alfabeto en orden descendente (Jovell y Navarro-Rubio, 1995). A continuación se reproduce la última versión de los grados de recomendación de la USPSTF (2014), en la que no solo se tiene en cuenta el nivel de evidencia sino también el balance de riesgos-beneficios de su implementación en la práctica clínica:

- A. Extremadamente recomendable. Buena evidencia de que la medida es eficaz y los beneficios superan ampliamente a los perjuicios.
- B. Recomendable. Moderada evidencia de que la medida es eficaz y los beneficios superan a los perjuicios.

- C. Ni recomendable ni desaconsejable. Recomienda no usarlo de forma rutinaria. Al menos moderada evidencia pero los beneficios son muy similares a los perjuicios y no puede justificarse una recomendación general.
- D. Desaconsejable. Al menos moderada evidencia pero los perjuicios superan a los beneficios.
- E. Evidencia Insuficiente, de mala calidad o contradictoria, y el balance entre beneficios y perjuicios no puede ser determinado.

Aunque los sistemas de clasificación propuestos parten de una base similar existen algunas discrepancias que pueden ir en decremento de la comprensión de estas graduaciones por parte del clínico. Por ello, se está trabajando desde la MBE en unos criterios unificados y más flexibles que aúnen los matices reflejados por los principales sistemas de clasificación antes referenciados, en una iniciativa denominada Grading Of Recommendations Assessment GRADE (Atkins et al., 2004; Guyatt et al., 2008).

La MBE ha recibido duras críticas desde poco después de su concepción (Jiménez, 2012). Críticas que le acusaban de tratar de imponer autoritariamente prácticas validadas en escenarios externos a la clínica real, como si de un libro de recetas se tratase, desdeñando la variabilidad de la realidad clínica y otras fuentes de evidencia distintas al ensayo clínico controlado (Tannenbaum, 1995; Charlton, 1997). Los impulsores de la MBE no tardaron en aclarar que el objetivo de esta, era promover el uso consciente y juicioso de la mejor evidencia clínica para la tomar decisiones individual, pero sin tratar de sustituir la experiencia clínica del profesional (más bien complementarla), ya que ambos elementos era necesarios para generar la mejor asistencia (Sackett, Rosenberg, Gray, Haynes y Richardson, 1996).

Sin embargo, Feinstein y Horwitz (1997) pusieron en duda también la consideración de mejor evidencia para el ámbito clínico, ya que esta para la MBE descansa singularmente en el ensayo clínico controlado. Charlton (1995) señala que los grandes ensayos clínicos no están exentos de sesgos de selección: se realizan en contextos especializados, la proporción de sujetos que entran a formar parte del estudio con respecto al grupo de pacientes elegibles es baja, suele existir una considerable mortandad experimental, pueden existir contaminaciones entre los grupos del estudio y otros elementos que dificultan la generalización y aplicabilidad de los

resultados. Existen datos relacionados con la evolución individual de los sujetos que no se tienen en cuenta en ese tipo de diseños, tales como: respuestas a corto plazo, agentes terapéuticos previos, rasgos distintivos para subgrupos de pacientes o dificultades de seguimiento de los tratamientos (Fenstein y Horwitz, 1997). Además hay ámbitos donde los ensayos clínicos no son pertinentes o poco posibles de realizar, ya sea por razones éticas o por la imposibilidad de aislar ciertas variables, generando grandes discrepancias de disponibilidad de evidencias según de qué especialidades sanitarias se trate (Newton y West, 1996).

A pesar de estas críticas, incluso para los que señalan limitaciones en la MBE, presenta una perspectiva útil para mejorar la práctica clínica aunque no como paradigma dominante y sustitutivo (Cabrero y Richart, 2000). Vandembroucke (1996) destaca su utilidad para gestores sanitarios e industrias farmacéuticas, pero también considera que la evidencia que se genera no responde totalmente a cuestiones sobre pacientes individuales. Al respecto de esto, Gol-Freixaak y Del Llano (2012) sintetizan un listado con los beneficios y limitaciones de la MBE en la práctica clínica recogidos en la Tabla 1.

El contexto contemporáneo de la atención sanitaria está muy condicionada por ideas como responsabilidad, razón de coste-beneficio, eficacia, y otros elementos que exigen de una prestación sanitaria basada en fundamentos empíricos rigurosos y avalados científicamente (Hicks y Hennesy, 1997). Respondiendo a esta orientación la MBE se ha difundido hoy por casi todos los escenarios clínicos del mundo y ha trascendido hacia los escenarios de la salud pública y la política de la salud en las administraciones (Satterfield et al., 2009). Su objetivo es reducir al mínimo el abismo que separa la investigación científica de la practica clínica real, para lo cual es fundamental que los resultados de evaluar con rigor científico las intervenciones sanitarias, sean accesibles y considerados por el profesional en su toma de decisiones cotidiana (Lenfant, 2003; Rosser, 2008; Wilton y Slim, 2012).

**Tabla 1***Beneficios y Limitaciones de la MBA en la práctica clínica*

<b>Beneficios</b>	<b>Limitaciones</b>
- Marca como objetivo la integración de experiencia individual con la mejor evidencia externa.	- Pérdida relativa del pensamiento fisiopatológico de la práctica clínica.
- Tiene en cuenta la preferencia del paciente informado.	- Minusvalora la experiencia y el instinto del profesional sanitario.
- Fomenta y organiza el auto-aprendizaje.	- Es difícil contar con el tiempo, habilidades y medios para revisar y valorar críticamente la evidencia de forma cuidadosa.
- Transforma las necesidades de información en preguntas de investigación contestables.	- Falta de definición clara y práctica de a qué nivel una intervención se considera efectiva.
- Evalúa la solidez de la evidencia con las respuestas a dichas preguntas. Valorando críticamente la validez y utilidad de la evidencia.	- No se tiene en cuenta el impacto de la aplicación de las políticas de la MBE más allá de la práctica clínica: distribución de recursos, política sanitaria y gestión clínica.
- Pone énfasis en la evaluación de las actuaciones clínicas.	- Riesgo de caer en el fundamentalismo metodológico.
- Resalta el papel activo del profesional como juez crítico y formado de las fuentes que consulte.	- Agravamiento de la disputa entre economistas, gestores y médicos.
- Enfatiza en el plano individual del comportamiento bioético: beneficio y no maleficencia.	- Pasa a un segundo plano el comportamiento bioético de la justicia distributiva y de la eficiencia.

Nota: Adaptado del contenido de Gol-Freixaak y Del Llano (2012).

## **3.2. Conceptos relevantes en la evaluación de resultados en salud**

### **3.2.1. Eficacia, efectividad y eficiencia**

Eficacia y efectividad son dos términos clásicos utilizados en economía, para hacer referencia al umbral de posibilidades de producción de un bien o servicio para una determinada capacidad tecnológica. Siendo la eficacia lo que se puede conseguir si todo funciona bien y representando la efectividad el grado de consecución de ese potencial máximo en función de los datos de la realidad en la que se desenvuelve la organización (Bouza, 2000). Desde la

investigación epidemiológica (Badia y Bigorras, 2000) se adaptó esta definición al mundo sanitario de la siguiente forma:

- *Eficiencia*: es aquella virtud o cualidad de una intervención, que le hace capaz de producir un efecto deseado cuando se aplica en condiciones ideales.
- *Efectividad*: es la capacidad de una intervención de producir el efecto deseado en condiciones reales o habituales.

Aunque teóricamente se asume la diferencia entre los dos términos, en la literatura médica a menudo se obvia o se confunde (Jiménez-Navarro, Cabrera-Bueno, Muñoz-García y Teresa-Galván, 2009). Conde (2002) plantea que la correcta aplicación de ambas definiciones posee una gran utilidad práctica y operativa, ya que configura un marco conceptual que facilita el abordaje sistemático y estandarizado del análisis de resultados de la actividad sanitaria.

Para demostrar la eficacia de una intervención es necesario estudiar sus resultados bajo condiciones de máximo control experimental, lo que vienen a coincidir con lo que la MBE considera generar la mejor evidencia (Iraurgi y Laespada, 2011). Para la MBE esta evidencia idealmente debe provenir del referente que representan los ensayos clínicos controlados y aleatorizados (Tannenbaum, 1995). Ya que es la forma de asegurar la eficacia y seguridad de los nuevos avances sanitarios (González y Condón, 2001). Pero como ya se señaló en el apartado anterior, también se tiene que tener en cuenta que este tipo de diseños tiene una serie de limitaciones entre las que se puede destacar los referentes a la validez externa, es decir, el valor que tienen los datos conseguidos en una muestra homogénea y controlada para extrapolarlos a la práctica clínica que se tiene que enfrentar a una población con una enorme variabilidad de condiciones (Laporte, 2001).

Por otro lado, para poder constatar la efectividad de una intervención es necesario recurrir a los denominados ensayos pragmáticos. Este tipo de diseños tratan de incluir las variaciones que se producen en la práctica habitual, con sus actores y escenarios típicos, y tienen como objeto informar sobre las elecciones entre tratamientos según sus consecuencias e impactos en un contexto real (Wared y Hamel, 2011). La efectividad de la intervención generalmente se mide a través del impacto de los resultados en elementos como la discapacidad, la satisfacción de usuarios, los costes, la calidad de la atención o la calidad de vida percibida (Roland y

Togerson, 1998). Mantienen criterios de inclusión más laxos y utilizan muestras más grandes que los estudios de eficacia con el fin de asegurar su validez externa, estando más interesados en constatar que una intervención funciona en términos importantes para el paciente que en saber por qué funciona (MacPherson, 2004). En estos diseños se problematiza la validez interna, debido a que es posible un menor control de las variables implicadas, por lo que es arriesgado llegar a conclusiones de causalidad (Rubio-Terrés, Sacristán, Badía, Cobo y García-Alonso, 2004). Cuando no es posible un ensayo pragmático en el sentido estrictamente experimental que se le dio desde la epidemiología, lo más frecuente es recurrir a los estudios observacionales de cohorte o de casos y controles manteniendo el objeto y las limitaciones la perspectiva pragmática (Bakke, Carné y García 1994). Este contraste entre estudiar la eficacia o la efectividad para Conde (2000), se resume en cuatro puntos de discrepancia fundamentales:

- *Condiciones del estudio*: ideales en el caso de estudios de eficacia y reales-habituales en el caso de los de efectividad.
- *Excelencia en la ejecución*: garantizada en los estudios de eficacia pero no así en los estudios de efectividad.
- *Seguimiento de la evolución del paciente*: A corto plazo en los estudios de eficacia y con posibilidad de seguimiento a largo plazo.
- *Número de pacientes*: reducido en el caso de los estudios de eficacia y grande en el caso de los estudios de efectividad.

Para Thorpe y colaboradores (2009) los ensayos puramente explicativos (de eficacia) y los ensayos puramente pragmáticos (de efectividad) representan los dos polos de un continuo en investigación clínica. En la Tabla 2 se puede ver como caracterizan estos extremos y enumeran los dominios que el investigador debe conjugar a la hora de diseñar un estudio. Al ubicarlos en un continuo se sugiere que la sensibilidad pragmática y la explicativa pueden tenerse en cuenta en diferente grado dependiendo de la finalidad de la investigación y qué cuestiones se busque responder. Siendo difícil en la realidad calificar a los estudios como totalmente pragmáticos o totalmente explicativos

**Tabla 2***Caracterización Extremos del continuo pragmático-explicativo en investigación clínica*

<b>Dominio</b>	<b>Extremo Pragmático</b>	<b>Extremo Explicativo</b>
<u>Participantes</u>		
Criterios de elección de participantes.	Todos los participantes que tengan la condición de interés son incluidos, independientemente de sus características.	Se aplican rígidos criterios de selección. Especial cuidado con los factores de riesgo previos y la comorbilidad.
Flexibilidad en la intervención.	Flexible. Considerable margen de maniobra.	Inflexible. Estrictas instrucciones a seguir.
Experiencia del profesional que ejecuta la intervención.	Gama completa de profesionales y entornos clínicos, independientemente de sus características.	Profesionales con experiencia que hayan aplicado la intervención anteriormente con éxito.
<u>Seguimiento y Resultados</u>		
Intensidad del seguimiento.	No visitas formales de seguimiento. Uso de bases de datos administrativas.	Extensa recopilación de datos. Independientemente de si han existido eventos de algún tipo o no.
Resultado Primario.	Medida de lo clínicamente significativo para los participantes. No se requiere de pruebas especiales o de un entrenamiento específico para obtener esta medida.	Medida de consecuencia directa e inmediata de la intervención (significativa). Puede requerir de formación especializada o pruebas poco habituales.
<u>Cumplimiento / Adherencia</u>		
Cumplimiento de lo prescrito en la intervención por parte del participante. Adherencia al protocolo de intervención establecido por parte del profesional.	Puede haber una medida discreta del cumplimiento del participante o de la adherencia del profesional, pero no se habrán usado estrategias dedicadas a mantener o mejorar ni la una ni la otra.	Se sigue de cerca el cumplimiento de los participantes y este puede ser un requisito previo para entrar en el estudio. Se supervisa a los profesionales para que se ajusten al protocolo. Se genera un manual de procedimientos.
<u>Análisis</u>		
Análisis del resultado primario	Incluye a todos los pacientes, independientemente del nivel cumplimiento, la elegibilidad u otros factores. No omisión de ningún paciente. Necesidad de constar resultados, teniendo en cuenta todas las condiciones naturales.	Iniciales con todos los participantes, posteriormente otros análisis restringidos a subgrupos de la muestra para estimar el efecto máximo alcanzable. Responder a la pregunta "mecanicista" más restringida y concreta.

Nota: Traducido y adaptado de Thorpe y colaboradores (2009, p. 48).

Conde (2000) señala que ambas perspectivas han de entenderse como complementarias y necesarias al investigar las intervenciones en salud, y nunca como contrapuestas o jerárquicas. Se tiene que exigir a cualquier intervención novedosa que demuestre su eficacia antes de su aplicación generalizada, pero es obligatorio también, que después se estudie su efectividad en condiciones habituales, especialmente a largo plazo.

La eficiencia es otro término habitual en el discurso económico que hace referencia a la relación entre "inputs" (recursos invertidos o costes) y a los "outputs" (beneficios o resultados). En esta relación los outputs figurarán en el denominador y los inputs en el numerador, resultando ser la eficacia mayor cuanto más alto sea el valor de los resultados y más bajo el valor de los recursos invertidos (Coelli, Prasada y Battese, 1998). Aplicado al área clínica, la eficacia relaciona los beneficios de las diferentes intervenciones con los recursos (humanos y materiales) necesarios para llevarlas a cabo (Kernick, 2003). Normalmente se refiere a la utilización de los recursos estrictamente necesarios (coste) para que produzcan la máxima efectividad (Baños, Brotons y Farre, 1998). Desde esta perspectiva los análisis económicos más comunes en gestión sanitaria han sido (Sacristán, Badia y Rovira, 1995):

- Coste-efectividad: tiene por objetivo relacionar el coste monetario de una intervención o programa con uno o varios indicadores de su efectividad expresados en términos físicos y no monetarios como pueden ser años de vida ganados o reducción de mortalidad. Aunque pueden existir dificultades para cuantificar de ciertas variables intangibles de la asistencia sanitaria.
- Coste-beneficio: es aquel que evalúa los costes y los beneficios en las mismas unidades, normalmente económicas; esta comparación nos informa sobre la eficiencia económica de las distintas alternativas. Se determinará si los beneficios de un proyecto superan a los costes para tomar la decisión de efectuarlo o no. De nuevo el problema puede ser cuantificar de manera económica ciertos intangibles, ya que expresar en unidades monetarias las ganancias en materia de salud es realmente complejo.
- Coste-utilidad: tiene en cuenta los beneficios percibidos por el individuo respecto al nivel de calidad de vida que produce un servicio sanitario determinado y lo relaciona con el coste económico de ese procedimiento. Es de especial interés esta consideración, en situaciones en las cuales es tan importante la duración de la vida, como la calidad de la

misma, como son los cuidados paliativos. Uno de sus indicadores más reconocidos en este sentido es el denominado "Año de Vida Ajustado por Calidad" AVAC. Los AVAC se expresan en unidades de años vividos en perfecta salud.

- Coste-minimización: cuando se ha constatado que dos intervenciones producen resultados equivalentes, se comparan exclusivamente por sus costes para priorizar una frente a otra.

Integrar este análisis económico (eficiencia) con la reportado por la evidencia científica (eficacia y efectividad), permite realizar una gestión mucho más competente de los recursos sanitarios (Rubio-Terrés, Sacristán, Badía, Cobo y García-Alonso, 2004).

### **3.2.2. Calidad de la información: Validez, precisión, errores y sesgos**

En el anterior capítulo se hizo referencia a la cuestión de validez y fiabilidad de los instrumentos de medida desde la psicometría y las estrategias que se proponían para asegurar que se cumpliesen estas características en los usos de los instrumentos. En este apartado se volverá a hacer mención a la cuestión de la validez y la fiabilidad pero en esta ocasión de una forma más global refiriéndose a la idoneidad de la inferencia que se puede realizar de los estudios que evalúan variables clínicas y el sentido que se les ha dado a estos constructos desde la epidemiología.

La validez del conocimiento derivado de un estudio dependerá en gran medida de la ausencia de errores y de su capacidad de estimar o predecir el parámetro verdadero en la población objetivo, es decir del grado de concordancia la realidad de dicho parámetro y lo obtenido en el estudio (Hernández-Avila, Garrido, Salazar-Martínez, 2000). Se diferencian dos tipos de validez según que nivel de inferencia queremos aplicar los resultados obtenidos (Porta-Serra, Álvarez-Dardet, Bolúmar, Plasencia y Velilla, 1987):

- Validez interna: grado en el que los resultados de un estudio son correctos para los sujetos que han formado parte del estudio.
- Validez externa: la capacidad de extrapolación de los resultados de un estudio a una población diferente o más extensa que la estudiada.

Por otro lado, la precisión o fiabilidad hace referencia al grado de reproducibilidad de un estudio o de un procedimiento de medida. Se refiere a la dispersión de medidas repetidas,

independientemente de si esas resultan válidas o no. A menos dispersión de las medidas, más fiable será el estudio o la herramientas de medida (Szklo y Nieto, 2003).

Según se trate de un problema de validez o de precisión se pueden reconocer dos fuentes de error (Arignón y Jiménez, 2004):

- Error sistemático: También conocido como "sesgo". Se denomina así porque hace referencia a cualquier efecto tendente a producir resultados que se apartan sistemáticamente del valor real sin importar el tamaño muestral (no debido al azar). Resulta en una estimación incorrecta de una asociación o en el enmascaramiento de posibles asociaciones.
- Error aleatorio: Es el debido a la variabilidad inherente al proceso de muestreo, al proceso de medición y las de las propias variables estudiadas (azar). Muy dependiente del tamaño muestral y afecta a la capacidad de inferencia y generalización de los resultados.

Con respecto al error sistemático, los múltiples sesgos que se corresponden con este tipo de error se han agrupado clásicamente en tres categorías según su origen: Sesgos de selección, Sesgos de información y Factores de confusión (Kleinbaum, Kupper y Morggensten, 1982).

Los sesgos de selección son los que causan una distorsión del efecto medido, derivado de la forma en la que ha seleccionado a los sujetos. Una fuente de sesgo de selección puede ser cualquier factor que influya sobre la posibilidad de que los sujetos sean seleccionados para participar o permanecer en el estudio y que al mismo tiempo esté relacionado con la exposición o con el evento en estudio. Se pueden señalar los siguientes como los más habituales (Hernández-Avila, Garrido y Salazar-Martínez, 2000):

- Selección incorrecta del grupo control. No permitiendo comparabilidad entre grupos por diferir en más factores que los del estudio.
- Elección de una muestra no representativa de la población de interés.
- Pérdidas de sujetos en el seguimiento que difieren con los que no abandonaron en variables relacionadas con el factor en estudio.

- Supervivencia selectiva al seleccionar casos prevalentes (no representativos en todos los casos) en vez de incidentes.
- Sesgo en la detección. No extrapolables a la población debido a que su estado de exposición se dirige más a la detección del evento de interés.
- Participación de voluntarios o autoselección. Los sujetos voluntarios en un estudio son diferentes de los no voluntarios.

Los sesgos de información se corresponden con los errores generados en los procedimientos usados para obtener la información y clasificar a los sujetos del estudio. En este sentido una fuente de sesgo, puede ser cualquier factor que influya de manera diferencial sobre la calidad de las mediciones que se realizan en los grupos de estudio. Debido a esta distorsión el efecto medido se puede exagerar o subestimar realizando una clasificación incorrecta de los sujetos del estudio (Hernández, 2007). Los sesgos de información más habituales son los debidos a (Moreno, Moreno y García, 2012):

- El evaluador, que sesga los resultados debido a sus propias expectativas e hipótesis del estudio o al uso incorrecto de las herramientas de medición produciendo una mala clasificación.
- El evaluado, cuando el sujeto a estudio sesga la información que da al investigador, exagerando su información o recordando incorrectamente.

Los factores de confusión representan un sesgo potencial que se encuentra en la población de referencia producido por las relaciones que mantienen otras variables externas al estudio con la exposición o con los efectos que se investiga. Esto puede dar lugar a un mezcla de efectos y por tanto a una distorsión en la medida del efecto que se quiere estudiar con asociaciones espurias o enmascaramiento de asociaciones (Doménech, 2013a). Para que una variable se considere un factor de confusión debe cumplir los siguientes tres criterios (Rothman y Greenland, 1998):

- Estar asociado con el desenlace o efecto estudiado (en el sentido de factor protector o de riesgo de la variable dependiente) independientemente de su asociación con la condición de exposición.

- Estar asociado con la exposición (variable independiente) en la población de estudio pero sin ser resultado de la misma.
- No debe tratarse de una variable mediadora entre la exposición y el efecto. Es decir no debe ser un eslabón causal intermedio entre la variable independiente y la variable dependiente.

La mayoría de los sesgos de selección e información no pueden ser controlados después de recoger los datos, por lo que es preferible tener en cuenta sus posibles distorsiones a la hora de diseñar el estudio antes de la ejecución (Porta-Serra, Álvarez-Dardet, Bolúmar, Plasencia y Velilla, 1987). Por otro lado, los factores de confusión, si el investigador es consciente de ellos, pueden ser controlados tanto en el momento de diseñar el estudio como posteriormente en el análisis de resultados (mediante estrategias estadísticas), siempre que se haya recogido suficiente información (De Irala, Martínez-Gonzalez y Guilleén, 2001). En la Tabla 3 se enumeran las principales estrategias para controlar los errores sistemáticos, según lo expuesto por Arignón y Jiménez (2013), y Hernández-Avila, Garrido, Salazar-Martínez (2000).

**Tabla 3**

*Estrategias para controlar o minimizar sesgos en investigación clínica*

Tipo de sesgo	Fase de control	Estrategia
Sesgo de selección	Diseño	-Aleatorización en la asignación a grupos. -Criterios de selección y exclusión (Restricción)
Sesgo de Información	Diseño	-Enmascaramiento (simple, doble o triple ciego). - Protocolos de actuación y evaluación preestablecidos.
Confusión	Diseño	- Pareamiento de los grupos - Restricción - Aleatorización
	Análisis	- Estandarización - Estratificación de muestra - Modelos Multivariados

Nota: A partir del contenido de Arignón y Jiménez (2013), y Hernández-Avila, Garrido, Salazar-Martínez (2000).

Por otro lado, el error aleatorio no se puede controlar totalmente pero si calcular su impacto en los resultados del estudio, siendo la mejor manera de reducirlo la estimación correcta del tamaño muestral. A mayor tamaño muestral, menor variabilidad, y por tanto, mayor probabilidad de que los resultados reflejen la realidad de la población a la que la muestra representa (Rothman y Greenland, 1998).

### 3.2.3. Apepciones de la evaluación en entornos sanitarios

Desde el contexto sanitario, Last (2007, p.10) define evaluación como “los esfuerzos dirigidos a determinar de forma sistemática y objetiva la efectividad y el impacto de las actividades realizadas para alcanzar objetivos de salud”. Entendida de una forma tan global la evaluación en salud implica diferentes vías y perspectivas. Por ello, y con el fin de aclarar los distintos enfoques posibles de evaluación en organizaciones sanitarias, Nebot, López, Ariza, Villalbí y García-Altes (2011) recogen las principales tipificaciones en la literatura, tal y como se pueden ver en la Tabla 4.

**Tabla 4**

*Principales concepciones de evaluación en el ámbito sanitario*

<b>Evaluación según nivel</b>	
<i>Táctica de Estructura</i>	Adecuación de recursos a necesidades.
<i>Táctica de Proceso</i>	Adecuación de los servicios y actividades a los objetivos y al protocolo.
<i>Táctica de Resultados</i>	Consecución de los objetivos de la intervención.
<i>Estratégica</i>	Valoración de la pertinencia de los objetivos.
<b>Evaluación según finalidad</b>	
<i>Formativa</i>	Evaluación de pruebas piloto para desarrollo de un nuevo programa.
<i>Sumativa (de impacto)</i>	Valoración eficacia o efectividad de un programa consolidado.
<b>Evaluación según perspectiva</b>	
<i>De desarrollo</i>	Análisis de ejecución de las actividades y de servicios.
<i>De gestión</i>	Orientada a conocer y mejorar programas y costes.
<i>Experimental</i>	Valoración de resultados en condiciones controladas (eficacia y casualidad).
<i>Económica</i>	Estudio de la relación entre costes y los resultados de la intervención.

Nota: Adaptado de Nebot, López, Ariza, Villalbí y García-Altes (2011, p.5).

Aunque a lo largo del presente capítulo se ha hecho referencia de alguna forma u otra a las diferentes acepciones de evaluación, en este apartado nos centraremos en el nivel táctico de la evaluación (ya que es donde se ubica la evaluación desde el enfoque que es de interés para este trabajo). Este nivel táctico se corresponde con lo que Donabedian (1984) identificó como los tres componentes esenciales para evaluar la atención sanitaria: estructura, proceso y resultados o efectos.

La evaluación de la estructura se refiere a la evaluación de los recursos (humanos, físicos y financieros) que se implican para prestar un determinado servicio de salud. Y la manera en cómo estos se coordinan en una estructura organizacional específica. Representando la calidad de estos recursos, su disponibilidad y su accesibilidad como principales indicadores para evaluar la estructura (Cabañas, 1994). El análisis de la eficiencia, anteriormente referenciado, sería la medida de la relación entre estos recursos y los resultados obtenidos (Haddix, Teutsch y Corso, 2002).

La evaluación de procesos se corresponde con la valoración de la ejecución de las actividades que se han realizado para materializar un programa o intervención planificados, a partir de los recursos disponibles en la estructura. Suele comportar evaluar principalmente tres aspectos (Villalbí y Tresserras, 2011):

- Actividad: monitorización de actividades que se realizan a través de indicadores directos para conocer lo que se hace.
- Cobertura: proporción de cobertura de población objetivo a la cual está dirigido el programa o intervención.
- Calidad: constatar si la ejecución de actividades se ajusta a lo planificado y cumplen con lo previsto previamente.
- Satisfacción: satisfacción de los pacientes con el servicio recibido.

Aunque la consecución satisfactoria del proceso es necesaria para que una intervención sea efectiva, no garantiza por sí sola alcanzar los resultados esperados (Nebot, López, Ariza, Villalbí y García-Altes, 2011). La relación entre la estructura y los procesos se estudia a través de la productividad, es decir el número de servicios desempeñados por unidad de recurso existente en un momento dado (Pineault y Daveluy, 1988).

La evaluación de efectos puede hacer referencia a dos enfoques: a la valoración de si se han alcanzado los objetivos previstos, o a la ponderación de las consecuencias sobre población más allá de los resultados inmediatos de la realización de la intervención o del programa. Esta última consideración, conocida como evaluación del impacto pertenece al nivel estratégico de la evaluación. Sin embargo, a nivel táctico, la evaluación de los efectos se identifica con la valoración del cumplimiento de los objetivos de la intervención, es decir con la evaluación de resultados. Estos resultados definidos en términos de estado de salud o comportamientos relativos a la salud por parte de la población, a quién ha ido dirigida la intervención o el programa, englobando no sólo aspectos físicos y fisiológicos sino también psicosociales (Pineault y Daveluy, 1988).

Evaluar los resultados atañe directamente a la consideración de la eficacia y de la efectividad que hemos explicado anteriormente. También se hace referencia a estos conceptos denominándolos eficacia potencial (eficacia) y eficacia real (efectividad). Entendiendo la eficacia real como un producto entre la eficacia potencial, la capacidad de la intervención de alcanzar a la población objetivo y los comportamientos de los beneficiarios y profesionales (Fernández-Ríos y Sánchez, 1997).

#### **3.2.4. Principales diseños evaluativos utilizados en intervenciones en salud**

La investigación evaluativa tiene como objetivo evaluar los resultados de uno o más planes de acción que se han realizado en un contexto determinado. Se diferencia de otros tipos de investigación por su focalización en unos resultados específicos que se orientan hacia la solución de un problema concreto en un entorno preestablecido. En este tipo de investigación se contrastan los resultados después de una intervención, a través de diferentes estrategias de recogida de información y análisis (Hurtado, 2007).

Evaluar los resultados de una intervención en salud, implica cuantificar la magnitud del cambio en los indicadores de resultados preestablecidos, pudiendo descartar explicaciones alternativas a la intervención en que hayan podido influir en el cambio observado (sesgos y errores). Para ello es necesario definir una serie de indicadores adecuados y seleccionar un diseño evaluativo apropiado a las condiciones de lo que se pretenden estudiar (López, Marí-Dell'Olmo, Pérez-Giménez y Nebot, 2011).

La clasificación clásica de diseños de investigación evaluativa, hace referencia a tres grupos de diseños: experimental, cuasi-experimental y no experimental. Esta división se corresponde con las diferentes conjugaciones de elementos que atañen a la validez interna de los estudios, tales como: naturaleza y presencia del grupo control o de comparación, procedimiento y momentos de obtención de medidas resultado, número de medidas de resultado y tipo asignación a los diferentes grupos (Windsor, Branowski, Clark y Cutter, 1994).

Los diseños experimentales son los que ejercen un mayor grado de control sobre las variables implicadas y permiten atribuir con mayor seguridad el efecto observado a la intervención (validez interna). La variable independiente estará bajo el control del experimentador, existirá un grupo control y la asignación a los grupos será aleatoria con el fin de generar grupos equivalentes en la distribución de variables de confusión. A pesar de su idoneidad metodológica, este tipo de diseños conllevan dos importantes inconvenientes prácticos que hacen que sean difíciles de aplicar en la realidad de las organizaciones sanitarias. En primer lugar, el conflicto ético que se presenta ante la aplicación diferencial (uno de los grupos no la recibe) de la práctica si esta ya ha demostrado su eficacia. Y en segundo lugar, el alto coste en tiempo y recursos que supone realizar este tipo de estudios (Pineault y Daveluy, 1988). Campbell y Stanley (1995) señalan la siguiente clasificación dentro de esta categoría:

- Diseño de grupo control Pre-test Post-test: existen medidas antes y después de la intervención en los grupos, si se comparan la magnitud del cambio en los grupos se puede observar el efecto diferencial de la intervención a estudio. Como ya hemos expuesto anteriormente se trata del prototipo de diseño experimental epidemiología.
- Diseño de cuatro grupos de Solomon: Se centra en controlar la posible distorsión que puede provocar medir varias veces ciertas variables en los mismos sujetos. Para ello utiliza 4 grupos equivalentes que son una combinación de dos diseños de grupos aleatorios (uno con medidas pre-test y otro sin ellas). Es poco habitual por su gran coste.
- Diseño de grupo control con solo post-test: No se recogen medidas previas al tratamiento. Se asume que los grupos asignados aleatoriamente van a ser perfectamente equivalentes y que la diferencia de las medidas después del tratamiento representan el efecto de la intervención. Pero esta asunción supone un riesgo a la hora de atribuir el efecto si en la realidad la variable de interés no es equivalente inicialmente en ambos grupos.

Los diseños cuasiexperimentales, también conocidos como diseños de control parcial tienen características similares a los diseños experimentales, pero se diferencian de estos por la ausencia de asignación aleatoria de los sujetos a los diferentes grupos. Aunque es posible cierto control sobre variables distorsionadores, al no existir aleatorización es imposible controlar los factores de confusión que el investigador desconoce. Por ello, es necesario buscar la mayor equivalencia posible, entre los grupos que se comparen, respecto a todo lo que puede afectar a la variable resultado (León y Montero, 2003). Dentro de esta categoría se destacan los siguientes diseños para la investigación evaluativa (Iraurgi, 2000):

- *Diseños pre-post*: Uno de los diseños más habituales en la investigación de intervenciones en salud. Implica por lo menos una medida antes y otra después (en variable resultado), en ambos grupos para medir el efecto de la intervención de forma parecida a los diseños experimentales pre-post. La dificultad radica en generar un grupo de comparación realmente equivalente, lo que se convierte en un problema importante a tener en cuenta cuando se atribuya la causalidad del efecto observado y un desafío a superar en la fase de diseño. En el contexto clínico será difícil encontrar grupos equivalentes. Si no se cumple este supuesto se denominará el diseño como “de grupo no equivalente”. Los grupos se configurarán adecuándose a la disponibilidad de la realidad clínica. Aunque no se pueda conseguir la equivalencia total es posible buscar una mayor homogenización mediante diferentes técnicas (como bloqueo o emparejamiento). Además si se conoce previamente las variables intervinientes sobre la variable resultado se pueden ajustar posteriormente mediante técnicas de control estadístico. Pese a su menor validez interna será más idóneo metodológicamente, que no contar con ningún tipo de grupo de comparación.
- *Discontinuidad en la regresión*: Se caracteriza por realizar la asignación a los grupos basándose en una medida previa de lo que será la variable resultado (u otra variable relacionada), estableciendo un punto de corte a partir del cual se ubican los sujetos que van a ser intervenidos y dejando al otro lado de ese punto lo que sujetos del grupo de comparación. Si la intervención es efectiva variará la variable resultado en el grupo intervenido y se mantendrá sin cambio (o tendrá escaso cambio) en el grupo de comparación. De esta forma se prioriza la intervención en los sujetos que más la necesitan. Por lo que además de las medidas pre y post, necesitamos otra medida previa

de la variable de interés que necesariamente deberá estar expresada en una escala de medida continua, ya que la lógica del diseño se basa en modelos lineales.

- *Serie temporal ininterrumpida o múltiple*: Supone que existe una continuidad de medidas de la variable resultado que van más allá de lo inmediatamente anterior o lo inmediatamente posterior a la intervención. Al contar con estas múltiples medidas en ambos grupos (intervenido y control) se puede comparar los dos grupos a lo largo del tiempo. Permitiendo así controlar sesgos históricos y observar tendencias temporales independientes de la intervención, aumentando de esta forma su validez interna. El efecto de esta forma se constatará en el cambio en la serie de valores del grupo intervenido frente al no cambio del grupo de comparación. La complicación es precisamente la obtención de esta situación de medida continua prolongada en el tiempo sobre la variable de interés, y que además exista en los dos grupos equivalentes.

Los diseños no experimentales o también conocidos como pre-experimentales son aquellos diseños donde no existe una asignación de sujetos a un grupo control. Esto problematiza la validez interna del estudio y por tanto la posibilidad de establecer una relación causal entre la intervención y los resultados. Su menor coste y su simplicidad pueden hacerlos adecuados para realizar análisis más detallados de los procesos que tienen lugar en el desarrollo de la intervención (evaluación de proceso o formativa), pero será inadecuada para comprobar la eficacia de la intervención (evaluación sumativa). Cuantas más mediciones, prolongadas en el tiempo, se recojan antes y después de la intervención, más robusto se podrá considerar el diseño ya que será menos vulnerable a algunas amenazas a su validez, tal y como se ha explicado en el diseño de series temporales (Pineault y Daveluy, 1988; López, Marí-Dell'Olmo, Pérez-Giménez y Nebot, 2011).

A esta tipología clásica de diseños de evaluación de resultados (experimental, cuasi experimental y pre experimental), se le puede añadir los aportes de los diseños observacionales. En este tipo de diseños el investigador no ha tenido ningún control sobre la variable independiente, que en el caso de la investigación evaluativa es la asignación a la intervención. Debido a sus limitaciones metodológicas y su ajuste a la realidad clínica preestablecida, su utilidad se ubica en el área de la efectividad de las intervenciones y no en su eficacia. Desde la epidemiología se señalan útiles para este fin evaluativo dos tipos de diseños observacionales:

estudios de cohortes y estudios de casos y controles (Kleinbaum, Kupper y Morggenstern 1982). Su naturaleza se corresponde con el significado que se les da en la investigación en ciencias sociales a los estudios “ex post facto” prospectivos y retrospectivos respectivamente (Montero y León, 2007):

- *Sentido prospectivo*: parten del conocimiento de la ejecución de la intervención en los sujetos que se están analizando, para recoger más adelante los resultados de esa intervención. Análisis en sentido de variable independiente a dependiente.
- *Sentido retrospectivo*: parten del conocimiento de los resultados de las intervenciones en los sujetos que se está analizando, para poder ver cómo se les ha intervenido para llegar a esos resultados. Análisis en sentido de variable dependiente a independiente.

Los diseños experimentales se presentan como el modelo idóneo metodológicamente para evaluar resultados. A pesar de ello, las dificultades y limitaciones que conllevan en la realidad clínica (ya mencionados en este apartado y en los anteriores), han generado la reivindicación de los denominados diseños evaluativos “débiles” (los diseños que no son estrictamente experimentales), como herramientas legítimas para analizar la realidad compleja de los efectos de los programas e intervenciones en salud. Por supuesto, se deben buscar todas las estrategias posibles para controlar las variables que sesguen o confundan la evaluación, además de ser conscientes de las limitaciones metodológicas que se tengan a la hora de atribuir la causalidad de los efectos encontrados (Nebot y Victoria, 2010). Desde del área de la salud pública se propuso un enfoque pragmático para la generación de evidencia con los diseños “débiles” que consiste en una valoración progresiva en dos etapas (Nebot, 2007):

1º *Adecuación del efecto*: comprobar que se han producido cambios deseables o favorables en las variables resultado en la población después de la intervención. No requiere de grupo control pero si de una determinada magnitud del efecto para considerar los resultados adecuados. No se demuestra causalidad pero se pueden empezar a tomar decisiones.

2º *Plausibilidad del efecto*: Una vez demostrada su adecuación, se tiene que constatar que ese efecto se ha debido efectivamente a la intervención. Con este fin, se buscará acumular diversas evidencias que sirvan para descartar explicaciones alternativas a la causalidad de

la intervención y los resultados, no solo en los resultados en sí mismos sino en también el análisis de los procesos.

### **3.3. Investigación de Resultados en Salud (IRS)**

La Investigación de Resultados en Salud (Outcomes Research en su acepción original anglosajona) es un término desarrollado en el ámbito médico para hacer referencia a la investigación que concierne a los resultados de la atención sanitaria en su contexto natural. Su consideración de los resultados va más allá de los indicadores fisiológicos o bioquímicos, y da un especial protagonismo a los elementos clínicamente relevantes tanto para los pacientes como para la sociedad en general, tales como calidad de vida o nivel funcional (Epstein y Sherwood, 1996).

La IRS examina la eficacia del tratamiento en los tratamientos en entornos clínicos reales de poblaciones heterogéneas y tiene que ver con una amplia gama de cuestiones, que afectan a todos los aspectos de la prestación de atención sanitaria: desde el encuentro clínico entre el paciente y el profesional hasta cuestiones organizacionales como la financiación o la regulación del sistema de salud. Y aunque los ensayos experimentales controlados y aleatorizados se reconocen importantes para la evaluación de algunos resultados, gran parte de la IRS se basa en diseños de investigación no estrictamente experimentales (Badia y Bigorras, 2000).

#### **3.3.1. Desarrollo de la IRS y aclaraciones conceptuales.**

Aunque los orígenes de la IRS se ubican en las reflexiones de Donabedian (1966) acerca de evaluar la calidad de la asistencia sanitaria (tal y como se ha expuesto en al inicio de este capítulo), la Agency for Health Care Policy and Research AHCPR de Estados Unidos, fue la responsable de los primeros grandes estudios de esta línea de investigación centrada en la efectividad. Generaron los grupos de investigación denominados Patient Outcomes Research Teams (PORT). Cada PORT se centró en una patología concreta con el fin de constatar los resultados en la clínica real de las diferentes intervenciones en ese área. Se atendió no solo a si la intervención funcionaba, sino también a sus costes, sus diferentes niveles de resultados, a su funcionamiento diferencial (según subgrupos de pacientes o de clínicos) y a los factores que explicasen variabilidad de resultados (Freud et al., 1999). Poco después desde esta misma

perspectiva se orientó en el Reino Unido, el National Institute for Clinical Excellence NICE para poder elaborar guías y recomendaciones clínicas (Towse, 1999). Actualmente, este enfoque de investigación se ha reconocido fundamental para dar respuesta a los grandes desafíos del entorno de la atención sanitaria y a las inquietudes de sus principales actores (clínicos, pacientes y gestores). Disponiendo este tipo de estudios un gran respaldo tanto desde las asociaciones profesionales como de las diferentes instituciones gubernamentales (Krumholz, 2009).

A nivel conceptual es necesario resaltar que el enfoque de la IRS no se engloba en la denominada Investigación Clínica, sino que resulta ser complementaria a esta perspectiva clásica previa. La IRS investiga en el estado funcional del paciente después del tratamiento de una forma global, incluyendo indicadores tales como la calidad de vida, mientras que la Investigación Clínica se centra en los elementos estrictamente referidos la eficacia inmediata del tratamiento, como la reducción o eliminación de síntomas (y patologías) y a la seguridad de su ejecución (Coyle, 2000). Inicialmente la característica clave para distinguir estos dos enfoques fue el énfasis sobre la efectividad de la IRS, en contraste con el objetivo de comprobación de la eficacia de la Investigación Clínica. Sin embargo, la aceptación de la perspectiva IRS en la investigación ha generado que el término abarque estudios cuyo objetivo vaya más allá de la sola constatación de la efectividad clínica de una intervención, cubriendo una amplia gama de preguntas de estudio, metodologías y resultados (Lee, Earle y Weeks, 2000).

La American Medical Association (1997) señaló la IRS como una actividad interdisciplinar, en la que participan investigadores de diferentes especialidades (médicos, epidemiólogos, economistas, sociólogos, psicólogos, estadísticos y expertos en ética), que utiliza métodos de investigación experimentales u observacionales para medir los resultados (de diferentes naturalezas) de las intervenciones sanitarias en condiciones de práctica habitual. Por su parte, Clancy y Eisenberg (1997) destacaron que la IRS estudia los resultados finales de los servicios de salud teniendo en cuenta los valores, preferencias y experiencias de los pacientes y los profesionales; con la intención de aportar evidencias científicas relativas a las decisiones tomadas por todos los participantes en el cuidado de la salud. Rapier (1996) sintetizó los principales elementos que caracterizan los estudios de IRS:

- Basados en la práctica clínica habitual (con muchos sujetos diferentes entre sí, con seguimiento prolongado, atendiendo a la consecución de resultados de funcionalidad).
- Análisis de la efectividad de las intervenciones (funcionamiento real de las diferentes intervenciones en condiciones de clínica habitual).
- Especial énfasis en los beneficios obtenidos por el paciente (atender a las variables que considera importantes el paciente como su propia satisfacción con los resultados).
- Utilización de métodos de investigación establecidos (aunque se enfatiza la perspectiva pragmática, se tienen en cuenta los desarrollos e inquietudes metodológicas de la investigación biomédica y el cumplimiento del rigor científico).
- Se puede contemplar tanto el enfoque individual (paciente) como el agregado (grupo) con respecto a los resultados de los estudios.

Para Badia y Lizán (2002; 2007) los resultados de los estudios IRS representan evidencias científicas relevantes, tal y como se reclama desde el movimiento de la MBE. Por lo que son posturas compatibles y en esencia con el mismo objetivo de mejorar la práctica clínica con la mejor evidencia disponible. Sin embargo, el énfasis de la MBE en categorizar los niveles de evidencia a partir del "patrón oro" metodológico de los Ensayos Clínicos Controlados choca con el objetivo de la IRS de estudiar los resultados de la práctica clínica habitual. Por lo que se reivindica la utilidad de los diseños más pragmáticos, para superar las limitaciones de validez externa de los Ensayos Controlados. No se propone una contraposición sino una complementariedad en forma de secuencia lógica, que sigue un esquema similar al ya expuesto en este mismo capítulo de demostrar la eficacia y seguidamente la efectividad.

De esta forma la MBE (y sus estudios metodológicamente robustos), generan información fiable sobre cuál es la intervención más eficaz en cada patología, dando un primer referente para la toma de decisiones clínicas ante las nuevas intervenciones o tecnologías sanitarias. Para que posteriormente los estudios de IRS evalúen las consecuencias de esas intervenciones en la clínica real y generen la información suficiente para que el clínico pueda tomar decisiones sobre la idoneidad de una intervención, teniendo en cuenta el contexto concreto de la intervención y a las condiciones particulares de cada paciente individual (Iraurgi y Laespada, 2011).

### 3.3.2. Tipos de Resultados en IRS

La consideración de los resultados de la atención sanitaria ha ido cambiando a lo largo del tiempo, teniendo cada vez más en cuenta mayor número de indicadores y de más diversa naturaleza. Esta evolución ha sido una lógica consecuencia de la modificación de la propia concepción del proceso salud- enfermedad, determinado por estilo y condiciones de vida del cada momento histórico. Algo evidente, por ejemplo, en el paso de la predominancia de la condición de enfermedad aguda en el pasado de occidente, al gran protagonismo de la enfermedad crónica y los procesos degenerativos del momento actual. Por ello, no es de extrañar que en la literatura se encuentren diversas tipologías de resultados que se focalicen en diferentes grupos de indicadores (Schlaepfer-Pedrazzini e Infante-Castañeda, 1990).

Desde la IRS se sintetizan los resultados de salud en tres grupos de indicadores: clínicos y/o epidemiológicos, de interés para el paciente y económicos. En la Figura 1 se puede observar como interrelacionan los distintos tipos de resultados para generar diferentes focos de estudio. Los primeros hacen referencia a las variables clínicas clásicamente contempladas (síntomas, signos, eventos clínicos, mortalidad, medidas fisiológicas, prevalencia, incidencias...). Los segundos son las cuestiones valoradas e informadas por el propio paciente que van más allá de la clínica inmediata (calidad de vida, satisfacción con el tratamiento, preferencias de tratamiento, adherencia al tratamiento...). Y los resultados de tipo económico que son los costes directos o indirectos en términos de recursos utilizados o recursos dejados de producir. Estos resultados se pueden ver determinados por diferentes factores y el objetivo de su análisis es aislar la relación entre el resultado y la intervención. Para ello es necesario recoger también información sobre los factores que puedan influir en los resultados sin ser características propias de la intervención, y así poder controlarlos (lo que se denomina ajuste por riesgo). Por lo que es deseable contemplar las características sociodemográficas, psicosociales y clínicas de los pacientes antes de la intervención (Epstein y Sherwood 1996; Badia y Bigorras, 2000).



**Figura 1.** Relación de los tipos de resultados en salud según modelo IRS

Adaptado de Badia y Bigorras (2002, p.27)

A lo largo de este capítulo ya se han hecho mención a las categorías de resultados clínicos y económicos, y a sus implicaciones teóricas y prácticas. Aunque no resta importancia a estas áreas, la IRS ha pretendido subrayar la gran relevancia de recoger además, los resultados de salud desde la perspectiva del paciente, y en las dimensiones que le son importantes para este. En los últimos años ha acontecido un cambio de paradigma en la relación profesional y paciente, en la cual se ha reconocido la necesidad de escuchar e implicar al paciente en la toma de decisiones que afectan a su salud y bienestar. Su impacto sobre la sociedad, ha trascendido de las consideraciones teóricas acerca de la calidad asistencial y se ha manifestado directamente sobre la legislación de los diferentes países con el reconocimiento de los derechos en relación a la autonomía del paciente (Nebot, Baqués, Crespo, Valverde, Canela-Soler, 2005; Soto, 2006; 2007; Pardo y Fuster, 2007).

En el mundo anglosajón este área de resultados se ha denominado Patient Reported Outcomes (PRO), y se ha incidido sobre la necesidad de afrontar su estudio de una forma precisa y con herramientas que aseguren que no se pierde el rigor científico en su análisis. En este sentido se han destacado principalmente dos focos de estudio: la Calidad de Vida Relacionada

con la Salud CVRS y el nivel de satisfacción de los usuarios con el tratamiento (Wright, 2000; Revicki, 2002; Doward y McKenna, 2004).

Las medidas de CVRS pretenden reflejar el impacto de la enfermedad y su tratamiento sobre todos los aspectos de la vida del paciente. Mientras que las escalas de capacidad funcional tradicionales, excluyen el criterio subjetivo del paciente (criterios objetivos ponderados por un evaluador externo), las de CVRS intentan medir de forma comprensiva los resultados importantes para los pacientes desde su punto de vista. Se ha constatado que habitualmente difieren en su valoración con lo observado por el profesional. Para la evaluación de resultados es de bastante interés la obtención de un valor cuantitativo que represente el estado de salud a partir de la medida de expresiones de tipo subjetivo. Estas medidas tienen enfoque multidimensional que va más allá de la constatación de funcionalidad física del paciente, que atañen tanto al funcionamiento: social (interacción a diferentes niveles del entorno), psicológico (indicadores cognitivos y afectivos de bienestar) y físico (sensaciones somáticas, alteraciones y limitaciones físicas), así como a percepciones e impresiones globales sobre el propio estado de salud. Los instrumentos de CVRS suelen tomar la forma de cuestionarios estructurados con ítems de opciones de respuesta cerrada (con determinadas alternativas o con opciones ordinales en escala Likert). Su naturaleza subjetiva no implica que se debe renunciar al rigor de la medida, así que estos instrumentos deben afrontar los requisitos de validez, fiabilidad y sensibilidad a los cambios, tal y como lo hacen los instrumentos psicométricos (Peiró 2002; Valderas, Ferrer y Alonso 2005; García, Alfaro y Moreno, 2009).

La consecución de la satisfacción del paciente con el tratamiento, es un elemento deseable por todos los agentes implicados en la prestación sanitaria, sin embargo, alcanzar un consenso acerca de las dimensiones a tener en cuenta a la hora de medirla y cuáles son los aspectos concretos que la definen es algo más complicado (Pascoe, 1983). El nivel de satisfacción con el tratamiento se puede definir como la valoración por parte del paciente acerca del proceso de administración del tratamiento y sus resultados (Revicki, 2004). Se entiende la satisfacción como un fenómeno multidimensional y se destacan ciertos aspectos de mayor peso empírico, tales como: comunicación, actitudes, clima de confianza y percepción del paciente (Bowers, Swan y Kohler 1994). A pesar de que los cuestionarios son una opción muy interesante para el estudio y medida de la satisfacción, se han generado muy pocos que sean metodológicamente correctos y

aún menos que dispongan de una validación en nuestra cultura e idioma. Por otro lado, se han observado indicios sobre el considerable impacto que tiene la satisfacción con el tratamiento sobre su efectividad. Por lo que se reconoce la necesidad actual de seguir investigando su relación con los otros resultados de salud y desarrollando instrumentos sensibles, fiables y fáciles de aplicar en la práctica habitual (Soto, 2007; Villar, Lizan, Soto y Peiró 2009; Rofail, et al. 2009).

## Capítulo 4

### Evaluación de Resultados en Intervenciones psicológicas

---

En el presente capítulo se destacan las particularidades de la evaluación de resultados de las intervenciones psicológicas. Muchos de los conceptos y desarrollos de la evaluación de resultados en salud son aplicables de forma análoga al caso específico de las intervenciones psicológicas, por lo que no se volverán a definir. Pero sí se matizarán las implicaciones concretas que genera esta en la ciencia psicológica. En primer lugar, se explica la evolución histórica de su consideración y conceptualización desde sus orígenes como campo de estudio específico. A continuación, se detalla la perspectiva de “Tratamientos psicológicos basados en la Evidencia” centrada en los estudios de eficacia diferencial de técnicas psicoterapéuticas, de gran relevancia en los años noventa. Finalmente, se explicita la situación contemporánea, con el protagonismo de la necesidad de evaluación de resultados en la práctica diaria y de constatar su relevancia clínica, así como las dificultades y retos que esto implica.

#### 4.1. Desarrollo histórico de la evaluación de resultados en psicología

Para Bakham (2002) la evolución histórica de este campo de estudio se puede sintetizar en dos procesos: el paso de justificar la utilidad de las intervenciones psicológicas en conjunto a especificar qué estrategias son mejores para cada circunstancia, y por otro lado, la evolución de analizar la eficacia y la eficiencia de las intervenciones a estudiar la efectividad y la significación clínica de las mismas. Por su parte, Orlinsky y Russell (1994) especifican cuatro etapas en el desarrollo de esta área:

- Primera: el establecimiento de un rol de científico investigador. Se empiezan a registrar y contabilizar los resultados de las intervenciones.
- Segunda: la búsqueda del rigor científico. Se buscaron métodos objetivos para medir los resultados e inicio del uso de ensayos experimentales para demostrar la eficacia de las intervenciones.

- Tercera: de la expansión, diferenciación y organización. Aumentó la sofisticación conceptual y metodológica, así como la innovación en la investigación con el uso de técnicas de revisión meta-analíticas y crecimiento de organizaciones dedicadas a la investigación de resultados.
- Cuarta: de consolidación, insatisfacción y reformulación. Se continúa con la sofisticación metodológica pero se generan dudas sobre la idoneidad acerca de los medios experimentales utilizados hasta ahora y se generan alternativas nuevas.

El primer estudio estadístico, en relación a la evaluación de resultados de las intervenciones, fue el realizado por Fenichel (1930) con los resultados de los pacientes tratados en el Instituto psicoanalítico de Berlín durante la década de 1920 a 1930. Se clasificó a los pacientes al término de su tratamiento (en no curado, mejorado, muy mejorado y curado) según su estado final. Esto representó un temprano afrontamiento al asunto de evaluar los resultados en psicoterapia y un claro ejemplo de la necesidad de alcanzar un consenso a la hora de definir la condición de mejoría real tras el tratamiento. En este estudio la asignación a categorías se realizó según el criterio clínico de los profesionales (subjetivo) y un porcentaje no muy alto de pacientes se consideró “curado” problematizando la eficacia de la psicoterapia. Además esa metodología no permitía diferenciar sujetos que mejoraban o empeoraban espontáneamente independientemente de la terapia (Gabbard, Beck y Holmes 2007; Rivera y Bernal 2008).

Aunque el principal revulsivo para el afrontamiento de este tema desde la psicología, vino dado por un polémico artículo de Eysenck (1952). El autor analizó 24 estudios de resultados de psicoterapia con pacientes neuróticos (incluido citado estudio de Fenichel) que aglutinaban más de 8000 casos. Posteriormente comparó las tasas de pacientes registrados como “curados” después de la psicoterapia con los datos de los estudios de Landis (1938) y Denker (1937) de recuperación espontánea de pacientes neuróticos sin haber recibido psicoterapia. El porcentaje de “curaciones” en ambos casos fue el mismo, lo que le llevó a concluir que con la evidencia disponible en ese momento, la psicoterapia no era más eficaz que el simple paso del tiempo. Después de esta provocativa declaración, matizó en sus conclusiones que los estudios analizados presentaban serias carencias evaluativas. Esto limitaba el alcance de las conclusiones que se podían sacar de esos datos, animando a la comunidad científica a afrontar el diseño y realización de estudios rigurosos que integrasen medidas de antes de la terapia, a modo de línea base de los

cambios que se puedan producir. La declaración de la no eficacia de la psicoterapia generó un intenso debate científico en las siguientes décadas, que a su vez provocó un incremento sustancial de trabajos y estudios acerca de la evaluación de resultados (Auping, 2000; Hilsenroth 2013).

En sus respuestas críticas al artículo de Eysenck, Strupp (1963) señaló la ambigüedad que existía en aquel entonces respecto a la definición de la variable independiente, es decir, en qué consistía un tipo particular de psicoterapia. Para este autor, sin una definición explícita de los principios y técnicas que la configuran, difícilmente se podían generar estudios de resultados satisfactorios. El primer manual que se centró en definir exhaustivamente una psicoterapia, fue el dedicado a la Terapia cognitiva de la Depresión (Beck, Rush, Shaw y Emery, ,1979). Posteriormente empezaron a proliferar los manuales de tratamiento, lo que significó no solo una respuesta a la demanda de Strupp sino un avance significativo en el campo de las intervenciones psicológicas (Luborsky y De Rubeis 1984).

Otro de los elemento que Strupp (1964) destacó como fundamentales para el estudio de la eficacia de las intervenciones, fue la discusión sobre la propia naturaleza de los resultados. Para este autor, a la hora de medir los beneficios de la intervención, era fundamental ir más allá del criterio de clínico subjetivo sobre la “mejoría” del paciente e incluir indicadores más amplios que la mera supresión de síntomas. Para Wampold (2013) el debate sobre la propia naturaleza de los resultados a medir, ha sido un elemento crítico en el afrontamiento de la determinación de los efectos de la psicoterapia, que sigue planteando divergencias entre las diferentes líneas de investigación. Aunque en los orígenes de varios modelos de psicoterapia como el humanista o el conductista existe un rechazo explícito al modelo médico de diagnóstico, se terminó realizando un acercamiento, desde las asociaciones profesionales, a la nosología psiquiátrica con el fin de competir con la psiquiatría biológica (Task Force on Promotion and Dissemination of Psychological Procedures, 1995). De esta forma el problema del desacuerdo en la definición de los resultados se trató de atajar poniendo énfasis en la terapia, para resolver “trastornos” definidos en clasificaciones psiquiátricas, tales como la distintas ediciones del “Diagnostic and Statistical Manual of Mental Disorders” DSM, solución que no todos los académicos encuentran totalmente satisfactoria por la dudosa validez de estas clasificaciones y el obviar otro tipo de resultados (Wampold 2001).

Por su parte Bergin (1963; 1967) rebatió al artículo de Eysenck subrayando otro elemento relevante a la hora de determinar la eficacia de la psicoterapia: el propio diseño de los estudios. Este autor generó explicaciones alternativas con los mismos datos expuestos por Eysenck, y concluyó que: la variabilidad de criterios, la diferencias de fiabilidad en la medidas, los potenciales sesgos evaluativos (sin medidas independientes u objetivas) y las discrepancias en duración, calidad y cantidad de tratamiento no permitían combinar los resultados de estas observaciones de la forma que se hizo y por supuesto, no podían representar evidencia de la no eficacia real de la psicoterapia. No es de extrañar que propusiese abandonar las estrategias retrospectivas, sujetas al sesgo de ese de tipo de estudios y que se señalase como estándar para evaluar los resultados terapéuticos al ensayo clínico controlado. Tomó en ese momento gran protagonismo la generación de estudios controlados que comparasen un grupo tratado con un grupo control sin psicoterapia de ningún tipo, con resultados favorables hacia la eficacia de la psicoterapia (Bergin y Lambert, 1978).

Posteriormente, Smith y Glass (1977) generaron un réplica contundente a la afirmación de equivalencia entre psicoterapia y no tratamiento al utilizar una nueva estrategia: el uso de la metodología de metaanálisis para estimar el tamaño del efecto de la intervención psicológica a partir de los numerosos ensayos controlados que se habían generado en esa época. Se concluyó que en promedio los resultados de las personas tratadas eran 75% mejores que los de las personas no tratadas, y se afirmó que no existían diferencias importantes entre los diferentes tipos de psicoterapia estudiados (Smith, Glass y Miller, 1980). En revisiones posteriores (basadas en análisis metaanalíticos), se han señalado tamaños del efecto superiores a 0,65 en eficacia de la psicoterapia, de forma general, respecto al no tratamiento (Matt y Navarro, 1997; García, Bados y Saldaña 1998; Chambless y Ollendick, 2001).

Gracias a este hito, se reconoció la gran utilidad de este tipo de técnicas de análisis para establecer generalizaciones confiables y su potencial para ofrecer pistas para explicar los resultados encontrados (y así superar las limitaciones del anterior paradigma más cualitativo-subjetivo), pero también se advirtió de que sus estimaciones no eran impermeables a sesgos y que era necesario asegurar la imparcialidad de la evaluación (Cook y Lebiton, 1980). De hecho, se señaló que los resultados reflejados por Smith y Glass, podían haber sobreestimado su verdadero tamaño del efecto al basarse en medidas promedio y no el porcentaje de sujetos

mejorados (Eysenck, 1978; Rachman y Wilson 1980). Por ello, no es de extrañar que aunque este tipo de revisiones metaanalíticas sirvieron para generar un relativo consenso acerca de la eficacia de la psicoterapia en términos generales, esta estrategia generó conclusiones contradictorias a la hora de establecer la eficacia diferencial entre las diferentes orientaciones y técnicas psicoterapéuticas (Shapiro y Shapiro, 1982; Matt y Navarro, 1997; Bados, García y Fuste, 2002).

La supuesta igualdad entre técnicas psicoterapéuticas, retratada por Luborsky, Singer y Luborsky (1975) como “el veredicto del pájaro Dodo” (en referencia a la carrera de Alicia en el País de las maravillas donde todos ganan), provocó dos líneas de investigación contrapuestas en su planteamiento base, pero que de alguna manera compartían un mismo objetivo: generar una investigación de resultados con un mayor nivel de especificación. Por un lado, los autores que dieron menos peso a las técnicas utilizadas y más a los factores comunes a toda psicoterapia tales como la relación terapéutica, y por otro los que se centraron en enfatizar los resultados de cada tipo de intervención para cada trastorno específico en vez de la perspectiva genérica anterior (Mitchell, Truax, Bozarth y Krauft 1973; Montgomery, 2007).

Las revisiones de Lambert (1980; 1992) se centraron en la mencionada línea de estudio de factores comunes. Para este autor solo un 15% de la varianza de los resultados se debía a la técnica psicoterapéutica utilizada. Señaló el gran peso sobre los resultados de las variables relacionadas con el paciente (su grado de perturbación psicológica, la disposición al tratamiento, las características sociodemográficas, la motivación, sus expectativas de cambio etc...). Y de entre los factores relacionados con el profesional, destacó por sus efectos en los resultados, el tipo de relación terapéutica que generaba (Lambert y Asay, 1984). Las investigaciones en este sentido han ocasionado que actualmente el estudio de los resultados en su vertiente de efectividad clínica, se haya fijado especialmente en lo reportado directamente por el paciente en los diferentes momentos del proceso, el análisis individualizado y en la importancia del uso retroalimentación de información con respecto al cambio terapéutico (Duncan y Miller 2000; Lambert; 2013). Más adelante en este mismo capítulo se profundizará en estos elementos.

Desde otra perspectiva, la investigación en el modelo de intervención conductista se orientó inicialmente en buscar la especificidad de la técnica psicoterapéutica para cada desorden de la conducta, a través de diseños de caso único para concretar los efectos en cada tipo de cliente

(Paul, 1967; Kazdin, 1978). Posteriormente se asumió el diseño de ensayo clínico controlado para estudiar la eficacia de las intervenciones, con el fin de controlar los posibles sesgos de otros tipos de diseños y aumentar la validez interna de los resultados encontrados. Se siguió manteniendo el objetivo de evaluar los tratamientos en trastornos concretos, y se señaló necesario utilizar muestras clínicas claramente delimitadas (Hollon 1996; Rivera y Bernal 2008). En este sentido Agras y Berkowitz (1980) propusieron asumir un modelo de investigación de intervenciones similar el utilizado en la validación de tratamientos farmacológicos. Este suponía primero identificar y examinar las intervenciones disponibles para cada trastorno para después someterlas a estudios de alto control metodológico (asegurar validez interna), constatar sus efectos y difundirlo para la práctica real, y así finalmente poder verificar su efectividad y aplicabilidad a entornos no experimentales.

Este planteamiento de investigación tuvo un gran apoyo institucional en las ciencias de la salud durante la década de los 90, como vimos en el capítulo anterior. En EEUU, la AHCPR también incluyó entre sus objetivos (además de las patologías físicas) determinar qué tratamientos eran eficaces para los trastornos mentales. La American Psychological Association (APA) creó un grupo de trabajo para dar respuesta a estas demandas de eficacia diferencial denominado Task Force on Promotion and Dissemination of Psychological Procedures. El informe resultante y sus sucesivas revisiones subrayaron la importancia de validar empíricamente los tratamientos psicológicos y los criterios señalados se ajustaron al paradigma de la demostración de eficacia mediante ensayos clínicos. Esto significó el precedente directo del movimiento que se ha denominado Práctica Psicológica Basada en la Evidencia. (Task Force on Promotion and Dissemination of Psychological Procedures TFPDP, 1995; Chambles, et al., 1996; Chambless, et al., 1998; Becoña, Labrador y Echeburúa, 2000).

## **4.2. Tratamientos Psicológicos Basados en la Evidencia (TPBE)**

A finales del siglo XX, ya no se albergaba duda sobre la eficacia genérica de la psicoterapia en salud mental (Wampold 2001; Beutler, 2009). Fue entonces cuando se incidió especialmente en el estudio de la eficacia diferencial de las técnicas terapéuticas, para así poder establecer guías clínicas de intervenciones psicológicas, de forma análoga a las que se establecieron en medicina (Mustaca, 2004). Parece existir un consenso entre los partidarios de este enfoque, a la hora de

señalar los siguientes factores contextuales que explican y justifican el auge de esta línea de investigación desde la década de los 90 (Echeburua y Corral 2001; Bados, García y Fusté 2002; Pérez, Fernández, Fernández y Amigo 2003):

- La evolución de la ciencia psicológica y del estudio de la salud mental en general crean un entorno de gran complejidad. Aparecen multitud de nuevas técnicas psicoterapéuticas. Se requiere de la delimitación de las estrategias más eficaces y de los componentes activos de estas. Los avances en psicopatología, diagnóstico y el descubrimiento de bases biológicas en determinados cuadros clínicos, permite generar una investigación con más nivel de especificación. El modelo cognitivo conductual que aboga por el empirismo en la validación de psicoterapias se consolida a nivel académico y presiona a los demás modelos a demostrar su eficacia.
- El desarrollo y la extraordinaria promoción de los tratamientos psicofarmacológicos en decremento de la opción psicoterapéutica. Se desarrolla una nueva generación de fármacos con una mayor eficacia para la supresión de síntomas y un perfil de efectos secundarios menos desfavorable. Gracias al impulso de la industria farmacéutica se da la alienación de una parte importante de los psiquiatras entorno a la psiquiatría biológica y al uso exclusivo de esta modalidad de tratamiento. En un entorno de recursos limitados la demostración de la eficacia de la psicoterapia, se hace aun más acuciante para poder competir en el mercado de la salud mental.
- La generalización del paradigma de acceso a los servicios sanitarios a través de un tercer pagador público (el estado) o privado (los seguros). Los pagadores priorizaran tratamientos que resulten eficaces a corto plazo, con el máximo beneficio posible para el paciente pero al mismo tiempo con el menor coste posible. No solo implica a la eficacia de las intervenciones sino a su eficiencia. Además, para asegurar tanto el cumplimiento como la satisfacción del cliente-paciente, se priorizan intervenciones no demasiado complejas y prolongadas en el tiempo cuyos resultados sean fácilmente constatables. Esta forma de gestión es la denominada “Managed Care” y se ha generalizado a nivel internacional (Hayes y Greff, 2001).

#### 4.2.1. Establecimiento de la eficacia en TPBE

La iniciativa tomada por la APA con la creación del grupo Task Force on Promotion and Dissemination of psychological Procedures (TFPDP, 1995), también fue seguida en otros países (Parry, 1996; Strauss y Kaechele, 1998; Hunsley, Dobson, Johnston y Mikail, 1999). El grupo de trabajo de la APA pretendía establecer guías clínicas con listados de tratamientos, que poseían respaldo empírico suficiente como para poder asumir su eficacia. Planteando estos listados sujetos a revisiones periódicas, para ir incluyendo las nuevas evidencias que surjan. Todo ello con el fin de promulgar las prácticas más adecuadas científicamente en el entorno clínico y así mejorar la calidad del desempeño profesional. Además se señaló su posible uso como medio de influencia del colectivo profesional sobre la política gubernamental y las decisiones que impliquen al sector (Crits-Christoph, Frank, Chambless, Brody y Karp, 1995; Chambless et al., 1996; Elliott, 1998).

En este contexto, para que una intervención se pudiese considerar adecuada y de calidad para un determinado trastorno, lo primero que tenía que hacer era demostrar su eficacia en rigurosos estudios experimentales (Chambless y Ollendick (2001). En base al cumplimiento de esta condición de apoyo empírico, la APA generó una taxonomía de clasificación de tratamientos en tres niveles de evidencia con los criterios que figuran en la Tabla 5 (Chambless et al., 1998; Mustaca, 2004).

La consideración de la eficacia terapéutica se puede establecer en base a diferentes elementos por lo que no ha resultado unívoca en la literatura (Vallejo, 1998). A continuación se señalan diferentes sentidos de la eficacia en TPBE según qué componentes se tengan en cuenta (Echeburúa, Salaberría, Corral y Polo-López, 2010):

- *Especificidad e intensidad*: cuáles son los síntomas que mejoran y en qué medida mejoran.
- *Plazo y duración*: cuánto se tarda en comenzar la mejoría sintomática y cuanto se prolongan sus efectos.
- *Interacción*: cómo interactúa con otros tratamientos.
- *Balance*: efectos secundarios, abandonos y en general ventajas e inconvenientes en relación a otros tratamientos).

**Tabla 5**

*Crterios y niveles de evidencia en los “Tratamientos psicológicos basados en la Evidencia”*

<p><b>I.- Tratamientos bien establecidos.</b> <b>(Una de las dos opciones)</b></p>	<p><b>1.-</b> Al menos 2 experimentos con diseños intergrupo que demuestren eficacia por una o ambas de las siguientes vías: A) Superioridad al placebo u otro tratamiento. B) Equivalencia con tratamiento establecido con experimentos de suficiente potencia estadística (muestra grande).</p> <p><b>2.-</b> Gran cantidad de experimentos de caso único que demuestren eficacia Cumpliendo: Buenos diseños experimentales. Comparación del tratamiento con otra intervención.</p>	<p>Tanto para Opción 1 como 2:</p> <ul style="list-style-type: none"><li>- Experimentos con tratamientos manualizados.</li><li>- Características de pacientes tratados claramente especificado.</li><li>- Los efectos del tratamiento deben ser demostrados por 2 o más diferentes investigadores o grupos de investigación.</li></ul>
<p><b>II.- Tratamientos probablemente eficaces.</b> <b>(Una de las tres opciones)</b></p>	<p><b>1.-</b> Al menos 2 experimentos que demuestren eficacia con superioridad al grupo control en espera.</p> <p><b>2.-</b> Uno o más experimentos que cumplan todos los criterios de tratamiento bien establecido excepto el hecho de haber sido constatado por más de un investigador o grupo de investigación.</p> <p><b>3.-</b> Una pequeña cantidad de experimentos de diseño de caso único que cumplan todos los criterios de tratamiento bien establecido (más de tres casos).</p>	
<p><b>III.-Tratamientos experimentales.</b></p>	<p><b>1-</b> El tratamiento no cumple los criterios metodológicos de los dos niveles anteriores, pero existen resultados prometedores que deben seguir siendo estudiados.</p>	

Nota: Tabla generada a partir del contenido de Chambles et al. (1996) y Mustaca (2004).

La utilización de ensayos clínicos para determinar la eficacia de las intervenciones, requiere del cumplimiento de una serie de condiciones que se traducen de la siguiente forma en el caso de los TPBE (Chambles et al., 1998; Echeburúa y Corral, 2001):

- Una definición sin ambigüedad del tipo de problemática sobre la que actúa el tratamiento que se está evaluando. Clasificando a los pacientes según criterios diagnósticos operativos tales como los del DSM IV. Se descartan diagnósticos múltiples porque es necesario la homogeneidad de las muestras.
- Los tratamientos deben estar exhaustivamente descritos y estandarizados. Es necesaria la manualización de la intervención para asegurar que se ejecuta con todos los pacientes de forma semejante. Y fijarse una duración determinada del proceso.
- Cuando se compara el tratamiento con otras condiciones (grupo control o tratamientos alternativos) es necesaria la asignación aleatoria de los sujetos a cada condición y un número mínimo de sujetos en cada una (más de 30). Y el ensayo deber ser concurrente en la aplicación de las diferentes condiciones y prospectivos en la consecución de los datos.
- Realización de varias medidas de evolución sintomática y seguimiento de varios meses después del tratamiento. Los evaluadores deben desconocer la condición experimental a la que pertenecen los pacientes evaluados.

Según todos estos criterios, el mencionado grupo de trabajo de la APA, estableció listados de tratamientos eficaces y probablemente eficaces con respecto a diferentes trastornos del DSM IV. Se observó inicialmente un predominio absoluto de la orientación cognitiva conductual, atenuado posteriormente con la inclusión en sucesivas actualizaciones de técnicas de otros modelos, aunque sin perder en ningún momento, esta orientación, su protagonismo. En las mismas revisiones se explicitó que el no estar contemplado como técnica eficaz, no implica necesariamente ineficacia real, sino que aún no existe evidencia empírica para respaldar sus efectos, animando afrontar esta validación desde los criterios señalados (Chambles et al., 1996; Chambless et al., 1998; Chambles y Ollendick, 2001).

Bajo un paradigma similar el Nacional Institute for Clinical Excellence (NICE) inglés generó una serie de guías clínicas para trastornos mentales específicos, donde recomendó que las técnicas psicológicas identificadas como eficaces (señalando también a la orientación

cognitiva conductual) fuesen la opción inicial para el tratamiento de los trastornos anímicos (NICE, 2004a; 2004b; 2005a; 2005b; 2006).

#### **4.2.2. Utilidad clínica de los TPBE**

Pese al gran peso que se dio a la determinación de la eficacia en condiciones de riguroso control experimental, dentro del movimiento de la TPBE esto se señaló como condición necesaria pero no suficiente para guiar a la práctica clínica. El segundo elemento que se destaca es la determinación de la utilidad clínica de la intervención o efectividad (Nathan y Gorman, 1998). Para Bados, Garcia y Fusté (2002), la utilidad clínica de una intervención psicológica implica tres características: viabilidad, generalización a situación real y eficiencia.

En relación a la viabilidad se ha constatado que en la práctica clínica un cuarto de los pacientes que busca ayuda psicológica no acepta el tratamiento propuesto en la primera sesión; y alrededor de un 65% de los que comienzan abandonan el tratamiento antes de la décima sesión (Berenger y Quintanilla, 1994; Castañeda y Mira, 1998). La complejidad, el coste económico y la duración de los tratamientos son variables que van a condicionar la continuidad del tratamiento por parte del paciente (Bados, Garcia y Fusté, 2002).

Según la revisión de Pekarik (1993), las expectativas iniciales de los pacientes se orientan en su mayoría hacia una intervención de menos de diez sesiones, que se fije en solucionar problemas muy concretos, lo que contrasta con la preferencia de la mayoría de los terapeutas de intervenciones más prolongadas en el tiempo. También se ha observado que la probabilidad de incumplimiento de pautas o prescripciones del terapeuta aumenta cuanto más complejas y difíciles de entender le sean al paciente (Foa, Rothbaum, Riggs, Murdock, 1991). Otro elemento que condiciona la viabilidad es la asunción de la técnica terapéutica eficaz por el propio profesional. Se ha constatado que los profesionales son muy resistentes a cambiar los métodos y formas de actuar una vez los han asimilado y puesto en práctica durante mucho tiempo. Y que la existencia de manuales de tratamiento claro y específico son determinantes para vencer la resistencia al cambio de los clínicos (Wilson, 1996; Addis, 1997; Becoña, 1999; Barraca, 2005).

Con respecto a la generalización de la eficacia a situaciones reales, Seligman (1995) planteó como estrategia de comprobación de efectividad, diseños de investigación naturalísticos que

compensen su bajo control experimental con un gran tamaño de la muestra, y donde se recojan resultados desde la perspectiva del paciente y lo útil que le ha resultado la terapia. Destaca este tipo de estudios, como el complemento ideal a los estudios de eficacia a la hora de generar verdadera evidencia. Como ejemplo de esta propuesta analiza los resultados de una gran encuesta realizada en EEUU por la revista Consumers Reports en relación a la experiencia de los de los servicios de salud mental. El autor llegó a la conclusión de que: las terapias de mayor duración redundaron en mejores resultados para los pacientes, los especialistas (psicólogos y psiquiatras) obtienen mejores resultados a largo plazo que otro tipo de asistentes en problemáticas psicológicas, las distintas modalidades específicas de psicoterapia se mostraron similares en resultados, se comprobó un nivel de satisfacción alto de los usuarios con respecto al trabajo psicoterapéutico y se constató más beneficio de la terapia en pacientes que optaban por una actitud activa respecto al proceso.

Aunque se reconoce el citado estudio de Seligman como un intento importante de investigar los resultados de la psicoterapia en el ámbito natural, se han señalado importantes limitaciones metodológicas que impiden una generalización real de sus conclusiones: sólo respondió un porcentaje bajo de las miles de personas a las que se le mandó la encuesta, no existe constatación de diagnóstico clínico mezclándose en los mismos resultados diferentes niveles de gravedad psicopatológica con problemáticas subclínicas, los datos obtenidos son retrospectivos y basados exclusivamente en autoinformes generales, no existen medidas previas ni grupos de control y las respuestas provienen en su mayoría de adultos de clase media (Hunt, 1996; Bados y García-Grau, 2011). Se ha terminado concluyendo que las encuestas planteadas de esta forma son importantes estudios de la satisfacción de los consumidores, pero no una medida real de la utilidad clínica de la psicoterapia, hallándose una sobredimensión del cambio terapéutico conseguido cuando se compara el uso de medidas retrospectivas generales como estas, con medidas prospectivas más específicas (Nielsen et al., 2004; Feixas et al., 2012).

La propuesta desde el ámbito de los TPBE, es constatar si los resultados de la aplicación de tratamientos con apoyo empírico es equivalente en el ámbito clínico con respecto al ámbito experimental controlado. Esta segunda fase de la validación de un tratamiento, permitiría generalizar adecuadamente los resultados de eficacia (Chambles y Ollendick, 2001). En este sentido Shadish, Matt, Navarro y Phillips (2000) realizan una revisión meta analítica de estudios

sobre resultados de intervenciones en contextos reales y concluyen que las técnicas que poseen apoyo empírico, siguen manteniendo eficacia en el contexto clínico, destacándose con un mayor tamaño del efecto las medidas de resultado directamente relacionadas con las metas explicitadas en el tratamiento y las intervenciones de mayor duración. Sin embargo, la revisión de Hansen, Lambert y Forman (2002) no fue tan optimista y de hecho mostró que las tasas de mejoría eran bastante inferiores en los estudios de situaciones clínicas que los resultados de los ensayos clínicos controlados.

Bados y García-Grau (2011) destacan varias razones por las que los pacientes en estudios de efectividad mejoren menos y más lento que en los estudios de eficacia: la comorbilidad de problemáticas, la contaminación de experiencias pasadas con otras intervenciones en salud mental, menor nivel socioeconómico, diferentes condiciones culturales y étnicas que los sujetos con los que se ha validado empíricamente la intervención (normalmente occidentales blancos de clase media). Se muestra muy relevante estudiar la efectividad en diferentes contextos y condiciones de clínica real, pero parece existir escasos estudios al respecto y la mayoría de lo que hay se centran en unas pocas problemáticas específicas (Margni, Molina, Sartori, Rebull y Vazquez, 2012; Ballesteros, Fernandez y Labrador, 2013).

En lo que atañe a la eficiencia, se puede señalar que no se debe entender como una simple contención del gasto coherente con intereses economicistas de los modelos de gestión actuales sino una verdadera consideración de costes y beneficios a varios niveles (Labrador, Echeburua y Becoña, 2000): terapéutico (duración, costes beneficios a nivel individual), sanitario ( ahorro en otros servicios y en usos siguientes del mismo servicio) y social (menor absentismo laboral, menos uso de recursos sociales, prevención de problemas sistémicos en familias etc...). En relación a esto se ha constatado, en distintas revisiones, que la utilización de servicios médicos se redujo significativamente después de que los usuarios hubiesen recibido psicoterapia (Olfson, Sing y Schlesinger, 1984; Cummings, O'Donohue y Ferguson, 2002).

Otros estudios han mostrado cómo las técnicas cognitiva conductuales para trastornos anímicos se muestran más eficientes que la terapia exclusivamente farmacológica, resultando esta última más cara a lo largo del tiempo (Antonuccio, Thomas y Danton, 1997; Otto, Pollack y Maki, 2000; Pyne, et al., 2005; Cuijpers, van Straten, Hollon y Andersson, 2010). En otras revisiones se ha observado que la terapia psicológica podría tener un impacto beneficioso en

diferentes costes médicos y laborales cuando se incluye en el tratamiento de los trastornos psiquiátricos graves (Gabbard, Lazar, Hornberger y Spiegel, 1997; Dixon-Gordon, Turner y Chapman, 2011). De forma genérica se ha observado que los tratamientos psicológicos están equilibrados desde el punto de vista coste beneficio, redundando además en una mayor productividad de los individuos en la sociedad (Chiles, Lambert y Hatch, 2002).

Layard (2006a) revisando los datos de anteriores estudios, realiza una estimación del gran ahorro que se produciría en el sistema nacional de salud británico si se introdujesen en él terapeutas formados en terapias basadas en la evidencia (los explicitados en las guías NICE), para tratar los problemas de ansiedad y depresión (en su mayoría técnicas del modelo cognitivo conductual) de la población. Este autor extendió su argumentación sobre la eficiencia de las TPBE, concretándolo en el tratamiento para la depresión, estableciendo que los costes para el gobierno británico de invertir psicoterapeutas formados en técnicas eficaces, se verían superados por los ahorros potenciales para la administración y los beneficios para la sociedad y los individuos (Layard, 2006b; Layard et al., 2006; Layard, Clark, Knapp y Mayraz, 2007). Finalmente el gobierno británico se hizo eco de demandas de Layard y otros académicos e inicio en el 2007 el programa Improving Access to Psychological Therapies IAPT en proyecto contratar y formar a 3600 psicólogos terapeutas para que atendiesen a los problemas de depresión y ansiedad de los usuarios del sistema nacional de salud, según los criterios de tratamientos del NICE (Cohen, 2008). Ya existen datos publicados que reportan indicadores de eficiencia favorables en la primera fase del programa (Richards y Suckling, 2008; 2009; Clark et al., 2009).

#### **4.2.3. Limitaciones de la TPBE**

Las revisiones TPBE, pretenden aumentar la calidad de intervenciones de los profesionales, a través de la generación guías clínicas de tratamientos para diagnósticos específicos. Esta tarea representa un recurso importante para la práctica clínica pero desde su inicio se han planteado tanto desde el ámbito clínico como investigador una serie de críticas y limitaciones a su pretensión de mejorar la práctica real, que se pueden agrupar en tres puntos clave (Echeburúa, Salaberría, Corral y Polo-López, 2010; Primero y Moriana, 2011):

- La forma de definir las problemáticas clínicas en base a criterios de clasificaciones psiquiátricas categóricas.

- Las grandes diferencias entre las condiciones de práctica real y los ensayos experimentales controlados.
- Las variables contempladas en los estudios de TPBE para demostrar la eficacia de las estrategias de intervención.

Las guías clínicas bajo el paradigma de TPBE se articularon sobre la base de clasificaciones internacionales de trastornos psiquiátricos (como CIE y DSM). Estas clasificaciones forman categorías de trastornos de forma análoga a otros sistemas taxonómicos dentro de la medicina. Se basan en la inferencia de una entidad natural discreta a través de una determinada agrupación de signos y síntomas (síndromes). Esta estrategia útil en otras problemáticas de salud puede resultar insuficiente con respecto a las problemáticas psicológicas donde los mismos síntomas pueden tener causas diferentes y cursos diversos. Al contrario que en otros diagnósticos en salud, no es evidente un correlato estructural que pueda confirmar la presencia del trastorno. Tampoco se da la condición de categorías excluyentes de las clasificaciones biológicas. Y su identificación se ve afectada por factores socioculturales. Se ha criticado que estas clasificaciones sobredimensionan la importancia de la fiabilidad a la hora de definir sus garantías científicas prestando poca atención a la validez de sus criterios. Todo esto puede llevar a problematizar la generalización de los tratamientos (Millon, Krueger y Simonsen, 2008; Duero y Shapoff, 2009; Primero y Moriana, 2011).

Con su última edición el DSM (American Psychiatric Association APA<sub>2</sub>, 2013) se reformula la clasificación incluyendo nuevas categorías de validez muy dudosa, lo que ha generado un rechazo de diversas instituciones, volviendo a cobrar fuerza otras alternativas diagnósticas de tipo dimensional o funcional (Marsá y Aragona 2009; Primero y Moriana, 2011; Pérez, 2012; Domingo, 2013). Por otro lado, al margen de la validez de las categorías, se ha constatado en diversos estudios en contextos reales, que la mitad de los pacientes que acuden a un servicio de salud mental no pueden diagnosticarse según las clasificaciones categoriales porque sus problemas no encajan o porque presentan niveles de expresión sintomática por debajo de los necesarios para entrar en una categoría (Stirman, DeRubeis, Crits-Christoph y Brody, 2003; Westen, Novotny y Thompson-Brenner, 2004). Actualmente, dentro de las demandas terapéuticas, existen situaciones de sufrimiento o infelicidad que no se ajustan a los cuadros clínicos habituales (problemáticas subclínicas) y el ir generando nuevas categorías de

“trastorno mental” de forma arbitraria aporta poco al estudio y desarrollo de estrategias de intervención eficaces para ayudar a estas personas, ya que esta dinámica solo responde a intereses de mercado de las industrias farmacológicas (González-Pardo y Pérez-Alvarez, 2007; Echeburúa, Corral y Salaberría, 2010).

El apoyo empírico necesario para que una intervención se considere TPBE, implica la demostración de su eficacia bajo condiciones de riguroso control experimental, es decir, donde se asegure la validez interna. Sin embargo, ganar validez interna conlleva perder validez externa, y viceversa. Controlar los factores que pueden sesgar la investigación hace difícil emular las condiciones naturales donde se ejecutara esa intervención. Por lo que no es extraño constatar que hay cierta distancia entre lo que se consigue en los estudios experimentales y la que sucede en la práctica clínica real. Al contrario que en los estudios experimentales en la clínica real los pacientes presentan habitualmente comorbilidad de problemáticas siendo poco común los diagnósticos puros. Sus características personales serán muy diversas y también presentarán diferentes niveles de gravedad siendo imposible identificar grupos homogéneos de pacientes. Además pueden haber tenido experiencia en tratamientos previos condicionando así los resultados de la actual intervención. Por otro lado, los terapeutas contarán con diversos niveles de experiencia y competencia que le llevarán a obtener resultados diferentes con la misma técnica, además de enmarcarse en un contexto donde la intervención no está limitada por una duración y contenidos prefijados como en la situaciones experimentales, ajustándose a durante el propio proceso a las necesidad de caso en particular (Pérez 1996; Botella 2000; Echeburúa y Corral, 2001; Westen y Morrison, 2001; Muñoz-Solomando, Kendall y Whittington, 2008).

Aunque se han generado estudios experimentales específicos con respecto la técnica terapéutica según variables de los pacientes tales como los rangos de edad, (Gatz et al., 1998; Kazdin, 2003; Muñoz-Solomando, Kendall, y Whittington, 2008), pueden existir un infinidad de combinaciones de características que ante un mismo diagnóstico y técnica terapéutica generaran resultados diversos (Perepletchikova y Kazdin, 2005). Se hace necesario complementar la evidencia de eficacia experimental con estudios de efectividad que observen los resultados de las intervenciones en la práctica habitual para poder validar los tratamientos adecuadamente (Kazdin 2001; Frias y Llobell, 2003). En los últimos años han empezado a generarse estudios que tratan de cubrir esta necesidad cuyas conclusiones señalan al mantenimiento de la eficacia

de las intervenciones con apoyo empírico (Westbrook y Kirk, 2005; Wales, Palmer y Fairburn, 2009; Minami et al., 2009). Sin embargo, tal y como se ha mencionado anteriormente los contextos representados en estos estudios de efectividad son aún limitados y es necesario un mayor volumen de investigación en diferentes condiciones y entornos (Margni, Fernanda, Sartori, Rebull y Vazquez, 2012; Ballesteros, Fernandez y Labrador, 2013).

La línea de investigación que se planteó con los TPBE se centró fundamentalmente en la técnica de intervención y en los resultados expresados en indicadores de reducción sintomática o desaparición del cuadro clínico. Han dejado en segundo plano otras variables relevantes para el cambio terapéutico como las características del terapeuta, del paciente y su interacción. No han tenido en cuenta, en la valoración del cambio, resultados de otra naturaleza más general o de espectro más amplio como constatación de cambios relevantes en el día a día del paciente, el cumplimiento de sus objetivos o su calidad de vida (Gardfield 1998; Wampold 2001b; Addis, Wade Hatgis 1999, Fernández y Pérez, 2001; Rodríguez 2004; Lerner, 2004).

Por otro lado, las condiciones establecidas para los estudios de eficacia plantean una serie de requisitos difíciles de cumplir desde modelos teóricos donde no existe un protocolización del tratamiento tan clara, un diagnóstico categorial o la mencionada perspectiva de resultados, generando una baja representatividad de técnicas fuera del modelo conductual en las guías clínicas de TPBE (Bohart, O'Hara y Leitner, 1998; Shedler, 2010; Parker, 2011; Rogers, Maidman y House, 2011). Se ha criticado también la forma de seleccionar la evidencia, problematizando la superioridad de la perspectiva cognitivo conductual en lo referente a eficacia demostrada: señalando sesgos de selección del revisor o el investigador, focalización en cierto tipo de resultados sin información de otras áreas, valorando tratamientos en forma de paquetes heterogéneos en vez de por componentes y tratando los efectos de la psicoterapia como de si un medicamento se tratase (Elliot 1998; Elkins 2009; Wampold, 2013).

Sobre si existen técnicas psicoterapéuticas superiores a otras o sobre la magnitud de la importancia del método utilizado sobre los resultados, sigue existiendo gran discrepancia entre los académicos y no se ha llegado a una sentencia definitiva (Cautilli, 2006; Lilienfeld, 2007; Baker, McFall, y Shoham, 2008; Ehlers, Bisson, Clark y Yule, 2010; Norcross, 2011; Stiles, 2013). Sin embargo sí parece existir un interés general en aumentar el foco de estudio a otras variables ya sea como moderadoras del efecto o como elementos determinantes de los

resultados, identificando los factores que marcan la diferencia en el proceso terapéutico (Stiles, Honos-Webb y Knobloch, 1999, Castonguay y Beutler, 2006, Kazdin y Whitley, 2006, Kadzin, 2011, Froján, 2011, Lambert, 2013).

#### **4.2.4. TPBE en psicología de la salud**

El rol clásico de las intervenciones psicológicas en salud se había visto restringido al ámbito de la salud mental y a los llamados trastornos mentales. Sin embargo, a partir de la década de los setenta acontece un cambio paradigmático en la consideración de la salud general, que genera un nuevo campo de estudio en la psicología, y que se explica por los dos siguientes puntos (Engel, 1977; Reig 2005; Amigo, Fernández y Pérez, 2010):

- El modelo biomédico tradicional basado en el reduccionismo de la enfermedad a unívocos correlatos químicos y físicos, se muestra insuficiente para explicar el proceso de salud-enfermedad. Se asume una nueva conceptualización al constatar que los factores psicológicos y sociales son también determinantes. Este modelo biopsicosocial trata de superar la falsa dualidad mente cuerpo y entiende la salud de una forma sistémica y producto de la interacción de diversas variables.
- En la segunda mitad del siglo XX, toman protagonismo los trastornos crónicos relacionados con el estilo de vida de las sociedades industrializadas. En este tipo de problemáticas el peso de factores tales como hábitos de vida, actitudes o comportamientos insanos es muy evidente no solo en su etiología, sino en el curso y pronóstico. Implicando además un gran impacto sobre la calidad y estilo de vida de las personas. Ante estas problemáticas, las estrategias para proveer de salud a la población necesariamente van más allá del ajuste de desequilibrios biológicos.

En respuesta a estos elementos se reconoce en la APA una división específica para la denominada Psicología de la Salud (Ballester, 1998). Matarazzo (1980; 1982) delimitó el papel de esta área con las siguientes focos de aportación: promoción y mantenimiento de la salud, prevención y tratamiento de la enfermedad, e identificación de los correlatos etiológicos y diagnósticos de la salud, la enfermedad y las disfunciones relacionadas. Lo que se tradujo en cuatro tipos de estrategias concretas (Oblitas, 2008): promoción de estilos de vida saludables,

prevención de enfermedades, evaluación y mejora del sistema sanitario y desarrollo de intervenciones para tratamiento de enfermedades.

Respecto a esta última estrategia, la constatación del impacto de los factores psicológicos sobre el funcionamiento del sistema inmune y el éxito de técnicas para el control de ciertas enfermedades desarrolladas en el ámbito de la medicina conductual (tales como el biofeedback), provocó que se planteasen numerosas intervenciones psicológicas para el tratamiento de patologías específicas y la normalización del rol del psicólogo en el ámbito médico y de la salud “física” (Mustaca y Bentosela 1995; Segerstrom y Miller 2004; Taylor, 2007).

Para Oblitas y Becoña (2000) la demanda y urgencia de este tipo de intervenciones respecto a los distintos trastornos de salud y el rápido desarrollo de este campo en muy poco tiempo, ha provocado que en muchos casos la aplicación de los distintos programas de intervención haya ido por delante de la evaluación de la eficacia de los mismos. Se plantea la necesidad de identificar las estrategias más eficaces con criterios semejantes a los utilizados para las intervenciones en salud mental del paradigma de TPBE, representando esto un elemento indispensable para su adecuada implementación en un contexto interdisciplinario (Compas, Haaga, Keege, Leitenbery y Williams, 1998). Pero se reconoce la complejidad de evaluar sus efectos por la integración con los medios médicos, ya que la intervención psicológica representa solo una parte de la totalidad del tratamiento. Además se señala la necesaria consideración de objetivos y medidas de resultado más allá de la reducción sintomática, tales como la calidad de vida de los pacientes o su adhesión al tratamiento prescrito (Berrío-Costa y Hernández-Zubieta, 2003).

Al respecto de esta labor de evaluación, Pérez, Fernandez, Fernandez y Amigo (2003b) realizan una extensa revisión de la evidencia de eficacia de intervenciones en psicología de la salud, en la Tabla 6 se pueden ver las áreas destacadas respecto a la evidencia de resultados positivos. Estos autores concluyen que pese a mostrar resultados prometedores en diferentes problemáticas, la investigación en este área es más limitada que en los trastornos mentales y que sería deseable la consecución de un mayor número de estudios acerca de los resultados de estas intervenciones.

**Tabla 6**

Revisión de la evidencia acerca de la eficacia de intervenciones psicológicas en salud en TPBE.

<b>Problemáticas</b>	<b>Estrategias intervención</b>	<b>Objetivos</b>
<i>Asma bronquial</i>	Relajación	Control reactividad emocional
	Educación terapéutica	Automanejo enfermedad
	Biofeedback	Regulación función pulmonar
<i>Diabetes</i>	Terapia familiar	Adhesión al tratamiento
	Programas conductuales	Automanejo enfermedad
	Entrenamiento afrontamiento	Control glucemia
	Educación terapéutica	Control peso y alimentación
<i>VIH</i>	Reestructuración cognitiva	
	Educación terapéutica	Adhesión al tratamiento
	Counselling	Regular estado emocional
	Modificación de conducta	Afrontamiento de enfermedad
<i>Cáncer</i>	Modificación de conducta	
	Relajación / Visualización	Control de efectos y síntomas
	Desensibilización sistemática	Automanejo y afrontamiento
	Terapia de grupo	Regulación emocional
	Counselling / Educación	
<i>Problemas cardiovasculares</i>	Educación terapéutica	Adhesión al tratamiento
	Modificación de conducta	Control factores de riesgo
	Biofeedback	Control parámetros cardio
	Relajación/ meditación	Manejo de estrés y ansiedad
<i>Dolor crónico</i>	Relajación	
	Programas multicomponentes	Control sintomático
	Biofeedback	Manejo de enfermedad
<i>Fibromialgia y otras enfermedades reumáticas</i>	Relajación	
	Programas multicomponentes (cognitivo conductuales)	Control sintomático Afrontamiento de enfermedad
<i>Enfermos terminales</i>	Counseling	Cuidados paliativos/ calidad Satisfacción con atención
<i>Problemas tracto digestivo</i>	Biofeedback	
	Programas multicomponentes (cognitivo conductuales)	Manejo del estrés Manejo de contingencias
	Técnicas de exposición	Reducción ansiedad
	Desensibilización sistemática	Control sintomático

Nota: Generada a partir del contenido de Pérez, Fernandez, Fernandez y Amigo (2003)

En la actualidad el contemplar los factores psicológicos se ha mostrado fundamental en el cuidado y tratamiento de diferentes patologías médicas, especialmente de las que tienen un curso crónico (Meeus y Nijs, 2007; Weiner, 2008; Tobón, Sandín, Vinaccia y Núñez, 2008; Havelka, Lucanin y Lucanin, 2009; Larenas-Linnemann, 2009; Salgueiro et al., 2009). Por ello, el seguir investigando en las aportaciones específicas a cada problemática se señala como una prioridad en psicología de la salud (Bennet, Pery, Rozensky, 2002; Amigo, Fenández y Pérez, 2010). Por otro lado, diversos autores plantean que esta disciplina no debería amoldarse a los esquemas diagnósticos médicos, sino investigar a partir de las necesidades particulares de los pacientes, sus características y sus condiciones en el proceso de la salud y enfermedad, independientemente de la identificación de un trastorno concreto (Marks et al., 2008; Fergusson, 2013).

Sobre las líneas de actuación futuras en esta disciplina, Garzón y Caro (2014) señalan la necesidad de realizar una integración práctica de conocimientos que aúne la experiencia clínica, los modelos teóricos y la investigación. Remarcan, estas autoras, en especial la importancia del acercamiento del ámbito clínico y académico para generar medios útiles y eficientes, ya que el contexto macroeconómico actual lo hace muy necesario. Taylor (2007) también coincide en señalar como objetivos principales el establecimiento de eficiencia y efectividad de las diferentes estrategias psicológicas en salud, siendo para ello indispensable la evaluación de resultados en el ámbito clínico natural.

#### **4.3. Practica Psicológica Basada en la Evidencia (PPBE)**

Con el fin de superar las limitaciones de los TPBE, la American Psychological Association (APA, 2006) publicó un informe denominado “La practica basada en la evidencia en Psicología” que seguía manteniendo el mismo objetivo de integrar los avances de la ciencia psicológica con la práctica. Sin embargo, planteó un punto de vista más holístico sobre las evidencias a considerar: los resultados de la investigación de los diferentes elementos, la experiencia clínica y las características de los pacientes. Se propone un proceso de toma de decisiones en la práctica donde se aúnen las múltiples corrientes de evidencias de investigación, incluida la eficacia de los tratamientos. El foco se pone en el paciente en vez de en la técnica y la evidencia se instrumentaliza para obtener los mejores resultados en cada caso (Kazdin, 2008; Botella, 2009).

Aunque esta perspectiva se acomoda a la misma definición de toma de decisiones adoptada en la MBE (Sackett, Straus, Richardson, Rosenberg y Haynes, 2000), la naturaleza de las intervenciones psicológicas no puede ser directamente equiparable a los tratamientos médicos, no pudiendo realizar un trasvase directo de las metodologías propuestas en la MBE. En psicología las medidas de resultados, son más difíciles de objetivizar que en otras áreas de salud y las intervenciones poseen más fuentes de variación que los fármacos (Vazquez y Nieto, 2005). No es de extrañar pues que en la PPBE caracterizada por la APA se ponga un énfasis especial en la integración bidireccional de la práctica y la investigación (APA, 2006; Henton, 2012).

Recientemente la APA ha aprobado una resolución sobre la situación contemporánea de la investigación de eficacia y efectividad de la psicoterapia. A parte de informar la evidencia de su eficacia en diferentes contextos (y sus ventajas frente a otras intervenciones en salud mental), señala como línea de investigación prioritaria el estudio de los factores que expliquen la consecución de resultados positivos o negativos, teniendo en cuenta variables de paciente, del terapeuta, de la problemática, de la técnica o de la combinación de estas, con el fin de seguir mejorando la calidad de las intervenciones y permitir a los clínicos tomar mejores decisiones (APA, 2012; Campbell, Norcross, Vasquez y Kaslow, 2013).

#### **4.3.1. Factores comunes en intervenciones psicológicas**

En cierta contraposición al paradigma del TPBE, se desarrolló otra línea de investigación de resultados terapéuticos. Según la cual los factores comunes (o también llamados inespecíficos) de las diferentes formas y modelos de psicoterapia son los ingredientes activos más relevantes para explicar los cambios acontecidos a la intervención, teniendo mucho menos peso el tipo de técnica terapéutica específica utilizada (Drisko 2004; Imel y Wampold 2008). Según las revisiones realizadas en este sentido un 70% de varianza de los resultados se debe a los factores comunes (Duncan, Miller, Wampold, y Hubble, 2010). Clásicamente se han destacado tres categorías de variables inespecíficas: las relacionadas con el terapeuta, las relacionadas con el paciente y las relacionadas con la interacción terapéutica entre ambos (Romero, 2008).

Las variables del terapeuta interactúan con las características del paciente de una forma compleja haciendo difícil discriminar, lo que aporta cada una al proceso terapéutico (Santibáñez et al., 2008). Aunque se ha evidenciado el gran impacto de las características del terapeuta en los resultados, estas por separado han recibido menos atención de los investigadores que otros factores comunes y se han llegado a escasas conclusiones definitivas (Beutler, 1997). En la revisión realizada por Corbella y Botella (2004) se agrupan las variables estudiadas en la literatura en cuatro ejes interconectados:

- La persona del terapeuta: características sociodemográficas (edad, etnia y género), características profesionales (experiencia, formación y afiliación teórica) y personalidad (estilo cognitivo, locus de control, flexibilidad, dominancia, bienestar emocional, y valores).
- Posicionamiento del terapeuta en referencia al marco conceptual: respecto a la orientación teórica (dogmático o crítico), respecto a la investigación (incredulidad, aceptación o distanciamiento) respecto a experiencia profesional (experiencial neutro o analítico).
- Actitud frente a terapia: actitud integradora y actitud restringida.
- Estilo personal del terapeuta (modos de comunicación dominantes que utiliza derivado de ejes anteriores): función instruccional (de rígido a flexible), atencional (de activo a receptivo), expresiva (de próximo a distante), operacional (de directivo a no directivo), evaluativa (de orientados al resultado a orientados a la comprensión), fomentativa (de dirigidos a la acción a dirigidos al insight) e involucración (de terapeuta muy comprometido a poco comprometido).

La investigación de las características del terapeuta según los ejes expuestos, no orienta hacia un perfil ideal general, sino a las relaciones de cada tipo terapeuta con cada tipo de paciente y sus efecto en los resultados (Fernández-Alvarez, Garcia, LoBianco y Corbella, 2003; Corbella y Botella, 2004; Hill, 2006). Pese a ese relativismo respecto a las características de los pacientes, si se ha podido señalar la idoneidad genérica de una serie de habilidades interpersonales del terapeuta, que combinadas se han asociado a mejores resultados en psicoterapia, que son las siguientes: fluidez verbal, percepción interpersonal, modulación

emocional y expresividad, calidez y aceptación y focalización en el otro (Anderson, Gole, Patterson, Lambert y Vermeersch, 2009; Wampold, 2011).

Con respecto a los factores del paciente, se han llegado a proponer más de 200 variables relevantes para los resultados terapéuticos, de las cuáles un centenar ya han sido sujetas a escrutinio de alguna investigación (Clarkin y Levy, 2004). La estimación inicial de Lambert (1992) identificó estos factores como las más influyentes en los resultados finales de la psicoterapia. Los estudios con las características sociodemográficas (genero, edad, educación y nivel socioeconómico) de los pacientes, no reportan conclusiones relevantes respecto a diferencias de resultados entre los diferentes grupos (Santibáñez et al., 2008). Por otro lado, la problemática clínica, si se ha encontrado como factor determinante en especial el nivel de perturbación psicológica que le ocasiona al paciente, generándose peores resultados cuanto mayor sea esta (Trull y Jerry, 2003). En referencia a las variables de personalidad se anima a evaluarlas y considerarlas en el proceso terapéutico, ya que la eficacia de las estrategias de intervención puede verse comprometida por estos factores (Bihlar y Carlsson, 2001). Un ejemplo de esto son los estudios sobre estilos de afrontamiento, en los cuales se han mostrado como un moderador evidente de la relación entre técnica terapéutica y los resultados (Beutler y Harwood 2000).

Por lo anterior Norcross y Wampold (2011), señalan que la función real de estudiar el impacto de las variables del paciente, es poder generar evidencias que permitan adaptar la intervención psicológica a las particularidades del paciente individual, de acuerdo con las generalizaciones identificadas en la investigación. Estos mismos autores realizan una revisión de metaanálisis, con respecto a la eficacia de acomodar la psicoterapia a determinadas variables del paciente, concluyendo que existe evidencia consistente respecto la consideración de las características: nivel de reactancia, preferencias respecto a intervención, espiritualidad/religiosidad y cultura. Según su revisión la adaptación de los variables estilos de afrontamiento y el nivel de disposición al cambio son probablemente efectivos aunque no se pueden afirmar con el mismo nivel de evidencia que las anteriores características. Y por último, señalan como líneas de investigación prometedoras las investigaciones respecto a estilos de apego y expectativas aunque aún no se dispone de suficiente apoyo empírico para realizar conclusiones fiables.

Los aspectos interpersonales de la relación terapéutica, han sido reivindicados desde hace décadas como determinantes para la eficacia de la terapia, independientemente de la problemática o la estrategia de intervención. Este elemento ha recibido la atención desde la investigación y se ha constatado el gran impacto que representa el tipo de relación terapéutica sobre el proceso y los resultados; señalándolo como un factor crítico a la hora de generar cambios relevantes en el paciente (Corbella y Botella, 2004; Wampold, 2010).

La consideración de su importancia es tal que de forma reactiva a la generación del grupo Task Force on Promotion and Dissemination of Psychological Procedures (y sus revisiones de TPBE), se creó desde la División 29 de la American Psychological Association otro grupo de trabajo denominado Empirically Supported Therapy Relationships, con el objetivo de identificar, operativizar y difundir la información disponible basada en la evidencia científica, acerca del papel de la relación terapéutica en la eficacia de las intervenciones (Ackerman et al., 2001). Las últimas conclusiones divulgadas por este grupo de trabajo (basadas en diversos metaanálisis) las recoge Norcross (2011) y se pueden sintetizar en los siguientes puntos:

- Se demuestra que la relación terapéutica hace contribuciones consistentes a los resultados de la psicoterapia. Esta explica por qué los pacientes mejoran o no mejoran tanto como la estrategia de intervención ejecutada.
- Los esfuerzos para promulgar las mejores prácticas o prácticas basadas en la evidencia deben incluir la consideración de la relación terapéutica. Las guías de práctica y tratamiento deben abordar explícitamente conductas de terapeuta y cualidades que promueven una relación terapéutica facilitadora.
- Respecto a la fuerza probatoria de las evidencias se destaca como elementos indudablemente eficaces: el desarrollo de una buena alianza terapéutica (con evidencia en terapia individual, familiar y juvenil), la cohesión grupal (en terapia grupal), buen nivel de empatía del terapeuta y el hecho de recoger feedback del estado del paciente durante el proceso terapéutico.

El constructo de alianza terapéutica se ha mostrado de singular relevancia en el estudio de relación terapéutica; mostrando en la investigación correlación significativa con diversos índices de resultados (Hovath, 2005). Aunque el concepto tiene su origen en el modelo psicoanalítico, este ha sido asimilado por la mayoría de las escuelas psicoterapéuticas despojándolo de la lectura

trasferencial que poseía inicialmente. Se considera que la alianza terapéutica es una construcción conjunta entre el paciente y el terapeuta por lo que las expectativas y opiniones que ambos se van formando en el curso de la terapia son relevantes para su establecimiento (Safran y Muran 2000; Corbella y Botella 2003). Independientemente de los matices de cada modelo teórico, la alianza terapéutica se entiende como el encaje y colaboración entre el paciente y el terapeuta. Y esta integra tres componentes que determinan su calidad y fuerza: vínculo emocional positivo, acuerdo en tareas y acuerdo en objetivos (Santibáñez, Roman y Vinet, 2009).

Como ya se ha expuesto anteriormente la supuesta igualdad de eficacia de las técnicas psicoterapéuticas incentivó el estudio de los factores comunes, definiéndose esta línea de investigación de forma antagónica con la de los TPBE. Por su parte, algunos partidarios del estudio de la eficacia diferencial de las diferentes técnicas, han puesto en duda los métodos utilizados para declarar la no diferencia entre estrategias de intervención y han señalado que se ha sobreestimado el impacto e importancia del estudio de los factores comunes (Luborsky et al., 2002; Chambless, 2002; 2007; Fernández-Hermida, 2012). Pese a ello, actualmente existe un reconocimiento general de que la investigación de resultados, no se puede reducir a uno u otro foco, y que se hace necesario integrar las distintas evidencias para identificar los componentes que generan el cambio terapéutico (Reisner, 2005; Castonguay y Beutler, 2006; Wampold, 2010; Pérez, 2013).

#### **4.3.2. Investigación en la práctica clínica**

La distancia entre la investigación científica y la práctica clínica ha representado una realidad persistente y preocupante para los académicos (Fall, 1991; Goldfried y Wolfe 1996; Becoña, 1999; Beutler 2009; Teachman et al., 2013).

El movimiento de los TPBE, se planteó con el objetivo de reducir esta distancia aunque con un enfoque unidireccional desde la investigación a la práctica. Esperando que los clínicos asimilasen y aceptasen sus conclusiones independientemente de si estas eran relevantes para su desempeño diario (Persons y Silberschatz, 1998; Goldfield y Wolfe, 1998). Lampropoulos y colaboradores (2002) señalan que este tipo de “imperialismo empírico” no ha sabido abordar realmente las preocupaciones y cuestiones a las que los clínicos se tienen que enfrentar en su desempeño. Por ello, y aunque los TPBE han representado potente revulsivo para la

investigación de los diferentes componentes del proceso terapéutico, el impacto de los listados de tratamientos empíricamente validados ha sido limitado tanto el ámbito aplicado como en el formativo (Woody, Weisz y McLean, 2005).

Por otro lado, se ha criticado la idoneidad de los ensayos clínicos controlados como principal patrón de referencia para evaluar los resultados de las intervenciones, ya que la realidad compleja de las terapias psicológicas no se puede operativizar como si de un fármaco se tratase y sus resultados van más allá de la remisión de una patología concreta (Elkins, 2009; Mollon, 2009; Henton y Midgley, 2012). Además de esto, tal y como ya se ha señalado anteriormente, no es correcto asumir la generalización (validez externa) de estos estudios directamente, ya que es necesario comprobar su transportabilidad desde el entorno experimental artificial al ámbito clínico natural, objetivo también reconocido desde el ámbito de los TPBE (Crits-Christophy Gallop 2006; Shafran, 2011).

Ante estas limitaciones y necesidades, el enfoque de investigación denominado “Evidencia basada en la práctica”, se plantea como una línea que complementa a la investigación experimental, pudiendo generar de esta forma una base de evidencia con una imagen más completa de la psicoterapia. Hace referencia a los estudios no experimentales, realizadas en colaboración con profesionales no vinculados al entorno académico en el contexto de la investigación naturalística de la práctica clínica (Barkham, Hardy y Mellor-Clark, 2010). Henton (2012) señala que los principales paradigmas de actuación de esta perspectiva son: los estudios de caso único no experimentales (incluyendo los estudios de procesos) y las investigaciones de efectividad.

Sobre los diseños de caso único, Roussos (2007) declara que han vuelto a despertar un gran interés en el contexto de la investigación de la psicología clínica en los últimos años (en el pasado ya tuvieron protagonismo en el inicio de esta disciplina) y que representan un medio con gran potencialidad para aunar práctica clínica e investigación. En la Tabla 7 se pueden observar las ventajas y limitaciones destacadas por este autor respecto a este tipo de diseños. Otros autores también destacan su potencialidad para abordar directamente cuestiones relativas a: los resultados terapéuticos, a generar y probar teorías y a sistematizar análisis fenomenológicos de la experiencia. (Flyvbjerg, 2006, Barkham, Hardy y Mellor-Clark, 2010). En la última década se han generado varias propuestas para aprovechar los datos provenientes de este tipo de estudios

tales como el desarrollo de sistemas computacionales específicos para gestionar los datos clínicos recogidos (Lueger, 2006) o el planteamiento de modelos de sistematización para permitir la comparación de información proveniente de diferentes estudios (Fishman, 2005; Virués-Ortega y Moreno-Rodríguez 2008).

**Tabla 7**

*Ventajas y limitaciones de los diseños de caso único en investigación de resultados en psicoterapia*

Ventajas	Limitaciones
- Capacidad de trabajo en <i>situaciones atípicas</i> (muestras clínicas de difícil acceso o problemáticas poco comunes).	- La <i>generalización</i> de conclusiones se puede problematizar a pesar de contar con múltiples observaciones.
- <i>Adecuación ética- legal</i> . Más fácil de cumplir al tratarse de un solo individuo y no existir un grupo control.	- La <i>variabilidad natural de cada sujeto</i> no se controla y compensa como en el caso de estudios con grupos.
- Posibilita <i>estudiar</i> las relaciones de <i>multi-causalidad</i> en detalle. Y recoger <i>múltiples medidas</i> del proceso terapéutico.	- El <i>sesgo de selección</i> del sujeto por parte del evaluador puede distorsionar los resultados.
- Facilidad para proporcionar argumentos contra facticos <i>con excepciones a teorías generales</i> .	- Difícil control de <i>explicaciones alternativas a cambio</i> del sujeto (validez interna). Necesario exhaustivo registro de variables.
- Generar <i>nuevas hipótesis</i> y convertir en <i>ideas concretas</i> principios teóricos abstractos de forma más fluida.	

Nota: Tabla generada a partir del contenido de Roussos (2007).

Kadzin (2001) señala tres aspectos metodológicos fundamentales a tener en cuenta a la hora de realizar estudios de caso único:

- *Evaluación continua*: son necesarias observaciones repetidas y confiables para asegurar que se genera un cambio en base a los factores estudiados (de tratamiento o no). Es deseable una monitorización con múltiples medidas de los mismos indicadores a lo largo del proceso.
- *Evaluación de la línea base*: las observaciones y medidas de indicadores deben comenzar antes de que la intervención comience. Con el fin de describir el estado real inicial de

sujeto y predecir el comportamiento futuro del paciente independientemente del tratamiento. Para poder así delimitar correctamente el impacto de la intervención.

- *Estabilidad del rendimiento*: se debe comprobar si las medidas pretratamiento son estables y no presentan gran variabilidad o una direccionalidad determinada. Ya que si esto se da, la obtención de conclusiones acerca del tratamiento se verá seriamente limitada. Si conocemos previamente la variabilidad de los datos previa a la intervención, mejor se podrán diferenciar las asociaciones espurias en los cambios del paciente.

Por su parte, los estudios de efectividad poseen un gran potencial para proporcionar información, en áreas que los estudios experimentales controlados tienen dificultad para abordar: rentabilidad real de la intervención, influencia de factores organizacionales o del servicio e implicación de diferentes variables moderadoras de la clínica del día a día (Henton, 2012). El conocimiento de cómo funciona la terapia en la clínica habitual puede ayudar a optimizar los tratamientos y proporcionar información de cómo ajustar las intervenciones a las necesidades individuales y dar información que sirva para la gestión de los servicios. Por ello, desde la perspectiva de eficacia de tratamientos se plantea la necesidad de implementar sistemas de seguimiento de resultados en la rutina de la clínica de los servicios de salud (Werbart, Levin, Andersson y Sandell, 2013).

En coherencia con ello se han desarrollado distintas iniciativas de este tipo de sistemas tales como: el Treatment Outcome Package en Estados Unidos (Kraus, Seligman y Jordan, 2005), el Rotune Outcome Monitoring en los Países Bajos (de Beurs et al, 2011), el Quality Assurance of Psychotherapy en Suecia (Werbart, Levin, Andersson y Sandell, 2013) o el Clinical Outcomes in Routine Evaluation en Reino Unido (Stiles, Barkham, Mellor-Clark y Connell, 2008).

Otra iniciativa para abordar el estudio de la psicoterapia en ámbitos naturales es la descrita por Castoguy (2011) como “Redes de Investigación de la Práctica” (PRN - Practice Research Networks), que se apoya en la colaboración activa entre investigadores y clínicos en el desarrollo de estudios clínicamente relevantes y científicamente rigurosos con el fin de romper la estructura unidireccional de la investigación descrita al inicio de este apartado. Esta colaboración conlleva la participación de los clínicos en la selección de temas a ser investigados, delineación de hipótesis a ser probadas, la construcción y la aplicación de diseño de la investigación, así

como difusión de los resultados. Suele implicar además una puesta en común de los datos acerca de resultados de los servicios locales mediante una medición estandarizada (Van Weel, 2002). Referentes en psicología de este tipo de redes (Parry, Castonguay, Borkovec y Wolf, 2010) son: la denominada “SPR-UK North PRN” en Reino Unido vinculada a las universidades de Sheffield y Leeds (Barkham, Hardy y Shapiro, 2011) y la “Penn State PRN” en Estados Unidos coordinada por la Universidad estatal de Pennsylvania (Castonguay et al., 2010).

#### **4.3.3. Medir resultados en la práctica clínica habitual**

En el marco de trabajo planteado en la “Practica Psicológica Basada en la Evidencia” PPBE, se destaca el papel de la evaluación como un aspecto esencial de cualquier intervención. No se trata de la fase inicial del proceso, sino que es un aspecto continuado que ayuda a tomar dediciones sobre el curso del tratamiento. Monitorizar la evolución del paciente a través de sus cambios en diversos indicadores de resultado, es imprescindible para poder adecuar la terapia a sus necesidades particulares. Independientemente de su función para generar evidencias científicas, la evaluación de resultados es una herramienta para el proceder profesional diario. Y está en el modelo de la PPBE debe realizar de un modo sistemático empleando medidas estandarizas, fiables y validas que complementen el juicio clínico, ya que de esta forma se obtienen reflejo estable, verificable y replicable del estado del paciente que de otra forma no se podría conseguir (Kadzin, 2008, Rivera y Bernal, 2008; Estupiñá, 2012). Esta evaluación sistemática, permite superar ciertas limitaciones documentadas habituales en la observación del estado del paciente por parte del profesional tales como: sesgos de memoria, atención selectiva, heurísticos de representatividad etc... (Garb y Boyle, 2003).

Los diversos estudios y revisiones de Lambert (2010) acerca del efecto del feedback en el proceso terapéutico, representan otro argumento adicional para el uso de la evaluación de resultados en la práctica habitual. Este autor constató que el hacer explícito tanto al terapeuta como al paciente la evolución del estado de este último (a través de indicadores relevantes para la consecución de los objetivos terapéuticos), no solo mejoran los resultados finales sino que se optimiza el proceso terapéutico, se hace más eficiente en número de sesiones y se aumenta la adherencia al tratamiento, independientemente del modelo o estrategias de intervención del terapeuta. La evaluación de resultados en este contexto representa en sí misma un facilitador del cambio terapéutico.

Con el fin de evaluar el cambio de estado en el proceso terapéutico de una forma fiable, se han generado diversos instrumentos que permiten obtener al profesional o al académico diferentes dimensiones de resultado. Es importante para el clínico seleccionar un instrumento adecuado y breve para su utilización en la práctica habitual (Lambert y Hawkins 2004). A continuación se destacan los instrumentos genéricos de autoinforme que más atención han recibido en la literatura de evaluación de resultados en psicoterapia (Beutler, 2001; Corbella y Botella, 2004; Biescad y Timulak, 2014):

- Cuestionario OQ-45 (Outpatient Questionnaire 45): Lo componen 45 ítems en escala Likert de cinco puntos que se agrupan en tres áreas de resultado: sintomatología, relaciones interpersonales y rol social (Lambert, Hansen, Umpress, Lunnen, Okiishi. y Burlingame 1996; Lambert et al. 1996).
- Cuestionario CORE (Clinical Outcomes in Routine Evaluation): Lo compone 34 ítems en escala Likert de cuatro puntos, que reporta medidas acerca de 4 dimensiones: bienestar subjetivo, problemas/síntomas, funcionamiento social y riesgo de hetero o auto agresión. (Barkham, et al. 1998; Feixa et al., 2012).
- Cuestionario SCL-90-R (Symptom Checklist 90 Revised): Lo componen 90 ítems en escala Likert de 5 puntos que se agrupan en 9 escalas sintomáticas específicas y en la que se pueden calcular 3 índices de gravedad sintomática general. (Derogatis y Cleary, 1977; Derogatis, 2002).

A la hora de analizar el cambio producido en una intervención psicológica se pueden tratar de responder tres preguntas diferentes a partir de los mismos resultados (Kirk, 2001):

- 1.- ¿El resultado encontrado es realmente un cambio en el sujeto o se debe al azar y la variabilidad natural de sus puntuaciones?
- 2.- ¿Ese efecto encontrado qué magnitud tiene?
- 3.- ¿Ese efecto es importante y relevante para el paciente?

Según la revisión de Faulkner, Fidler, y Cumming (2008) muchos estudios experimentales controlados valoran la eficacia de las intervenciones limitándose a responder únicamente a la primera pregunta, que se corresponde con las pruebas estadísticas de significación clásicas. Lo que representa una seria distorsión de las conclusiones a la que se puede llegar y resta mucha

utilidad real a este tipo de estudios. Kadzin (2008) señala que la respuesta a estas tres preguntas implican la comprobación de: la significación estadística, el tamaño del efecto y la significación clínica de los resultados encontrados. Siendo la comprobación de esta última de importancia capital el profesional en su desempeño habitual. Sobre el desarrollo metodológico de estos conceptos versará el siguiente capítulo así como la propuesta concreta del presente trabajo.

## Capítulo 5

### Paradigmas estadísticos en la evaluación de resultados clínicos

---

En el capítulo desarrollado a continuación se exponen los principales medios estadísticos de análisis de resultados clínicos para la constatación de cambio terapéutico en el paciente. Se abordarán los significados e implicaciones específicas de la significación estadística, la magnitud del efecto y la significación clínica. Finalmente, se expondrá la metodología basada en indicadores fiables de cambio (y las diferentes alternativas propuestas a la misma) que representa el eje principal de la presente tesis doctoral.

#### **5.1. De lo estadísticamente significativo a lo clínicamente relevante**

##### **5.1.1. Significación estadística**

La investigación en ciencias de la salud implica estudiar la relación entre los diferentes elementos de una muestra, para poder alcanzar una serie de conclusiones sobre esas mismas relaciones en la población de la cual proviene la muestra a estudio. En el caso de los estudios de resultados clínicos, esto supone generalizar los resultados de unas condiciones de intervención concretas en una muestra de pacientes específica a la totalidad de pacientes que presenten la misma problemática. Las condiciones de intervención normalmente son los elementos manipulados experimentalmente, configurando las variables independientes del investigador mientras que los indicadores de resultados o efectos recogidos en el estudio, cobran el rol de variables dependientes. Basándose en los datos del estudio, el experimentador debe establecer en qué medida existe una relación entre los diferentes valores de la variable independiente y los valores de las variables dependientes. Como ya se presentó anteriormente, en el contexto empírico esta relación puede verse enmascarada o distorsionada por dos tipos de fuentes de error: los errores sistemáticos o sesgos y los errores variables o aleatorios. Los primeros son variables de influencia constante que deben tenerse en cuenta y controlarse al diseñar y ejecutar la investigación (ver capítulo 2). Sin embargo, los segundos son los generados por variables

ajenas a la investigación y cuya influencia es aleatoria y difícil de prever. Esta última fuente de error se encuentra relacionada con la propia variabilidad del proceso de muestreo, y la natural de las variables dependientes y su medición. En este sentido, los errores aleatorios dificultan la generalización de los resultados de los estudios. La estadística analítica, o también denominada inferencial, se presenta como una herramienta práctica para que el investigador pueda afrontar este último hecho, permitiéndole determinar en qué grado la relación entre variables observadas se puede deber al azar, y/o a los valores de la variable independiente (Álvarez, 2007; Box, Hunter y Hunter, 2008; Arignón y Jiménez, 2013).

Con respecto a este uso de la estadística, las propuestas que mayor impacto han tenido en la comunidad científica han sido: el proceso de inferencia inductiva a través de pruebas de significación de Fisher (1925a) y el razonamiento hipotético-deductivo mediante pruebas de hipótesis de Neyman y Pearson (1928a). Frecuentemente ambas posiciones, teóricamente enfrentadas, se han presentado combinadas en los manuales de estadística bajo el epígrafe del paradigma del contraste de hipótesis nula en una metodología híbrida, siendo este formato el utilizado de forma rutinaria en la investigación y en las publicaciones científicas (Rodríguez, 2005).

Fisher (1925a; 1991) propone la prueba de significación, como una herramienta para aportar argumentos en un proceso de inferencia inductiva (de lo particular a lo general). Se utiliza para valorar la credibilidad de una hipótesis concreta a partir de unos datos empíricos. La prueba en sí implica evaluar la verisimilitud de observar en una población de referencia teórica los datos descubiertos en el estudio, donde la hipótesis del experimentador no se cumple. Por ejemplo, constatar en qué grado la diferencia de las medias en un indicador de resultado entre dos grupos de tratamiento de un estudio, es probable que exista en una distribución de referencia teórica donde la media de diferencias es 0 y por tanto la diferencia observada en el estudio se pueda deber a la variabilidad del muestreo y no al tratamiento. Fisher designa esta probabilidad como grado de significación o valor  $p$ , y establece que el criterio a partir del cual se puede considerar ese valor  $p$  bajo se debe haber establecido antes del estudio. Aunque el autor considera este criterio flexible y no necesariamente fijo a todas las situaciones experimentales, realizó la mención al valor de  $p > 0,05$  como sugerencia. Si  $p$  es menor que el criterio establecido, el experimentador podrá descartar que los resultados observados se deban al azar, sino no podrá

hacerlo y será necesario seguir investigando. Fisher advierte que el valor de significación no puede ser utilizado como una inferencia de la probabilidad de error en caso de repetir el estudio. Este procedimiento se puede concretar en cuatro pasos (Iraugui, 2009a):

- 1º- Formular una hipótesis nula. Se plantea una hipótesis donde no existe relación entre las variables del estudio. Es decir una hipótesis contraria a la del experimentador donde las diferencias observadas se deban al azar y no a la relación entre variables. Esta es la hipótesis que se operativizará y pondrá a prueba.
- 2º- Fijar el nivel de significancia. Es necesario fijar un umbral a partir del cual el valor p, resultante de poner a prueba la hipótesis nula, permita decidir si esta hipótesis de azar es creíble o no. Los valores por debajo del umbral se tienen que tomar como evidencias en contra de la hipótesis nula pero no es un criterio absoluto. Este umbral es la magnitud del riesgo que está dispuesto a asumir el experimentador de cometer el error de rechazar la hipótesis nula si es verdadera.
- 3º- Elegir una prueba estadística apropiada a la naturaleza de las variables contrastadas. Según se puedan cumplir o no unos supuestos acerca de la distribución poblacional se tratará de una prueba paramétrica o no paramétrica.
- 4º- Comparar el valor p resultante de la prueba con el nivel de significancia preestablecido y tomar una decisión de poder rechazar o no la hipótesis nula (pero no para aceptarla). Bajo la lógica de Fisher no se consideran más opciones ya que el proceso solo sirve para obtener argumentos contra la hipótesis nula.

A partir de la propuesta inicial Fisher, pero bajo otra perspectiva, Neyman y Pearson (1928a, 1928b, 1933) desarrollaron su propio procedimiento de contrastación de hipótesis. Estos autores parten de la teoría de la probabilidad, para establecer la prueba como una regla de decisión entre dos hipótesis complementarias (rechazar una implica aceptar la otra), siendo esta consideración del contraste en lo difieren fundamentalmente con Fisher. Para ello es necesario plantear las dos hipótesis: por un lado la hipótesis nula ( $H_0$ ), que como en el caso de Fisher plantea la no relación real entre las variables estudiadas, y por otro, su suceso complementario a nivel de probabilidad, la Hipótesis Alternativa ( $H_1$ ), que se expresa como lo contraria a la  $H_0$ . En esta situación todos los valores muestrales posibles pertenecen a uno de los dos subconjuntos

que forman las hipótesis a contrastar. Se debe determinar a cuál de los dos subconjuntos pertenece la muestra del estudio. Por lo que se generan dos regiones de aceptación de cada una de las hipótesis. La región de aceptación de  $H_1$  se denomina región crítica.

La hipótesis alternativa no se somete directamente a la prueba de significación sino que se realiza con la  $H_0$  ya que es la que se puede operativizar y establecer matemáticamente, pero dada su complementariedad los resultados obtenidos permiten tomar la decisión de aceptar una y rechazar la otra. Si el valor del estadístico de contraste se sitúa dentro de la región crítica se acepta  $H_1$  y se rechaza  $H_0$  y si se sitúa fuera se realiza la decisión inversa. Sin embargo, esto puede implicar asumir dos errores diferentes según la decisión que se tome: el de rechazar la hipótesis nula cuando esta es verdadera, denominado Error tipo I con la probabilidad de cometerlo *riesgo*  $\alpha$  o el de aceptar la hipótesis nula cuando la hipótesis alternativa es verdadera Error tipo II con la probabilidad de cometerlo *riesgo*  $\beta$ . En la Tabla 8 se ilustra lo que suponen estos elementos en la decisión de la prueba. Los riesgos  $\alpha$  y  $\beta$  deben haberse establecido a priori según la naturaleza del estudio y la consecuencias prácticas de cometer cada error (Peña, 2000; Lenhard ,2006).

**Tabla 8**

*Errores y riesgos asociados a la prueba de hipótesis de Neyman y Pearson*

DECISIÓN DE PRUEBA	HIPÓTESIS VERDADERA	
	H0	H1
Aceptar H0	<b>Decisión correcta</b> Pr = $1 - \alpha$ <i>Verdadero negativo</i>	<b>Error tipo II</b> Pr = $\beta$ <i>Falso negativo</i>
Aceptar H1	<b>Error Tipo I</b> Pr = $\alpha$ <i>Falso positivo</i>	<b>Decisión correcta</b> Pr = $1 - \beta$ <i>Verdadero positivo</i>

Nota: Generada a partir del contenido de Peña (2000) y Lenhard (2006).

Doménech (2013b) señala que debido a que tanto el planteamiento de Fisher como el de Neyman y Pearson implican ambos la puesta a prueba de la hipótesis nula mediante los mismos estadísticos y la comparación de resultados con un criterio de decisión (ya sea valor  $p$  o región crítica) es fácil cometer el error de suponer que en el fondo tratan de la misma metodología pero utilizando terminología diferente (mantienen concordancia algebraica). Respecto a esto la Tabla 9 se presentan las diferencias entre los dos procedimientos destacadas por el autor.

**Tabla 9**

*Diferencias entre prueba de significación de Fisher y prueba de hipótesis de Neyman y Pearson*

<b>Prueba de significación ( Fisher)</b>	<b>Prueba de Hipótesis (Neyman y Pearson)</b>
- Solo se plantea $H_0$ .	- Se plantean dos hipótesis complementarias $H_0$ y $H_1$ .
- Vía de decisión= Valor $p$ . Establecer previamente umbral Significativo / no significativo.	- Vía de decisión= Región crítica. Regla de decisión entre $H_0$ y $H_1$ . Aceptar una rechaza la otra.
- Valor $p$ no se interpreta como probabilidad sino como indicador de credibilidad $H_0$ .	- Riesgos $\alpha$ y $\beta$ son probabilidad de cometer errores en la decisión. Error tipo I o Error tipo II.
- No se concluye que se acepta $H_0$ , solo que se puede rechazar o no rechazar según los datos del estudio.	- Necesita de muestreo aleatorio y se basa en la teoría de la probabilidad.
- No necesita de supuesto de muestreo aleatorio.	- La lógica del proceso es contraria a la inferencia inductiva.
- Valor $p$ no es una estimación de probabilidad al repetir el experimento.	- Los riesgos $\alpha$ y $\beta$ representan frecuencias de errores si se repetiría el estudio muchas veces.
- Valor $p$ sólo se asocia a un único estudio.	- Para limitar el riesgo $\beta$ se calcula previamente el número de sujetos del estudio según se fije $\alpha$ , $\beta$ y la magnitud de efecto que debe detectar la prueba.
- El valor $p$ se usa de forma flexible junto a otras informaciones en el proceso de inferencia inductiva.	

Nota: Generada a partir del contenido de Doménech (2013b).

Para Nuzzo (2014) la confusión entre ambos planteamientos ha llevado a generalizar en la comunidad científica una serie de interpretaciones erróneas acerca del significado y utilidad de la significación estadística de las pruebas de contraste de hipótesis nula. Esta autora considera además que se ha sobreestimado la importancia del valor  $p$  y su relevancia para la investigación en general, llevándose a obviar en demasiadas ocasiones otros posibles indicadores de interés como los referentes a la magnitud del efecto encontrado. De esta manera se desvirtúa la propuesta de Fisher de usar la prueba de significación como una herramienta más en el proceso

de descubrimiento científico y no como el procedimiento definitivo para validar hipótesis. En coherencia con esto, Rebas (2003) recuerda a los clínicos cuatro precauciones que se deben tener en cuenta a la hora de interpretar el resultado de una prueba de significación:

- El rechazo de una hipótesis nula no sugiere causalidad. Esto como se ha señalado anteriormente se consigue a través del diseño del estudio, asegurando la validez interna de las conclusiones de relación de variables.
- Un resultado no significativo no demuestra por sí solo que la hipótesis nula sea cierta. Poner a prueba la hipótesis nula posibilita al investigador valorar si los datos permiten dudar de la credibilidad de esta hipótesis o no son suficientes para descartarla.
- El valor  $p$  no es una medida de magnitud del efecto. Un resultado muy significativo hace referencia a la poca credibilidad de la hipótesis nula y por tanto a que el efecto encontrado no se puede achacar al azar, pero no es una medida de la cuantía de ese efecto.
- El resultado estadísticamente significativo no implica relevancia clínica. En relación con lo anterior la significación estadística tampoco informa de la importancia que puede tener ese efecto encontrado para la toma de decisiones clínicas.

Otro elemento relevante a la hora de interpretar los resultados de las pruebas de significación estadística es ser consciente de su dependencia del tamaño muestral. El número de sujetos en un estudio afecta a la significación estadística a través del error estándar. El error estándar se entiende como la desviación estándar de un estadístico en todas las posibles muestras de un tamaño determinado escogidas de una población (Everitt y Skrondal, 2010). Como este dato suele ser desconocido se estima a partir de la muestra de datos que se está estudiando. En su cálculo en el denominador siempre va a figurar el número de sujetos, por lo que el error estándar será más pequeño cuanto mayor sea la muestra independientemente de otros factores. Esta estimación es la utilizada en las pruebas estadísticas de contraste de hipótesis nula como indicador de error muestral. Esto provoca que cualquier diferencia entre variables pueda llegar a ser estadísticamente significativa (por pequeña que sea), si se cuenta con el número suficiente de efectivos en la muestra y que, por el contrario, efectos importantes pasen desapercibidos cuando se dispone de pocos sujetos en el estudio (Álvarez, 2007; Manterola y Pineda, 2008).

Las dos estrategias presentadas acerca del contraste de hipótesis estadísticas (Fisher y Neyman-Pearson), se engloban en el denominado enfoque clásico o frecuentista de la probabilidad, donde se utiliza únicamente la información proveniente del estudio y donde se fijan criterios de decisión a priori que permanecen inmutables en el proceso. Esta es la perspectiva más común a la hora de realizar inferencias estadísticas. Sin embargo, existe una alternativa a este enfoque que pese a no tener demasiada aceptación e impacto en el ámbito científico general está cobrando fuerza en los últimos años. Se trata del enfoque Bayesiano, denominado así por basarse en los planteamientos matemáticos de cálculo de probabilidades condicionadas del Teorema de Bayes. En este método también establece tanto  $H_0$  como  $H_1$  pero ninguna tiene una función especial como es el caso de  $H_0$  en el enfoque frecuentista. La inferencia bayesiana plantea el contraste de hipótesis entre diferentes alternativas, a través del mencionado Teorema para ponderar las evidencias a favor de una u otra hipótesis. Es necesario disponer de una información de la probabilidad previa (al margen de los datos del propio estudio) de las diferentes alternativas en el sentido de probabilidad subjetiva, la cual viene a representar la certidumbre del observador sobre la plausibilidad de la ocurrencia de cada suceso según los conocimientos acumulados hasta ese momento. La posibilidad de incluir conocimientos previos en el análisis es la aportación más destacada de este enfoque, ya que permite realizar un proceso de inferencia más natural y acumulativo (se van ajustando las probabilidades según las evidencias de los diferentes estudios). Pero también es su mayor debilidad al no existir una fuente fiable acerca de cómo establecer las probabilidades a priori en todas las situaciones, existiendo el peligro de obtener conclusiones sesgadas (Christiansen, 2005; Alamilla-López y Jiménez, 2010; Monterey, 2012; Doménech, 2013b).

### **5.1.2. Magnitud y relevancia de los efectos**

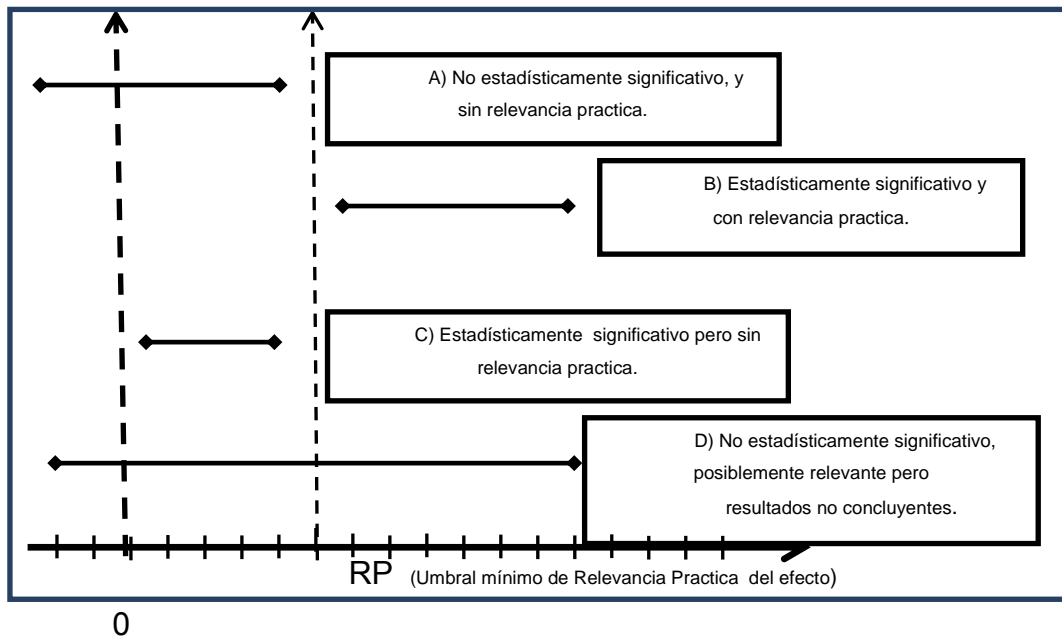
Desde el ámbito académico se ha advertido en numerosas ocasiones que limitar la validación de hipótesis científicas a únicamente realizar y reportar pruebas de significación de hipótesis nula (paradigma comúnmente denominado como Null Hypothesis Significance Testing, NHST), es una práctica incompleta e incluso inadecuada para poder llegar a conclusiones útiles en el avance práctico de una disciplina (Cohen, 1994; Schimdt, 1996; Murphy, 1997; Ioannidis, 2005; Lambdin, 2012; Nuzzo, 2014).

Con respecto a la investigación clínica, la NHST se muestra insuficiente para dar respuesta a las necesidades planteadas en este tipo de estudios (descartar que las diferencias encontradas se deban al azar no es información suficiente para la toma de decisiones), donde lo interesante es la valoración del efecto del tratamiento y la obtención de indicadores para considerar la relevancia sustantiva del cambio terapéutico (Howlin, 1997; Frías, Pascual y García, 2002). Sin embargo, el reporte exclusivo de NHST ha seguido siendo una práctica recurrente en una proporción mayoritaria de la investigación en psicología y sólo en los últimos años está empezando a cambiar, fundamentalmente debido a la presión de algunos editores de las revistas científicas y a las asociaciones profesionales (Vacha-Haase y Nilsson 1998; y Vacha-Haase Ness 1999; Faulkner, Fidler, y Cumming 2008; Astrid, Scherndl y Kühberger, 2012).

Debido a las controversias generadas con el uso del NHST en la investigación, la APA creó un grupo de trabajo específico para que realizase una serie de recomendaciones en referencia al análisis estadístico de los datos y su interpretación, y así mejorar la calidad de la información utilizada en los estudios de psicología. Se destacaron dos sugerencias para complementar al NHST y superar sus limitaciones: calcular los intervalos de confianza alrededor de los diferentes estimadores del estudio y presentar indicadores de los tamaños de efecto de los resultados (Wilkinson y APA-Task Force on Statistical Inference, 1999). En sucesivas publicaciones al respecto, dicha asociación ha seguido enfatizando la idoneidad de presentar estos dos elementos, declarando que deberían ser un mínimo exigible en cualquier publicación científica (APA, 2008; 2010). Requisitos que algunos editores han empezado a incorporar en sus revistas o por lo menos a señalar su importancia (Fidler y Thompson, 2001; Vacha-Haase y Thompson, 2004; Carretero-Dios y Pérez, 2007). Desde el ámbito de la medicina también se han subrayado la conveniencia de su inclusión en cualquier informe de resultados de investigación clínica (Schulz, Altman y Moher, 2010; International Committee of Medical Journal Editors, 2013). Por su parte, el Centro Nacional para la Educación estadística de Estados Unidos (NCES, 2002) señala que una vez se han encontrado resultados estadísticamente significativos, es útil generar indicadores de la magnitud del efecto de las diferencias halladas en el estudio, para así poder permitir una interpretación sustantiva del efecto (valorar su importancia).

El intervalo de confianza (IC) es un rango de valores calculados por métodos estadísticos, entre los cuales se espera se encuentre el verdadero valor de un parámetro (valor en la población), que se está tratando de estimar a partir de una muestra (por ejemplo, una media o una diferencia entre dos medias), con una probabilidad definida de antemano, denominada nivel de confianza y representada como  $1-\alpha$ , siendo el nivel de significancia  $\alpha$  la probabilidad de equivocarse en esta estimación. Normalmente se elige un nivel de significación del 95% ( $\alpha=0,05$ ), por lo que el intervalo incluirá el valor verdadero 95 de cada 100 veces. Su uso en investigación clínica tiene la ventaja, respecto a los valores p, de reflejar los resultados en la misma escala en la que los datos han sido medidos e indicar, además, la dirección del efecto observado. Si se puede suponer normalidad de la distribución del parámetro en la población, el intervalo se calculará sumando y restando al efecto resultante del estudio su error estándar multiplicado por el valor Z correspondiente al nivel de confianza. Debido a esto el intervalo será más estrecho (y por lo tanto más preciso), cuanto más grande sea la muestra y más pequeña la dispersión de las puntuaciones dentro de la misma. Los IC son estimadores de la magnitud y de la precisión del efecto encontrado, y útiles a la hora extrapolar los resultados del estudio a la práctica clínica. Gracias a sus características matemáticas también se pueden utilizar como medidas de significación, observando si incluyen dentro del intervalo resultante el valor correspondiente a una situación de no significación (por ejemplo, 0 en el caso de las diferencia de medias o 1 en el de la razón de riesgos). Finalmente, si se puede ubicar un umbral previo de importancia del efecto estudiado, el IC puede ayudar a valorar la relevancia práctica de los resultados encontrados (Du Prel, Hommel, Röhrig y Blettner, 2009; Balluerka y Molina, 2013).

En relación a lo anterior, en la Figura 2 se expone un ejemplo grafico de las situaciones posibles en un estudio de comparación de grupo por diferencia de medias donde la relevancia clínica se establece al superar un nivel predeterminado (Du Prel, Hommel, Röhrig y Blettner, 2009).



**Figura 2.** *Uso Intervalos de Confianza para valorar relevancia práctica de resultados.*

Adaptado y modificado de Du Prel, Hommel, Röhrig y Blettner (2009, p. 338).

El constructo de tamaño del efecto (effect size) se ha establecido en relación al NHST como el grado en el que la  $H_0$  no se cumple, es decir la cuantificación de en qué medida los datos del estudio se separan de las especificaciones de  $H_0$  (Vacha-Haase y Thompson, 2004). También se ha conceptualizado como la medida de la fuerza de relación entre dos variables, normalmente entre una variable independiente, como el tipo de tratamiento, y otra, una variable dependiente como una medida de resultado (Vaske, Gliner y Morgan, 2002). De forma más específica se le ha caracterizado como la magnitud estandarizada o normalizada del efecto encontrado en el estudio (NCES, 2002, Olejnik y Algina, 2003).

Para Preacher y Kelley (2011; 2012) existen numerosas definiciones de tamaño del efecto en la literatura acertadas para el contexto en el que son realizadas, pero demasiado restrictivas para resultar adecuadas como definición genérica. Estos autores lo caracterizan de forma más inclusiva como: “*un reflejo cuantitativo de la magnitud de un fenómeno que se utiliza para el propósito de hacer frente a una cuestión de interés*” (Preacher y Kelley, 2012 p.140). Vinculan de esta forma el índice estadístico calculado al problema de investigación (se hace explícito la

intencionalidad de su cálculo para afrontar preguntas) y lo presentan de forma separada a la lógica NHST.

Por su parte, Nakagawa y Cuthill (2007) señalan tres connotaciones diferentes del uso del término tamaño del efecto:

- Índice estadístico que estima la magnitud de un efecto (por ejemplo, la  $d$  de Cohen).
- El valor real del índice estadístico calculado a partir de la muestra del estudio (por ejemplo,  $d=0,8$ ).
- La interpretación (cualitativa) de la relevancia del efecto a partir el valor del índice resultante (por ejemplo, efecto importante).

Sobre el uso específico de los diferentes índices estadísticos de tamaño del efecto y el cálculo de sus valores, se profundizará en el siguiente apartado. Se puede destacar de forma genérica la ventaja que implica utilizar índices estandarizados (con una escala común independiente de las unidades de medida del estudio y del tamaño de la muestra) y la posible transformación matemática de unos a otros. Esto permite realizar una comparación directa de resultados de diferentes estudios (siempre que mantengan un diseño semejante) y la puesta en común acumulativa de evidencia acerca de los efectos de un tratamiento o condición. Tales características les hace indispensables para la realización de estudios metaanalíticos (Gonzalez-Ramírez y Botella 2006).

Con respecto a la valoración de la relevancia del efecto, Preacher y Kelley (2012) indican que resulta una labor compleja y difícil de objetivizar. Para estos autores se vincula directamente con la idea de “trascendencia sustantiva”, entendida como el grado en el cual las partes interesadas (científicos, profesionales, consumidores, gestores, pagadores...) considerarían que un efecto sea importante, digno de atención y con repercusiones para la toma de decisiones. Su significado es más subjetivo que la significación estadística y depende en gran medida del contexto y las circunstancias específicas del fenómeno estudiado. Mientras que en una situación dada un determinado tamaño del efecto puede resultar trivial, en otra, esa misma magnitud puede tener un gran impacto. Por ello, señalan que el tamaño del efecto no es sinónimo de relevancia práctica o sustantiva, aunque el conocimiento del tamaño del efecto ayuda a poder realizar un

juicio sobre la importancia práctica, así que su reporte como complemento a la significación estadística es tan recomendado.

En el ámbito asistencial, la mencionada relevancia sustantiva de los resultados es denominada significancia clínica (Kendall, Butcher y Holmbeck, 1999). Desde el punto de vista médico la valoración del tamaño del efecto de los resultados, la naturaleza del problema a tratar y las implicaciones de la aplicación de la intervención a estudio (coste, duración, efectos secundarios...) definirán esta significación. Por su parte, el paciente ubicará la importancia clínica en la resolución de los síntomas disruptivos (en “curarse”), en su calidad de vida y la satisfacción con la atención recibida (Fernández y Díaz, 2001; Álvarez-Cienfuegos et al., 2012). Ochoa (2010) señala que no existe un consenso a la hora delimitar su definición y componentes. Sin embargo, a partir de la revisión que realiza en la literatura médica compila una serie de implicaciones a tener en cuenta al identificar la significación clínica de los resultados de un estudio:

- Debe hacer referencia a una variable de resultado válida y con sentido clínico.
- Debe estar valorada, al menos, por alguna de las partes implicadas en el proceso asistencial.
- Supone un beneficio para el paciente, la aplicación es viable en el entorno real y aceptable en términos de coste-beneficio.
- El efecto encontrado ha alcanzado significación estadística en un diseño que controle los posibles sesgos.
- Debe ser previamente definida en consenso entre los principales actores del desempeño profesional.
- Cuanto mayor sea el tamaño del efecto, más verosímil resultará.
- Es diferente para cada situación, por lo que se adapta a las circunstancias.
- Depende en gran medida de la perspectiva desde la que se juzgue (clínicos, pacientes, investigadores, agencia financiadoras etc...).

En psicología, la determinación de la significación clínica de las intervenciones se ha ubicado en constatar si estas generan un cambio beneficioso, importante y evidente en el día a día de la vida del paciente y de quienes interactúan con él. Por lo que una intervención clínicamente importante efectuará un cambio en el paciente que se pueda apreciar de forma palpable en su estado vital (Kazdin, 1999, Ogles, Lunnen y Bonesteel, 2001). Esta perspectiva posee una connotación de valoración individual que ha llevado a Bothe y Richardson (2011) a sugerir distinguir dos términos diferentes para hacer referencia a la importancia clínica. Por un lado, la significación clínica, valorada a partir de medidas grupales seleccionadas e interpretadas por los profesionales. Y, por otro, la significación personal que hace referencia a la determinación de relevancia a partir de las variables, medidas y cambios que son de valor para los pacientes individuales.

En referencia a lo anterior, Kazdin (2001) indica que los métodos de evaluación habitualmente utilizados en los ensayos clínicos no capturan la relevancia del impacto de la intervención sobre el individuo. Incluso la consideración del tamaño del efecto es insuficiente: la medida seleccionada puede tener poco valor práctico para el paciente y el cambio individual puede verse enmascarado por las medidas promedio del grupo. En la Tabla 10 se recogen las principales estrategias propuestas (en la investigación en psicología) para evaluar significatividad clínica respecto al individuo, según la revisión de este autor. Resalta, además, la necesidad de seguir profundizando y reflexionando sobre si los indicadores utilizados para valorar este constructo contribuyen realmente a generar cambios en el ajuste diario del paciente, y aconseja utilizar más de un indicador en su determinación.

**Tabla 10***Medios para establecer significación clínica del paciente en psicología*

<b>Método</b>	<b>Definición</b>	<b>Criterios y Medidas</b>
<i>Método de Comparación</i>	<p>Se evalúan los resultados del paciente en relación con:</p> <ul style="list-style-type: none"> <li>- Los desempeños de otros, por ejemplo con muestra normativa o con muestra del paciente (métodos normativos).</li> <li>- Con las propias medidas del paciente (métodos ipsativos).</li> <li>- Con criterios absolutos preestablecidos como cumplimiento de criterios diagnósticos, presencia de síntomas.</li> </ul>	<ol style="list-style-type: none"> <li>1.- La similitud con muestras normativas en el final del tratamiento.</li> <li>2.- Desviación estadística sobre una medida de las puntuaciones de una muestra.</li> <li>3.- Cantidad de cambio en pre- post.</li> <li>4.- No cumplir ya criterios de trastorno psiquiátrico.</li> <li>5.- La eliminación completa del problema o síntoma.</li> </ol>
<i>Evaluación Subjetiva</i>	<p>Impresiones, juicios y opiniones del paciente o de quienes interactúan con él que señalan a que el cambio es claramente importante y marca la diferencia.</p>	<p>Valoración de:</p> <ol style="list-style-type: none"> <li>1.- Funcionamiento actual.</li> <li>2.- Si el problema original sigue siendo evidente y afecta al funcionamiento normal de la persona.</li> <li>3.- Los cambios producidos en el tratamiento generan una diferencia real para la vida del paciente.</li> </ol> <p>El Cambio se refleja en medidas prácticas como: el absentismo escolar, los días perdidos de trabajo, la hospitalización, la supervivencia o el coste.</p>
<i>Impacto Social</i>	<p>Cambio en una medida que es reconocida o considerada como de importancia crítica en la vida cotidiana. Normalmente esta medida no consiste en un inventario psicológico o una medida estandarizada.</p>	

Nota: Adaptado y modificado de Kazdin (2001, p.456).

## **5.2. Análisis estadísticos habituales en la evaluación del cambio terapéutico**

El elemento determinante de los estudios de evaluación de resultados clínicos es la valoración del cambio producido en los diferentes grupos observados. Este énfasis en el cambio implica necesariamente una perspectiva longitudinal con medidas repetidas de las variables dependientes y unas estrategias de análisis de datos adecuadas a esta circunstancia. En los estudios de eficacia de intervenciones psicológicas se ha optado clásicamente por los métodos de análisis derivados del Modelo Lineal General (fundamentalmente los métodos de análisis de la varianza) y en la operativización de los resultados evaluados en forma de variables continuas (Brown, Costigan y Kendziora 2008; Chartier y Faulkner 2008; Rivera y Bernal 2008). Este apartado se centrará en los estadísticos de significación y de tamaño de los efectos ubicados en ese contexto.

### **5.2.1. Pruebas de significación**

Para evaluar el cambio terapéutico es necesario recoger como mínimo dos medidas de cada sujeto estudiado, una antes y otra después de la intervención. El método de análisis más elemental es hallar la diferencia entre esas dos medidas como indicador del cambio. Para poder descartar que este cambio no se deba en realidad a una fluctuación por azar, será necesario aplicar una prueba de significación a esa diferencia. Si se está analizando el cambio con un solo grupo de pacientes, la prueba t de Student-Fisher (Student 1925b; Fisher 1925) para datos emparejados (contraste de hipótesis nula de no diferencia según distribución t y n-1 grados de libertad), será el método de elección. Si no se cumple su supuesto de aplicación (normalidad de la distribución de las diferencias en la población), se recurrirá a la prueba no paramétrica de los rangos con signo de Wilcoxon (1945). Sin embargo, como ya se ha mencionado en capítulos anteriores, tener en cuenta un solo grupo a la hora de valorar el cambio dificulta atribuir éste a la intervención, ya que pueden incurrir numerosos factores que lo hayan generado al margen de la variable independiente (menor validez interna). Por ello, en los ensayos clínicos controlados se incluye un grupo control equivalente al grupo tratado que no haya sido intervenido (Domenech, 2013c).

La inclusión de varios grupos en un estudio hace necesario métodos de análisis que sean compatibles con esta circunstancia. El uso de la t de Student no será posible si se pretende

comparar el rendimiento de una intervención, en más de dos grupos (niveles de la variable independiente) o con respecto a más de dos medidas repetidas ya que no se puede operativizar mediante diferencia de medias. El análisis de la varianza (ANOVA) y los métodos derivados de ella, permiten dar respuesta a estas necesidades gracias a su flexibilidad para afrontar el análisis de datos con cualquier número de variables independientes (con sus respectivos niveles o categorías) y variables dependientes (diferentes indicadores de resultado medidos un determinado número de veces). En vez de realizar una comparación de la medias de la variable dependiente de forma directa, utiliza la varianza entre las diferentes medias para calcular un estadístico de contraste que ponga a prueba la posibilidad de que las diferencias observadas se deban al azar. Este estadístico de contraste es el denominado F de Snedecor (1934) . El cual consiste en dividir la varianza observada en la muestra entre la varianza estimada que sería propia de la variabilidad de la población, que no quedaría explicada por el valor de la variable o variables independientes (denominada varianza residual o error) y contrastarlo en la distribución F, con los grados de libertad de ambas varianzas. En su configuración más sencilla (una sola medida y dos grupos) la prueba F mantiene equivalencia con la prueba t ( $t_{g1; \alpha/2} = \sqrt{F_{1;g1; \alpha}}$ ) por lo que se puede considerar a esta última como un caso particular del análisis de la varianza (Pardo, Ruiz y San Martín 2009; Field, 2009). Cuando se están comparando más de dos grupos o categorías del factor es necesario recurrir adicionalmente a un análisis a posteriori (Field, 2009) que identifique entre que medias existe diferencia significativa (la prueba F realiza el contraste en conjunto), para este fin existen diversas alternativas que controlan la posible distorsión de realizar contrastes múltiples tales como las propuestas por: Bonferroni (1936), Tukey (1953) o Scheffé (1953).

En la Tabla 11 se señalan las principales variantes del análisis de varianza en los estudios clínicos en psicología. Estas permiten analizar los efectos intersujeto (diferencias entre grupos experimentales), los efectos intrasujeto (diferencias de los diferentes momentos de medida o bajo diferentes condiciones dentro de cada grupo) o los posibles efectos de interacción entre los factores del estudio. También posibilitan realizar análisis con factores aleatorios (los niveles de la VI incluidos en el estudio son una selección al azar de todos los niveles posibles) y con factores fijos (niveles de VI determinados a priori). Además se puede controlar el efecto de terceras variables extrañas que covarían con las variables resultado, y de las que se desea eliminar su efecto. (Brown, Costigan y Kendziora, 2008).

**Tabla 11***Modelos de ANOVA en investigación en psicología clínica*

<b>Modelo</b>	<b>Objetivo</b>	<b>Variables Independientes (VI)</b>	<b>Variables Dependientes (VD) *Continuas</b>	<b>Uso habitual en investigación clínica</b>
<i>ANOVA de un factor (One way)</i>	Comprobar las diferencias entre los grupos (niveles de VI).	1 VI con Niveles $\geq 2$	1 VD	Comparar las diferencias entre intervenciones en los promedios.
<i>ANOVA Multifactorial</i>	Comprobar las variaciones en el resultado como una función de los principales efectos de los factores y la interacción entre ellos.	VIs $\geq 2$ con Niveles $\geq 2$ cada una.	1 VD	Comparar el impacto de la intervención sobre los diferentes grupos o comparar los efectos combinados de las diferentes intervenciones.
<i>Análisis de la Covarianza (ANCOVA)</i>	Comprobar la variación en la VD como una función de las diferencias de grupo después de controlar los efectos de covarianza.	VIs $\geq 1$ categoriales VIs $\geq 1$ covariantes continuas	1 VD	Fácilmente extensible para estudiar posibles interacciones entre variables.
<i>ANOVA de medidas repetidas</i>	Comprobar la variación en la VD como una función de las medidas repetidas (tiempo).	1 VI con Niveles $\geq 2$	1 VD	Examinar el impacto de las intervenciones en los cambios en las puntuaciones de los sujetos a lo largo del tiempo.
<i>Modelo Mixto de ANOVA</i>	Comprobar la variación en la VD como una función de factores fijos y factores aleatorios.	VIs $\geq 1$ de efectos fijos VIs $\geq 1$ de efectos aleatorios	1 VD	Examinar el impacto de la intervención cuando es aplicada en unos grupos que son una muestra de todos los posibles grupos de factor (efecto aleatorio).
<i>Análisis Multivariado de la varianza (MANOVA)</i>	Comprobar las diferencias entre grupos en una combinación lineal de VDs	1 VI con Niveles $\geq 2$	VDs $\geq 2$	Examinar el impacto de la intervención en múltiples y relacionados tipos de resultados.
<i>Análisis Multivariado de la covarianza (MANCOVA)</i>	Comprobar la variación en múltiples VDs como una función de las diferencias de grupo después de controlar los efectos de covarianza.	VIs $\geq 1$ categoriales VIs $\geq 1$ Covariantes continuas	VDs $\geq 2$	Fácilmente extensible para estudiar posibles interacciones entre variables.

Nota: Tabla adaptado de Brown, Costigan y Kendziora (2008, p.290).

La aplicación de los modelos de análisis de la varianza implica asumir una serie de supuestos cuyo no cumplimiento puede complicar las inferencias realizadas a partir de ellos (Pascual, Frías, García 1996):

- Independencia de las observaciones realizadas. Es decir que una medida no proporciona información sobre otra medida individual (no existe correlación entre observaciones).
- Homogeneidad de varianzas (homocedasticidad). Las varianzas de las poblaciones de donde provienen los grupos experimentales sean iguales. Como prueba de contraste de este supuesto: Prueba de Levene (1960).
- Normalidad en distribución de variables dependientes. Las observaciones se extraen de poblaciones distribuidas según la ley normal. Como prueba de contraste de este supuesto: Shaphiro-Wilk (1965).
- Esfericidad (solo en situaciones de medidas repetidas). Las varianzas de las diferencias de cada par de medias de medidas repetidas sean constantes. Como posible prueba de contraste de este supuesto: Test de Esfericidad de Machly (1940).
- Supuestos de covarianza (solo en modelos que la incluyan): relación lineal entre variables covariante y dependiente, independencia entre variable covariante y variable independiente y homogeneidad de rectas de regresión (no efectos de interacción entre variable covariante e independiente).

En su aplicación al entorno clínico es difícil que todos estos supuestos se cumplan, y se ha estudiado el impacto de su vulneración. Resultando que el no cumplimiento de supuestos como el de normalidad o el de homogeneidad de varianzas afecta poco o moderadamente a la prueba F, especialmente si se cuenta con una muestra grande o parecido número de personas por grupo; aunque sí puede tener importancia para la elección de contrastes a posteriori. En cambio, se resalta el gran efecto del no cumplimiento de supuestos como el de independencia de observaciones de forma general para los modelos de ANOVA o el de esfericidad para situaciones de medidas repetidas (Harwell, Rubinstein, Hayes y Olds 1992; Pascual, Frías, García 1996; Berkovits, Hancock y Nevitt, 2000).

Para afrontar las situaciones de vulneración de supuestos, se han propuesto diferentes vías. Existen alternativas no paramétricas de menor potencia como la de Kruskal-Wallis (1952) o la

de Friedman (1937) para situaciones donde la muestra se aleje demasiado de la normalidad. En caso de no poder asumir igualdad de varianzas (sobre todo si el tamaño de los grupos que se compara es muy diferente), se recomienda utilizar las pruebas paramétricas equivalentes a la F de Snedecor: de Brown-Forsythe (1974) o de Welch (1951). En otro sentido, se han planteado ajustes de los grados de libertad de la prueba F para compensar la falta de homogeneidad o de esfericidad como los de Geisser y Greenhouse (1958) o Huynh-Feldt (1976). Otro procedimiento sugerido ha sido realizar transformaciones equivalentes de la variable dependiente para que se pueda ajustar a los diferentes supuestos. También se ha señalado la idoneidad de adaptar el análisis de datos a otros modelos de análisis de varianza alternativos como puede ser el caso del uso del MANOVA o el de modelos mixtos de ANOVA cuando no se cumple la esfericidad necesaria para un ANOVA de medidas repetidas (Pascual, Frías, García 1996; Moder 2010; Lantz 2013).

Por otro lado, las estrategias de análisis de la varianza (con los mismos supuestos expuestos anteriormente) se pueden generalizar a modelos de regresión lineal simple o múltiple, ya que se engloban dentro del Modelo Lineal General. Para ello, es necesario transformar las variables independientes (de tipo categorial en los modelos ANOVA) a un formato continuo, en lo que se denomina variables ficticias o dummy, que implica por ejemplo, convertir la variable binaria grupo tratamiento o grupo control en 1 y 0 respectivamente (Iraurgi, 2000; Chartier y Faulkner, 2008). En la investigación de intervenciones psicológicas ha cobrado cierta relevancia el uso de los modelos de regresión para poder analizar de forma exhaustiva el proceso de cambio terapéutico, concretamente a través de dos procedimientos: mediación y moderación. Estos estudian el papel de terceras variables en la relación entre la variable dependiente (resultado) y variable independiente (tratamiento) como agentes intermedios, representando el siguiente eslabón en la cadena causal o moderando el efecto de una sobre la otra (Kazdin y Nock, 2003; Hayes, Laurenceau y Cardaciotto, 2008).

Aunque los modelos ANOVA han sido la elección predominante como medio de análisis de datos en los estudios de la eficacia de las intervenciones, en los últimos años se ha empezado a extender el uso de los métodos desarrollados en el marco del denominado Análisis de Crecimiento Modelado (en su acepción original Growth-Modeling Analysis, GMA), y que engloba a los siguientes procedimientos (Feingold, 2009): Modelos multinivel o también

conocidos como Modelos de Regresión de Efectos Mixtos (Hedeker y Gibbons, 2006), Modelos de Crecimiento Latente (Singer y Willett, 2003) y las Ecuaciones de Estimación General (Hardin y Hilbe, 2003). Aunque cada uno de estos métodos tiene sus particularidades e implicaciones matemáticas diferentes en esta exposición se hará referencia a las características genéricas comunes que los agrupan como GMA, ya que una mayor profundización requeriría de una extensión excesiva.

El GMA se basa en la construcción, análisis y comparación de curvas de crecimiento. Las curvas de crecimiento se infieren a partir de las medidas repetidas de cada sujeto ajustándolas a trayectorias continuas que subyacen a lo observado mediante modelos de regresión (lineal o no lineal). Esta estimación se realiza con cada paciente modelizando la trayectoria de su cambio individual en el tiempo. Posteriormente se comparan las curvas de los diferentes grupos de tratamiento para hallar diferencias entre las trayectorias y por tanto diferencias con respecto a la respuesta en la variable dependiente en los distintos niveles de la variable independiente (Curran y Willoughby, 2003; Atkins 2005; Feingold, 2009).

Son modelos flexibles que permiten tener en cuenta tanto factores fijos como aleatorios, los efectos de las diferentes covarianzas e incluso relaciones no lineales entre variables. Con respecto a los modelos ANOVA tienen la ventaja de no necesitar que los datos estén balanceados (que haya el mismo número de observaciones por categoría), de que se puedan utilizar los datos de los sujetos que no estén completos (de los que no se pudo tomar todas las medidas que se tomaron en los demás sujetos) y de mostrarse más idóneos para las situaciones de medidas repetidas por tener en cuenta la covarianza de las puntuaciones de un mismo sujeto (en especial si recogen gran cantidad de medias repetidas). Sin embargo, en situaciones con muestras pequeñas o subgrupos reducidos estos modelos pueden proporcionar estimaciones sesgadas y alejadas de la realidad, necesitando además de múltiples medidas repetidas siendo insuficiente solo dos medidas (pre y post tratamiento) en la mayoría de los casos (Arnau y Balluerka, 2004; Arnau y Bono, 2008).

Otra alternativa para el análisis del cambio terapéutico se puede encontrar en los modelos de series temporales univariadas tales como el ARIMA (Autoregressive Integrated Moving Average) que se basan en la obtención de muchas medidas repetidas de la variable interés en un mismo sujeto planteado para evaluar el impacto de las intervenciones en casos únicos. Sin

embargo, debido a la gran cantidad de medidas repetidas necesarias y a la complejidad de su cálculo y modelización han sido de uso minoritario en la investigación psicológica (Arnau y Bono, 1993; Bono Arnau y Vallejo, 2008).

### 5.2.2. Índices de tamaño del efecto

Los indicadores de tamaño del efecto se han dividido clásicamente en tres clases generales: los de la familia de las diferencias entre grupos, los de la familia de los índices de correlación y los de la familia de las ratios. En esta exposición nos centraremos sobre las dos primeras ya que son los que sirven para analizar variables dependientes métricas que son de mayor relevancia en psicología clínica, mientras que los indicadores en forma de ratios (datos categoriales) que son más populares en la literatura médica no se abordarán aquí (Rosnow y Rosenthal, 2008, Talavera y Rivas-Ruiz, 2011 y Gibert, 2012; Morales, 2012). En la Tabla 12, (al final del apartado) se presentan las formulaciones de los índices a las que se va hacer referencia a continuación.

La primera propuesta en el sentido de medir el tamaño del efecto a partir de las diferencias entre grupos es la de Cohen (1988). Este autor señaló que observar las diferencias de medias directas tenía el inconveniente de permanecer en términos de la unidad de medida utilizada y relativa a los datos concretos de los grupos de estudio, dificultando poder sacar conclusiones de esa diferencia e imposibilitando la comparación con otros estudios. Por ello, propone realizar una estandarización de esa diferencia y expresar la magnitud en términos de número de desviaciones estándar de forma equivalente al procedimiento de generación de puntuaciones  $z$ . Este proceso implica dividir la diferencia entre las medias grupales (por ejemplo, del grupo experimental menos la del grupo control) entre una desviación estándar (DE) que represente la dispersión de la variable dependiente. En la notación original, la DE se calcula realizando la media no ponderada de las desviaciones estándar de ambos grupos, pero bajo la suposición de mismo número de sujetos en los dos grupos y distribución normal de la variable dependiente.

Para Glass, McGaw y Smith (1981) la DE que se debería utilizar es únicamente la del grupo control (sólo en los casos que implique diseño grupo control-grupo experimental) ya que consideran que la aplicación del tratamiento altera la variabilidad natural en el grupo experimental por lo que ésta debe excluirse en la estimación de dispersión de la variable dependiente, generando el indicador denominado *Delta de Glass*  $\Delta$ . Por su parte Hedges (1981)

considera que las dos anteriores propuestas están positivamente sesgadas (sobreevalúan el tamaño del efecto) por lo que propone dos modificaciones a la formulación original de Cohen. Primero la obtención de una media ponderada de las DEs de ambos grupos como DE de la variable dependiente, permitiendo de esta forma comparación entre grupos con diferente tamaño (aunque sí necesita del supuesto de homogeneidad de varianzas). Y, segundo, aplicar al indicador una corrección matemática de sesgo basada en los dos tamaños muestrales (ver tabla 4.x). Se genera de esta forma el *Índice g*. Según algunos estudios de simulación las correcciones de Hedges provocan una gran infra estimación del tamaño del efecto, siendo recomendable la formulación original pese a sus limitaciones (Marín y Sánchez 1996). De hecho, en la revisión de McMillan y Foley (2011), se observa que el *Índice g* es muy poco utilizado en la literatura en comparación a formulación original de Cohen.

Según la situación de experimental, los usos recomendados concretos de la  $d$  de Cohen pueden diferir en referencia a la DE utilizada. Cuando sólo se comparen medidas grupo control y experimental sin medidas repetidas, es preferible utilizar la *Delta de Glass*  $\Delta$  (es decir la DE del grupo control). En situaciones de muestras relacionadas sin grupo control (solo medias pre tratamiento y post tratamiento del mismo grupo tratado) se realizará la diferencia entre los dos momentos de medida y se dividirá entre la DE del momento pre tratamiento. Cuando existan grupo control y grupo tratamiento se recomienda utilizar solo las puntuaciones post tratamiento de ambos grupos y dividirlo entre la DE post tratamiento del grupo control (es decir, interpretar la magnitud el estado final). Así a menos que las medidas pre tratamiento de ambos grupos sean muy diferentes, entonces si es relevante incluirlas en el análisis (Morris, 2008; Morales, 2012).

Para la interpretación cualitativa del indicador resultante (de esta diferencia tipificada), Cohen (1988) propone la siguiente guía basada en su propia revisión de los tamaño del efecto en investigación general:  $d=0$  no efecto,  $d= \pm 0.20$  tamaño del efecto pequeño (poca diferencia entre grupos),  $d= \pm 0,50$  efecto moderado (el efecto medio en investigación) y  $d \geq \pm 0,80$  efecto importante (una diferencia grande entre grupos, siendo muy poco habitual un valor de  $d$  mayor que 1). Rosenthal (1996) añade a esta guía un umbral más en  $d= \pm 1,30$  como efecto muy grande.

En este sentido Lipsey (1990) propone utilizar la distribución de los tamaños del efecto de estudios anteriores en el área para poder interpretar cualitativamente el indicador. El autor realiza una revisión de más de 100 metaanálisis de estudios en psicología y educación. La

distribución de los tamaños del efecto de estos estudios se dividió en tres partes: el tercio inferior que se ubicaba entre los tamaños 0 y 0,32; el tercio intermedio figuró entre 0,33 y 0,55 y el tercio superior entre 0,56 y 1,26. Resultando los puntos intermedios de estos intervalos 0,15, 0,45 y 0,80 respectivamente, es decir esta validación empírica resultó muy cercana a los umbrales planteados por Cohen. Sobre el establecimiento umbrales, el propio Cohen (1988) advierte que deben tomarse como orientaciones generales y no como puntos de corte estrictos. Existen de hecho voces críticas con la utilidad de estos criterios que señalan que: el criterio cualitativo de la magnitud del efecto solo se puede establecer en relación con otras intervenciones que buscan producir el mismo efecto y que en cada contexto específico el valor del estadístico tendrá una importancia relativa diferente (Glass, McGaw y Smith, 1981; Coe y Merino, 2003).

Otra propuesta para la interpretación y cálculo del tamaño del efecto desde las diferencias entre grupos, es expresarlo en términos de un valor de probabilidad, aprovechando el supuesto de distribución normal en estos indicadores. McGraw y Wong (1992) lo operatizan a través del denominado *Índice CLES* (Common Language Effect Size), que consiste en generar una puntuación *Z* de las diferencias de medias de ambos grupos (ver tabla 4.x) y buscar la probabilidad (según las tablas de distribución normal tipificada) de obtener un valor menor al obtenido. Esa probabilidad será la de encontrar un sujeto extraído al azar del grupo de tratamiento que tenga una puntuación mayor que sujeto extraído al azar del grupo control.

En los modelos ANOVA con situaciones de contraste de muestra independientes es posible plantear la estimación del tamaño del efecto con un cálculo equivalente a la *d* de Cohen, en los contrastes a posteriori (contrastos de dos en dos grupos) o mediante adaptaciones de la formulación original incluyendo elementos de la prueba *F* y el tamaño de los grupos (Cortina y Nouri, 2000). Sin embargo, para estos modelos se disponen estimadores de magnitud del efecto particulares más cercanos en su naturaleza e interpretación a los índices de correlación que a los derivados de la diferencia de medias.

El de uso más habitual es el denominado eta cuadrado  $\eta^2$ , propuesto originalmente por Pearson (1911). Consiste en calcular qué proporción del total de la varianza en los grupos de estudio se puede atribuir al cambio de valor de la variable independiente (pertenecer a un grupo de tratamiento u a otro). Para ello, se divide la suma de cuadrados (SS) entre grupos o del efecto

(la que depende del valor de la variable independiente) entre la SS total observada (elementos calculados como indicadores de la variabilidad en la distribución de los datos del estudio e implicados en la generación del estadístico F). El coeficiente resultante es similar al coeficiente de determinación de una regresión lineal múltiple ( $R^2$ ), con una interpretación también parecida. Será un valor comprendido entre 0 (donde los cambios en la variable dependiente no explican ninguna proporción de la variación en la variable dependiente) y 1 (donde los cambios en la variable independiente explican totalmente las variaciones en la variable dependiente) (Levine y Hullet, 2002; Field, 2009). Cohen (1988) también estableció unos umbrales orientativos para las ciencias sociales para interpretación cualitativa de  $\eta^2$ :  $\eta^2 = 0,01$  efecto pequeño,  $\eta^2 = 0,06$  efecto mediano y  $\eta^2 = 0,14$  efecto grande.

Con posterioridad se planteó una ligera modificación de este mismo indicador que se ha venido a denominar  $\eta^2_p$  (eta cuadrado parcial). Sustituye en el denominador a SS total por la suma de las SS entre grupos y las SS del error (la variabilidad que no es debido a la pertenencia a un grupo), a efectos prácticos resulta en una estimación más permisiva de la magnitud del efecto, y con una interpretación algo más problemática que el  $\eta^2$  habitual, por lo que su uso no está muy extendido fuera del ámbito de la comparación entre estudios con un mismo diseño experimental (Levine y Hullet, 2002).

Por otro lado, Hayes (1963) plantea su propia corrección a la  $\eta^2$  cuadrado clásico, ya que lo considera sesgado para su uso como estimador del efecto en la población. Relativiza la magnitud del efecto teniendo en cuenta las características del estudio (número de grupos, el tamaño de la muestra y la variabilidad dentro de los grupos). A la formulación clásica añade la Media Cuadrática residual o del Error, (el denominador del estadístico F) y el número de grados de libertad del efecto (el número de categorías de la variable independiente menos 1). Este autor al indicador resultante lo designa como  $\omega^2$  (omega cuadrado). Es un ajuste que tiende a reducir la magnitud del efecto del estudio con respecto al  $\eta^2$ , y en la literatura se señala como alternativa adecuada en situaciones donde el tamaño de los grupos sea reducido, sin embargo puede resultar en exceso restrictiva si se cuenta con tamaño muestral grande, ya que el sesgo que se corrige ya no afecta en tanta medida (Lakens, 2013).

La familia coeficientes de correlación derivan del planteamiento de Pearson (1900) acerca de la obtención de un indicador de asociación lineal entre dos variables cuantitativas

denominado *r de Pearson*. Consiste en estandarizar la covarianza entre dos variables. Definiéndose covarianza como una medida de variabilidad conjunta de ambas variables, que toma valores positivos cuando la relación es creciente, o negativos cuando la relación es decreciente (inversa). La *r de Pearson* libera la covarianza de las unidades de medida de las variables al dividirlo por sus desviaciones respectivas desviaciones estándar. El valor del coeficiente estará comprendido entre -1 y +1. Donde +1 es una asociación lineal total (cuando variable cambia la otra cambia en el mismo sentido), -1 asociación lineal negativa total (cuando una variable cambia la otra cambia en sentido contrario) y 0 es ausencia de relación lineal entre variables. Por lo que la relación será más fuerte cuanto más se acerque a  $\pm 1$  (Guàrdia, Freixa, Però y Tubany, 2008). De nuevo Cohen (1988) ofrece unos criterios orientativos cualitativos para la interpretación de este coeficiente como indicador de magnitud del efecto en las ciencias sociales:  $r = \pm 0,10$  asociación pequeña;  $r = \pm 0,30$  asociación media y  $r = \pm 0,50$  asociación grande. Otra forma de estudiar esta asociación es elevando al cuadrado el coeficiente  $r$  y obtener el coeficiente de determinación  $R^2$  que expresa la proporción de la varianza común entre las dos variables, que multiplicado por 100 se explicaría en forma de porcentaje (Field, 2009).

La *r de pearson* original necesita que tanto variable dependiente como variable independiente estén medidas de forma continua y bajo supuesto de distribución normal. Sin embargo, se han propuesto diversos desarrollos para poder generar coeficientes de correlación con diferentes tipos de variables y distribuciones (Jackson, 2010), tales como: Coeficiente de Spearman (1910), Coeficiente de Correlación Phi (Yule, 1912) o la Presentación Binomial del efecto (Rosnow y Rosenthal, 2008). Se destaca en esta exposición el denominado coeficiente de correlación Biseral Puntual  $r_{bp}$ , por sus implicaciones en la evaluación de resultados clínicos. Este pone en relación la variable independiente categorial dicotómica pertenencia al grupo (dos niveles del factor, por ejemplo experimental y control) con la variable dependiente continua resultado (Brown, 2001). Para ello es necesario calcular la proporción de sujetos del estudio que pertenecen a o una u otra categoría-grupo ( $p$  y  $q$ ) y la puntuación media de la variable dependiente también en cada grupo ( $M_p$  y  $M_q$ ). Se genera un índice de asociación con una interpretación equivalente a la *r de pearson*. Una propiedad muy relevante de este coeficiente es la facilidad de conversión desde los indicadores de tamaños del efecto derivados de la *d de cohen*. Pudiendo transformar los diferentes índices de magnitud del efecto basados en diferencia de medias a una escala común en forma de coeficiente de correlación, facilitando así la

comparación entre estudios (Iraurgi, 2009b; Rosnow y Rosenthal 2008; Morales, 2012). En la Tabla 12, también se presenta tanto el cálculo básico del  $r_{bp}$  así como las formulas de conversión desde los otros índices.

**Tabla 12**

*Formulación de principales indicadores de tamaño del efecto grupal comentados en este apartado*

<b>Índices basados en diferencias de medias.</b>	$d = \frac{M_{tto} - M_{ctr}}{\sqrt{\frac{DE_{tto}^2 + DE_{ctr}^2}{2}}}$	$g = \frac{M_{tto} - M_{ctr}}{\sqrt{\frac{(n_{tto} - 1)DE_{tto}^2 + (n_{ctr} - 1)DE_{ctr}^2}{n_{tto} + n_{ctr} - 2}}} \times 1 - \frac{3}{4(n_{tto} + n_{ctr} - 2) - 9}$	
	$\Delta = \frac{M_{tto} - M_{ctr}}{DE_{ctr}}$	$CLES Z = \frac{M_{tto} - M_{ctr}}{DE_{tto} + DE_{ctr}} ; p(Z \leq Z_0)$	
<b>Índices de correlación como tamaño del efecto desde otros índices.</b>	$r = \frac{DE_{xy}}{DE_x DE_y}$	<p>-Calculo a partir de d: Si <math>n_1 = n_2</math> <math>r_{bp} = \frac{d}{\sqrt{d^2 - 4}}</math>; Si <math>n_1 \neq n_2</math> <math>r_{bp} = \frac{d}{\sqrt{d^2 + \frac{1}{pq}}}</math></p>	
	$r_{bp} = \frac{M_p - M_T}{DE_T} \sqrt{\frac{p}{q}}$	<p>-Calculo a partir de g: <math>r_{bp} = \sqrt{\frac{g^2 n_1 n_2}{g^2 n_1 n_2 + [(n_1 + n_2)(n_1 + n_2 - 2) ]}}</math></p>	
<b>Índices basados de Tamaño del efecto en ANOVA.</b>	$\eta^2 = \frac{SS_g}{SS_t}$	$\eta_p^2 = \frac{SS_g}{SS_g + SS_e}$	$\omega^2 = \frac{SS_g - (k - 1)(MC_e)}{SS_t + CM_e}$

*Nota: d= d de Cohen; Δ= Delta de Glass; g= g de Hedges; CLES Z= Z para Common Language Effect Size; r= Coeficiente de correlación de Pearson;  $r_{bp}$ = Coeficiente de correlación biseral puntual;  $\eta^2$ = Eta cuadrado;  $\eta_p^2$ = Eta cuadrado parcial;  $\omega^2$ = Omega cuadrado;  $M_{tto}/M_{ctr}$ = Medias de grupo tratamiento y grupo control;  $DE_{tto}/DE_{ctr}$ = Desviaciones estándar grupo tratamiento y grupo control;  $DE_{xy}$ = Covarianza entre variable x e y;  $DE_x/DE_y$ = Desviaciones estándar de variables X e Y;  $p/q$ = proporción de sujetos en una de los dos grupos o categorías;  $M_p$ = Media de variable resultado en uno de los dos grupos o categorías;  $M_T$ = Media total en variable resultado (los dos grupos);  $DE_T$ = Desviación estándar total en variable resultado (los dos grupos);  $n_{tto}/n_{ctr}/n_1/n_2$ = Número de sujetos: de grupos: tratamiento, control, condición 1 y condición 2;  $SS_g/SS_e/SS_t$ = Suma de cuadrados: entre grupos ( del efecto), del error (residual) y total;  $MC_e$ = Media cuadrática del error ( denominador de prueba F); k= Número de categorías o grupos de la variable independiente.*

### 5.3. Metodologías de análisis de cambio individual y de su relevancia clínica

En el primer apartado de este capítulo se ha explicitado que a pesar de la importancia y utilidad de las pruebas de significación y el cálculo de los tamaños del efecto, éstas no reflejan si a nivel del paciente individual se ha generado un cambio clínicamente relevante para su funcionalidad. Los estadísticos presentados en el epígrafe anterior valoran el cambio terapéutico a partir de medidas grupales y están condicionadas a la obtención de un determinado tamaño muestral determinado e incluso de varios grupos de estudio. En el capítulo 3 se habló de la necesidad de complementar los necesarios estudios controlados de eficacia con estudios de efectividad de la práctica habitual. Sin embargo, en el entorno clínico real es difícil contar con las condiciones experimentales óptimas para realizar el mismo tratamiento estadístico. Además, la prioridad del clínico es saber si la intervención que ha realizado tiene un efecto relevante en la situación particular de cada paciente. Por ello, se hizo necesario la generación de metodologías de análisis de resultados que diesen respuesta a estas necesidades (Iraurgi, 2010).

Según Ogles, Lunnen y Bonesteel (2001) la metodología de Índice de Cambio Fiable propuesta por Jacobson y Truax (1991) ha sido la que mayor aceptación ha tenido en el ámbito académico para abordar este problema, ya que resulta ser una combinación relativamente sencilla de aplicar las dos consideraciones más destacadas acerca del cambio clínicamente significativo (CCS) en Psicología:

- Recuperación: El paciente alcanza un estado de salud normativo al finalizar la intervención. Es decir, se hace empíricamente indistinguible de una persona sin el problema tratado.
- Mejoría: El paciente alcanza mejoras estadísticamente fiables como resultado del tratamiento.

Estas dos facetas del CCS se corresponden respectivamente con dos tipos de métodos que pretenden determinarla a partir de las respuestas de cada paciente (Turner, et al. 2010; Pardo y Ferrer, 2013):

- Los métodos basados en criterios externos a la propia herramienta de medida. Que consisten en relacionar el valor de la medida con un criterio clínico externo (comparación normativa, juicio profesional etc..).

- Lo métodos basados en criterios basados en criterios internos. Que se centran exclusivamente en la información de las puntuaciones de los pacientes y las características de la herramienta de medida.

Parece existir cierto consenso a la hora de señalar que la estrategia idónea de valorar el CCS es tener en cuenta ambos tipos de métodos (Sheldrick, Kendall y Heimberg 2001; Crosby, Kolotkin y Williams 2003; De Vet et al., 2007) tal y como lo contempla el procedimiento propuesto por Jacobson y Truax (1991).

### **5.3.1. Método de Índice de Cambio Fiable de Jacobson y Truax (ICF<sub>JT</sub>)**

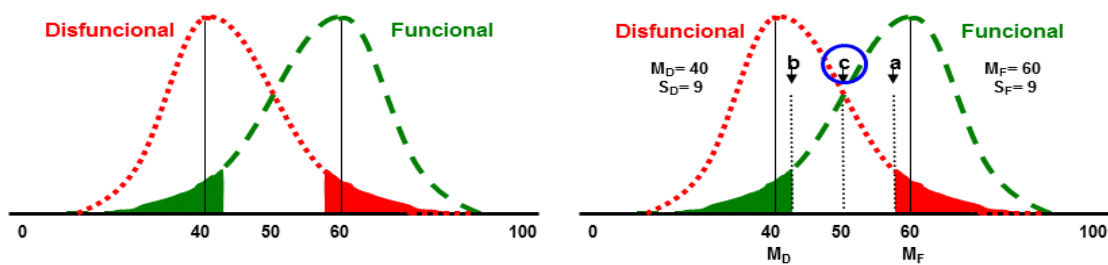
En este apartado se va explicar la metodología de Índice Cambio Fiable tal como se refleja en el artículo de Jacobson y Truax (1991) y posteriormente es reafirmada por Jacobson, Robers, Berns y McGlinchey (1999). Aunque la propuesta inicial es anterior (Jacobson, Follette y Revenstorf, 1984; 1986), el equipo de Jacobson asumió e incorporó las sugerencias y correcciones de Christensen y Mendoza (1986) al planteamiento matemático original (en referencia a la generación de un criterio interno de mejoría). La acepción inglesa original es Reliable Change Index (RCI), que aquí se ha traducido como Índice de Cambio Fiable, al que haremos referencia a partir de ahora por el acrónimo ICF<sub>JT</sub> para mayor fluidez de la explicación y para diferenciarlo de otros índices de cambio fiable alternativos que se presentarán posteriormente.

Ante la dificultad para alcanzar un consenso acerca de qué se puede juzgar como un cambio clínicamente importante, los autores sugieren definirlo como la constatación de un cambio hacia los valores de normalidad en los aspectos clínicos sometidos a intervención. Viene a significar la valoración de la “recuperación” de un funcionamiento normal con respecto a indicadores de salud. Para conceptualizar este proceso se valen de la consideración de la distribución poblacional del indicador medido. Si por ejemplo se obtiene un indicador (una medida) del nivel de salud del paciente, antes y después de la intervención, se tendrá que constatar que en el momento posterior el indicador ubica al paciente fuera de la población disfuncional a la que inicialmente pertenecía. Para operativizar esto, se propone utilizar un método de comparación normativa que permita calcular un determinado punto de corte, a partir del cual la puntuación del indicador posibilita considerar al paciente como perteneciente a la población funcional. En este

cálculo es necesario disponer de los datos de por lo menos una de las distribuciones de referencia, que representen o a la población funcional o a la población disfuncional.

Para poder comprender la lógica de las distribuciones de referencia se debe hacer la abstracción teórica de la distribución poblacional del indicador que se esté manejando. Como manifestación de un fenómeno real en la totalidad de la población se puede asumir que su distribución se aproximará a una curva normal, donde las puntuaciones más extremas sean las menos frecuentes (por ejemplo muy mala o muy buena salud), y la mayoría de las personas posean puntuaciones cercanas a un valor promedio más o menos centrado. Si de esa población extrajésemos todas las personas que no manifiestan un trastorno que implique el deterioro del indicador que estamos manejando, obtendríamos una distribución de población funcional o no enferma, que como es de esperar tendría puntuaciones promedio que se ubicarían con respecto a la distribución de población total hacia uno de sus lados (en el caso de un índice de buena salud hacia la derecha, valores más elevados). De la misma forma si se extrajésemos de esa misma población las personas que presentan el trastorno o grupo de trastornos de interés para el estudio, obtendríamos una distribución de población disfuncional que también se lateralizará con respecto a la distribución total, si continuásemos con el índice de salud, hacia la izquierda, valores más pequeños.

En la Figura 3 se ilustra este proceso lógico según el ejemplo de Iraurgi (2010) en relación a un indicador de salud percibida. Si el indicador fuese una escala sintomática y una disminución de su valor sería representativo de mejoría clínica, el orden de las curvas de distribución sería el inverso (derecha disfuncional e izquierda funcional). A nivel práctico no se va poder contar con los datos de la población total, pero sí con selecciones muestrales representativas de ésta. Es habitual contar con datos de muestras disfuncionales amplias e incluso normativas. Sin embargo, es más complicado disponer de muestras representativas de población funcional, por lo que puede ser necesario recurrir a muestras normativas de población general (provenientes, por ejemplo, de encuestas de salud general) donde existe la posibilidad de que haya sujetos con niveles clínicos o subclínicos del trastorno que se está tratando. No es raro que exista cierto solapamiento en las puntuaciones de ambas distribuciones.



$$a = M_D + 2DE_D$$

$$b = M_F - 2DE_F$$

$$c = \frac{M_D \cdot DE_F + M_F \cdot DE_D}{DE_D + DE_F}$$

$M_D$  = Media grupo Disfuncional  
 $DE_D$  = Desviación Estándar grupo Disfuncional  
 $M_F$  = Media grupo Funcional  
 $DE_F$  = Desviación Estándar grupo Funcional

**Figura 3** Puntos de corte de significación clínica en el modelo de Jacobson y Truax según criterio externo. Adaptado y modificado de Iraurgi (2010, p.108).

Teniendo en cuenta la relación entre estas dos distribuciones y de la disponibilidad de datos acerca de ellas, los autores postulan tres criterios para ubicar un punto de corte para considerar que según el indicador el paciente retorna a un funcionamiento normal (representados en las curvas de distribución de la derecha en la Figura 3):

- a) El nivel funcional posterior a la intervención debe caer fuera del rango de población disfuncional, donde el rango se define como la amplitud de dos desviaciones estándar con respecto a la media de la distribución disfuncional y en dirección a la distribución funcional. Si se trata de un indicador positivo de funcionalidad como es el caso del ejemplo d la figura X sección, se sumarán las dos desviaciones estándar para ubicar el punto de corte a superar (lo que delimita el área derecha coloreada en rojo en de la Figura 3). Si se tratara de un indicador negativo de funcionalidad, se restarían esas dos desviaciones para calcular un punto de corte por el que quedar por debajo (por ejemplo con síntomas).
- b) El nivel de funcionamiento posterior a la intervención debe caer dentro del rango de la población funcional o normal, donde el rango se define como dos desviaciones estándar a partir de la media de la distribución funcional en dirección a la disfuncionalidad. Si se

trata de un indicador positivo de funcionalidad como es el caso del ejemplo de la figura 5.1 sección, se restarán dos desviaciones estándar para ubicar el punto de corte a superar (área derecha coloreada en verde). Si se tratara de un indicador negativo de funcionalidad se sumarían esas dos desviaciones para calcular un punto de corte por el que quedar por debajo (por ejemplo síntomas).

- c) El nivel de funcionamiento posterior a la intervención, sitúa al cliente más cercano a la media de la población funcional que a la media de población disfuncional. Los autores señalan este criterio como el menos arbitrario. Se basa en la probabilidad relativa de que una determinada puntuación termine perteneciendo a una u otra distribución. Para ello, se supone la normalidad de ambas distribuciones. Si se cumple el criterio será estadísticamente más probable ubicar a la puntuación final en la distribución funcional. Para calcular el punto de corte se utilizan las medias (M) y desviación estándar (DE) de ambas distribuciones de referencia funcional (F) y disfuncional (D) para poder ubicar un punto intermedio entre ambas distribuciones (que este a la misma distancia relativa de ambas medias). Este punto se calcularía de la siguiente forma:  $C = [(M_D \times DE_D) + (M_F \times DE_F)] / (DE_D + DE_F)$ . En el caso del ejemplo de la figura X (indicador positivo de funcionalidad) el criterio se cumpliría al superar el punto de corte (hacia la derecha). Si se tratara de un indicador negativo de funcionalidad el criterio, se cumpliría al quedar por debajo de ese punto (hacia la izquierda).

Para ayudar a decidir entre la utilización de uno de estos tres criterios, los autores realizan las siguientes recomendaciones:

- Si se dispone de datos normativos de ambos tipos de población, serán preferibles los criterios “b” y “c” frente al a, ya que son los que explícitamente ubican al paciente en la población funcional.
- Si se carece de una distribución de referencia de población funcional, se optará por el criterio “a”. Sin embargo, hay que tener en cuenta que resulta un criterio muy estricto en problemáticas que exista un importante solapamiento entre las distribuciones del indicador de las poblaciones funcional y disfuncional.

- El criterio “b” será preferible cuando no existe apenas solapamiento entre la población funcional y la población disfuncional. Como puede ser el caso de problemas de salud muy raros o poco comunes.
- El criterio “c” será el idóneo cuando exista un solapamiento evidente entre población funcional y población disfuncional. Circunstancia en la cual, las personas pertenecientes a la población funcional que tienen un valor más bajo en un indicador positivo, pueden tener puntuaciones similares a las personas de la población disfuncional que puntúan más alto. Siendo esta situación la más habitual entre la mayoría de los problemas de salud.

Comprobar el estado de funcionalidad del paciente después de la intervención es indispensable para valorar la significación clínica. Sin embargo, es necesario aplicar otro paso adicional para comprobar que efectivamente ha existido un cambio respecto al estado inicial y que este cambio ha sido de magnitud suficiente para superar el margen de error de la medición (mejora de forma estadísticamente fiable). La propuesta de los autores consiste en un método interno que tiene en cuenta la diferencia entre la puntuación en el momento previo ( $X_{pre}$ ) y la puntuación en el momento posterior ( $X_{post}$ ) al tratamiento, y también la fiabilidad de estas medidas. Se asume que al valorar la magnitud del cambio hay que tener en cuenta que las medidas recogidas están afectadas por cierto grado de imprecisión, que puede sesgar la estimación de su tamaño del efecto. A la propia variabilidad del sujeto se debe añadir la generada por el instrumento. El indicador propuesto para este fin es el que se ha denominado Índice de Cambio Fiable ( $ICF_{JT}$ ). Consiste en dividir a la diferencia entre los dos momentos de medida, un Error Estándar de esa Diferencia (EED), calculado a partir del Error Estándar de la Medida (EEM). Para generar este EEM, se necesita a su vez la desviación estándar de la muestra de referencia ( $DE_{norm}$ ) y el coeficiente de fiabilidad del instrumento de medida que se esté utilizando ( $r_{xx}$ ).

La formulación completa se expresaría de la siguiente manera:

$$ICF_{JT} = \frac{X_{post} - X_{pre}}{EDD}$$

$$EDD = \sqrt{2 EEM^2} = EEM\sqrt{2} \quad EEM = DE_{norm}\sqrt{1 - r_{xx}}$$

$$ICF_{JT} = \frac{X_{post} - X_{pre}}{\sqrt{2(DE_{norm}\sqrt{1 - r_{xx}})^2}}$$

Si el valor absoluto final del  $ICF_{JT}$  es igual o superior a 1,96, se podrá asumir que el cambio reflejado en la puntuación del paciente es fiable. El umbral de  $\pm 1,96$  se corresponde en una curva normal a un nivel de confianza de 95% (es el valor z que equivale a dos desviaciones estándar). Si no se alcanza ese umbral, no se podrá descartar que el cambio registrado se deba al error de la medida y, por tanto, no represente un cambio real.

Una vez evaluado el cambio clínico bajo el prisma de los dos pasos descritos (externo e interno), se podrá llegar a las siguientes conclusiones sobre el estado del paciente según cumplan sus criterios o no:

- No cambiado: cuando no se produzca un cambio estadísticamente fiable en el  $ICF_{JT}$ , independiente de si su puntuación ubique al paciente en una población funcional o disfuncional.
- Deteriorado: cuando se da un cambio hacia la disfuncionalidad que se detecta estadísticamente fiable mediante  $ICF_{JT}$ . Es decir la diferencia de las puntuaciones resulten en sentido inverso al esperado por efecto de la intervención.
- Mejorado: cuando se da un cambio hacia la funcionalidad que se detecta estadísticamente fiable mediante  $ICF_{JT}$ , pero que no llega a alcanzar la puntuación necesaria para salvar el punto de corte de funcionalidad, calculado mediante el método de comparación normativa.

- Recuperado: cuando se cumplen los criterios de ambos métodos (significativo y fiable). Se da un cambio fiable según  $ICF_{JT}$ , y la puntuación alcanzada después de la intervención permite ubicar al paciente en la población funcional

### **5.3.2. Metodologías alternativas al $ICF_{JT}$ .**

Aunque el método propuesto de Jacobson y Truax ha sido el procedimiento que más aceptación ha tenido para evaluar significación clínica en estudios de resultados de intervenciones psicológicas, varios autores han sugerido otras metodologías equivalentes para ese mismo propósito. Se ha incidido de forma especial sobre alternativas al propio  $ICF_{JT}$  como estimador preciso de cambio, es decir formulaciones alternativas en el criterio del método interno. A continuación se describen las principales alternativas propuestas en la literatura, y en la Tabla 13 se recogen sus formulaciones matemáticas. En su mayoría son modificaciones al  $ICF_{JT}$  que siguen tomando como referencia el error típico de medida aunque calculando sus elementos de distinta forma con el fin de obtener estimaciones más precisas (Lambert, Hansen y Bauer, 2008; Iraurgi, 2010; Pardo y Ferrer 2013).

El índice Gullisken-Lord Novick (GLN) propuesto por Hsu (1995) se basa en los trabajos de los matemáticos homónimos al índice, para poder controlar el efecto de regresión a la media, fenómeno no contemplado por la formulación  $ICF_{JT}$ . Este efecto hace referencia a que en las situaciones de medidas repetidas con un mismo instrumento, las puntuaciones más extremas inicialmente tienden a presentarse más cercanas a valores medios en las siguientes medidas. Por ello, incluye en la formulación los valores de la población de referencia a las que el paciente tendería hipotéticamente a volver (media y desviación estándar). Si no se disponen de datos de la población de referencia, el autor sugiere utilizar las puntuaciones medias pre tratamiento del conjunto de pacientes evaluados. Este mismo autor propuso previamente otro índice (Hsu, 1989) similar al GLN, denominado Hsu-Linn-Lord (HLL), que en vez de utilizar una puntuación de población de referencia, trata de controlar el efecto de regresión a la media incluyendo en la formulación los valores promedio del grupo tratado antes y después de la intervención.

El problema de la regresión a la media también fue señalado por Speer (1992), aunque su propuesta se basó en los desarrollos de otros metodólogos cuyo nombre utilizó también para denominar su índice. El método Edwards-Nunnally (EN) consiste en calcular un intervalo de

confianza del 95%, donde se encuentra la posible variación de las puntuaciones pre tratamiento. El cambio se considerará significativo si la puntuación post-tratamiento se sitúa fuera de ese intervalo. Otra propuesta previa a esta pero también basada en la consideración de la verdadera puntuación pre tratamiento (teniendo en cuenta su posible variabilidad), es el Índice Nunnally-Kotsch NK (Nunnally y Kotsch, 1983). Consiste en sustituir la puntuación pre tratamiento individual por una estimación de su valor verdadero (controlando el efecto de regresión a la media) realizada a partir de los valores promedios observados en el grupo de pacientes en tratamiento en el momento previo a la intervención. Y en vez de usar la fiabilidad test-retest utiliza la consistencia interna de cada momento.

La propuesta HA, homónima de Hagerman y Arrindal (1999a; 1999b), va más allá en su intento para mejorar la precisión del método propuesto por Jacobson y Truax (1991). No sólo propone una modificación a la ecuación de cambio fiabilidad de la diferencias de las pre tratamiento y post tratamiento ( $ICF_{HA}$ ), sino que también realiza un nuevo planteamiento para modificar el criterio de cambio significativo basados en puntos de corte por comparación normativa ( $ICS_{HA}$ ). En el  $ICF_{HA}$  se incorpora la fiabilidad de las medidas tanto el momento pre tratamiento ( $r_{xx-pre}$ ) como el momento post tratamiento ( $r_{xx-post}$ ) para calcular un coeficiente de fiabilidad de las diferencia de ambas puntuaciones ( $R_{dd}$ ). La  $r_{xx-pre}$  y  $r_{xx-post}$  se calculan a partir de:

$$r_{xx-pre} = \frac{DE_{pre}^2 - EEM^2}{DE_{pre}^2}, \text{ y } r_{xx-post} = \frac{DE_{post}^2 - EEM^2}{DE_{post}^2};$$

$$\text{siendo } EEM = DE_{norm} \sqrt{1 - r_{xx-norm}}$$

De forma equivalente a anteriores propuestas, realiza una corrección de la diferencia entre las puntuaciones pre tratamiento y post tratamiento, para controlar el efecto de regresión a la media basado en los promedios grupales observados. La puntuación resultante de esta formulación se contrasta frente al rango de valores de  $\pm 1,65$  (valor z que deja por debajo o por encima el 5% de las puntuaciones de la distribución). Si esta fuera de ese rango se considerará un cambio estadísticamente fiable (a peor o a mejor dependiendo del sentido del indicador), sino

no se podrá considerar que haya existido un cambio fiable. Con respecto al  $CS_{HA}$  se plantea una forma de calcular el verdadero punto de corte que delimita el nivel funcional (VPC) estimándolo con un promedio de las puntuaciones medias en función de la dispersión de muestras y ajustado también por la precisión de medida del instrumento. Una vez generado este VPC se puede calcular el índice  $ICS_{HA}$  del paciente en particular, que implicará la diferencia entre la puntuación post tratamiento corregida y el VPC, matizados también por la precisión de medida. Este índice también se contrastará frente al valor  $\pm 1,65$  para determinar si el paciente supera el umbral de funcionalidad, y el cambio se puede considerar significativo. Aunando los resultados de los índices  $ICF_{HA}$  y  $ICS_{HA}$ , se llega a la misma clasificación de estados del paciente, con respecto a su cambio terapéutico, que con el procedimiento de Jacobson y Truax (no cambiado, mejorado, deteriorado y recuperado). Los autores también ofrecen otras formulaciones de estos mismos índices destinados a realizar su cálculo pero a nivel grupal en vez de individual. Un elemento importante a destacar es que se exige que la  $r_{dd}$  sea mínimo de 0.40 para poder aplicar el método correctamente.

El grupo de Wyrwich (Wyrwich, Tierney y Wolinsky, 1999; Wyrwich 2004) desarrolla un índice bastante similar al  $ICF_{JT}$  y en el cual también se calcula un error típico de medida. En esta exposición lo denominaremos según la acepción que le dan Pardo y Ferrer (2013): Diferencia Tipificada de Wyrwich DTW. El error se calcula a partir de la variabilidad y fiabilidad de las puntuaciones de la escala. Los autores señalan de forma específica la aplicación coeficiente de fiabilidad de Cronbach a las puntuaciones de una muestra de referencia o en su lugar a las puntuaciones pre tratamiento del grupo evaluado. El valor del DTW se comparará con el criterio de valor absoluto  $\geq 1,96$  (nivel de confianza del 95%), para comprobar si es un cambio fiable, como en el caso del  $ICF_{JT}$ . Aunque se proponen otros dos posibles criterios: valor absoluto  $\leq 1$  que se corresponde a un efecto de tamaño medio cuando la fiabilidad vale 0,75, y otro más exigente de valor absoluto  $\leq 2,77$  (multiplica 1,96 por la raíz cuadrada de dos para compensar que se utilizan dos momentos de medida).

Otro conjunto de propuestas alternativas al  $ICF_{JT}$  se vertebran a partir de la tipificación de las diferencias individuales entre la medida post y la medida pre tratamiento, y cuya interpretación se realiza bajo los mismos criterios establecidos por Cohen (1988) utilizados para valorar el tamaño del efecto de la diferencia de dos medias grupales (0,20 efecto leve, 0,50

efecto moderado y 0,80 efecto grande). El Tamaño del Efecto Individual (TEI) divide la diferencia de las puntuaciones post y pre tratamiento entre la desviación típica de las puntuaciones del grupo tratado en el momento pre tratamiento, interpretándose cambio significativo valores absolutos mayores que el umbral de efecto leve de Cohen (Kazis, Anderson y Meenan, 1989). La Diferencia Individual Tipificada (DIT) divide la diferencia de las puntuaciones pre tratamiento y post tratamiento entre la desviación típica de esas diferencias en el grupo tratamiento (Stucki, Liang, Fossel y Katz, 1995). Otro índice denominado Estadístico de Sensibilidad de Guyatt (ESG) presenta la misma formulación que el DIT pero la desviación típica de las diferencias que se utiliza debe ser la de un grupo control en donde no se espera cambio sistemático debido a que no son tratados (Guyatt, Bombardier y Tugwell, 1986).

Por su parte Crawford y colaboradores (Crawford y Howell, 1998; Crawford y Garthwaite, 2006) exponen una alternativa que no se basa directamente en la diferencia de puntuaciones sino en la comparación de la puntuación observada con un pronóstico de la puntuación post tratamiento en situación de no cambio. Se genera al realizar un análisis de regresión lineal de las puntuaciones post tratamiento sobre las puntuaciones pre tratamiento de una muestra de referencia donde no se ha producido un cambio real (grupo control). Con este pronóstico se crea un intervalo de confianza cuyos límites van a demarcar el rango de puntuaciones post tratamiento entre las cuales no se puede considerar un cambio significativo. Sí será un cambio significativo si la puntuación post tratamiento observada del paciente está fuera de ese intervalo. Los autores describen dos métodos con esta misma premisa pero que se diferencian en cómo calcular el error estándar del intervalo de confianza. El denominado Intervalo para Pronóstico Promedio IPP se genera bajo el supuesto de normalidad de la distribución de pronósticos del grupo de referencia. De una forma más restrictiva el denominado Intervalo para el Pronóstico Individual IPI se establece a partir de una distribución  $t$ , y el cálculo del error típico genera un valor de una magnitud mayor de su valor que con el IPP.

La propuesta de Hiller y Schindler (2011) es utilizar el Criterio de Mejoría Sintomática del 50% (CPM50%), ya que es un valor consensuado en investigación psicofarmacológica para señalar relevancia clínica. Para ello, a este criterio lo añaden con la reflexión Jacobson y Truax (1991) acerca del Punto de Corte con la Funcionalidad PCF, que utilizan para relativizar el cambio acontecido en el indicador de resultado. De esta forma se calcula una puntuación del

rango de mejoría posible, restando a la puntuación pre tratamiento el PCF. El valor resultante será el denominador que habrá que aplicar a la diferencia entre los dos momentos de medida. Si el resultado de esta división es mayor que 0,50 (50%) se podrá considerar el cambio de importancia clínica. Este método se establece bajo el supuesto de que los pacientes con puntuaciones iniciales más alejadas de la funcionalidad (mayor gravedad sintomática), si mejoran, lo harán siempre en mayor magnitud que los que empiecen con valores más cercanos a la funcionalidad, independientemente de los efectos reales del tratamiento. Los autores critican al  $ICF_{JT}$ , la linealidad de su criterio y no tener en cuenta la gravedad inicial a la hora de valorar el cambio. Por ello, el CPM 50% se muestra más exigente para clasificar cambio relevante con los puntuaciones iniciales altos, que con las puntuaciones cercanas a PCF. Sin embargo, añaden un criterio adicional para compensar lo excesivamente permisivo que resulta el método con puntuaciones adyacentes al PCF: además de la condición principal siempre tiene que existir una mejoría total mínima del 25% sobre la puntuación pre-tratamiento (sin tener en cuenta el PCF) para poder considerar cambio relevante. Es necesario tener en cuenta que el método se desarrolló únicamente para detectar cambio sintomatológico en sentido de mejoría, obviando cambios en sentido de empeoramiento u otro tipo de medidas de resultado.

Desde una perspectiva muy diferente a las anteriores se plantea el método denominado HLM (Hierarchical Linear Modeling), que se basa en las curvas de crecimiento generadas en los modelos jerárquicos lineales (Byrk y Raudenbush, 1992; Speer y Greenbaum, 1995). Para determinar el cambio del individuo se utiliza un procedimiento bayesiano complejo que permite realizar estimaciones incluso en caso donde existen datos ausentes. La utilización de las estimaciones empíricas de Bayes permite combinar en el análisis los datos acerca de la evolución tanto del individuo como del grupo. Sin embargo, el procedimiento requiere de la recogida de puntuaciones en tres momentos diferentes como mínimo. No es suficiente comparar solo la medida post-tratamiento con respecto a la medida pre-tratamiento.

**Tabla 13**

Formulación de principales alternativas a  $ICF_{IT}$  para el análisis del cambio individual en psicología

Método		Cálculo	
GLN/HLL	Gullisken-Lord-Novick / Hsu-Linn-Lord	$GLN = \frac{(X_{post} - \bar{X}_{norm}) - r_{xx(norm)}(X_{pre} - \bar{X}_{norm})}{DE_{norm}\sqrt{1 - r_{xx(norm)}^2}}$	$HLL = \frac{(X_{post} - \bar{X}_{post}) - r_{xx}(X_{pre} - \bar{X}_{pre})}{DE_{pre}\sqrt{1 - r_{xx(norm)}^2}}$
EN	Edward-Nunnally	$EN = [r_{xx(norm)}(X_{pre} - \bar{X}_{pre}) + \bar{X}_{pre}] \pm 2DE_{pre}\sqrt{1 - r_{xx(norm)}}$	
NK	Nunnally-Kotsch	$NK = \frac{X_{post} - [r_{xx(pre)}(X_{pre} - \bar{X}_{pre}) + \bar{X}_{pre}]}{\sqrt{[r_{xx(pre)}^2 DE_{pre}^2 (1 - r_{xx(pre)})] + [DE_{pre}^2 (1 - r_{xx(post)})]}}$	
HA: $ICF_{HA}$	Hagemann-Arrindell Índice Cambio Fiable	$ICF_{HA} = \frac{(X_{post} - X_{pre}) r_{dd} + (\bar{X}_{post} - \bar{X}_{pre})(1 - r_{dd})}{\sqrt{r_{dd}} \sqrt{2 EEM^2}}$	$r_{dd} = \frac{DE_{pre}^2 r_{xx(pre)} + DE_{post}^2 r_{xx(post)} - 2DE_{pre}DE_{post} r_{xx(pre*post)}}{DE_{pre}^2 + DE_{post}^2 - 2DE_{pre}DE_{post} r_{xx(pre*post)}}$
$ICS_{HA}$	Índice Cambio significativo	$ICS_{HA} = \frac{\bar{X}_{post} + (X_{post} - \bar{X}_{post}) r_{xx(post)} - VPC}{\sqrt{r_{xx(post)}} EMM}$	$EEM = DE_{norm}\sqrt{1 - r_{xx(norm)}}$ $VPC = \frac{\bar{X}_{pre}(DE_{norm}\sqrt{r_{xx(norm)}}) + \bar{X}_{norm}(DE_{pre}\sqrt{r_{xx(pre)}})}{DE_{norm}\sqrt{r_{xx(norm)}} + DE_{pre}\sqrt{r_{xx(pre)}}$
DTW	Diferencia tipificada de Wyrwich	$DTW = \frac{X_{post} - X_{pre}}{DE_{pre}\sqrt{1 - r_{xx(norm)}}$	

Nota:  $X_{post}/X_{pre}$ = puntuación post-tratamiento/pre-tratamiento del paciente;  $\cdot_{post}/\cdot_{pre}$ = media de puntuación post-tratamiento/pre-tratamiento del grupo evaluado;  $DE_{post}/DE_{pre}$ = desviación estándar grupo evaluado pre-tratamiento/post-tratamiento;  $\cdot_{norm}$  = media grupo normativo  $DE_{norm}$ = desviación estándar grupo normativo; n= tamaño de grupo evaluado;  $r_{xx(pre)}/r_{xx(post)}/r_{xx(norm)}$  = fiabilidad del momento post tratamiento, pre tratamiento o del grupo normativo;  $r_{xx(pre-post)}$ = coeficiente de fiabilidad test-re-test;  $r_{dd}$ =coeficiente de fiabilidad de las diferencia de puntuaciones; EEM= Error estándar de medida. VPC= Verdadero punto de corte del nivel funcional.

**Tabla 13**

Formulación de principales alternativas a  $ICF_{IT}$  para el análisis del cambio individual en psicología (Continuación)

	Método	Cálculo	
TEI/DIT/ESG	Tamaño del Efecto Individual /Diferencia Individual Tipificada /Estadístico de Sensibilidad Guyyatt	$TEI = \frac{X_{post} - X_{pre}}{\sqrt{\frac{\sum(X_{pre} - \bar{X}_{pre})^2}{n - 1}}}$	$DIT/ES = \frac{X_{post} - X_{pre}}{\sqrt{\frac{\sum[(X_{post} - X_{pre}) - (\bar{X}_{post} - \bar{X}_{pre})]^2}{n - 1}}}$
IPP/ IPI	Intervalo para el Pronóstico Promedio/Intervalo para el Pronóstico Individual	$ICPP = X_{pron} \pm Z_{\alpha/2} \sqrt{MCE \left[ \frac{1}{n} + \frac{(X_{pre} - \bar{X}_{pre})^2}{\sum(X_{pre} - \bar{X}_{pre})^2} \right]}$	$ICPI = X_{pron} \pm (t_{n-2;\alpha/2}) \sqrt{MCE \left[ 1 + \frac{1}{n} + \frac{(X_{pre} - \bar{X}_{pre})^2}{\sum(X_{pre} - \bar{X}_{pre})^2} \right]}$
CPM 50%	Criterio de Porcentaje de mejoría 50%	$CPM50\% = \frac{X_{pre} - X_{post}}{X_{pre} - PCF} \geq 0,50$	$CPMTmin25\% = \frac{X_{pre} - X_{post}}{X_{pre}} \geq 0,25$
HLM	Hierarchical Linear Modeling	$HLM = \frac{B^*}{\sqrt{V^*}}$	

Nota:  $X_{post}/X_{pre}$ = puntuación post-tratamiento/pre-tratamiento del paciente;  $\bar{x}_{post}/\bar{x}_{pre}$ = media de puntuación post-tratamiento/pre-tratamiento del grupo evaluado;  $DE_{post}/DE_{pre}$ = desviación estándar grupo evaluado pre-tratamiento/post-tratamiento;  $n$ = tamaño de grupo evaluado;  $X_{pron}$ = pronóstico de la puntuación  $X_{post}$  basado en la regresión lineal de la puntuación post sobre la puntuación pre en una muestra de referencia donde no exista cambio sistemático;  $MCE$ = media cuadrática del error, varianza de los residuos de la anterior regresión lineal;  $PCF$ = Punto de Corte con Funcionalidad  $CPM_{min25\%}$ = Criterio de porcentaje de mejoría total mínima 25%;  $B^*$ = Estimación empírica de Bayes;  $V^*$ =desviación estándar de la estimación empírica de Bayes.

Debido a esta proliferación de propuestas, se han realizado diversos estudios que comparar los diferentes métodos y su comportamiento ante los datos clínicos. El primero de estos estudios es el realizado por Speer y Greenbaum (1995) donde comparan: ICF<sub>JT</sub>, EN, HLL, NK y HLM en base únicamente a la clasificación según cambio fiable (mejorado, deteriorado o no cambio), en una muestra clínica heterogénea de 73 adultos. Se observó un relativo acuerdo entre los distintos métodos a la hora de detectar el cambio, a excepción del método HLL que se mostró muy divergente con respecto a los demás (con mayores tasas de deterioros y mucho menores de mejorías). Con respecto al método HLM se constató que resultaba más sensible que los demás métodos para clasificar pacientes como mejorados pero menos que los mismos para clasificar pacientes como deteriorados. Los autores concluyen recomendando el método HLM siempre y cuando la detección de deterioros no sea el objetivo del estudio. Aunque, en caso de no poder usar este método (por ejemplo por contar solo con dos medidas), señalan al ICF<sub>JT</sub> como preferente al haber constatado que sus alternativas computacionalmente más complejas no se muestran más sensibles que él.

Bajo el mismo planteamiento del estudio anterior McGlinchey, Atkins y Jacobson (2002) realizan una contrastación de métodos en base a los datos de una muestra de 129 pacientes tratados por depresión. En esta ocasión se tuvo en cuenta la clasificación del estado del paciente según la doble perspectiva de cambio estadísticamente fiable y clínicamente significativo (mejorado, deteriorado, no cambio o recuperado). Se volvió a comparar cinco métodos, sin embargo, con respecto al anterior estudio, se sustituyó los métodos NK y HLL por los métodos GLN y HA por ser enfoques más recientes y actualizados de las mismas perspectivas (justificación que perdura en los subsiguientes estudios). También se realizó un seguimiento de dos años para valorar la validez predictiva de cada método con respecto a posteriores recaídas (según criterio de recuperación o no recuperación). Se encontró que existían diferencias significativas en la tasas de clasificación de cada método. El método HA resultó ser el más conservador y el que más se diferenciaba del resto. El método ICF<sub>JT</sub> no difirió significativamente con respecto a los métodos GLN y HLM, pero sí con respecto al método HA y EN. Sin embargo, todos los métodos mostraron una sensibilidad equivalente a la hora de predecir recaídas, discriminando significativamente entre los pacientes que recayeron y los que no lo hicieron. Los autores concluyeron que a efectos prácticos, los métodos alternativos al ICF<sub>JT</sub> analizados no se mostraron superiores a éste.

Bauer, Lambert y Nielsen (2004) vuelven a poner a prueba los 5 métodos analizados en el estudio anterior con los resultados de 386 pacientes (de diagnósticos diversos) tratados en un servicio de atención psicológica universitario. El porcentaje de acuerdo entre la clasificación de los diferentes métodos oscila entre el 71% a 85%. El acuerdo más fuerte fue entre los métodos ICF<sub>JT</sub> y GLN que prácticamente no se diferencian. El método EN resultó el más permisivo para clasificar a los sujetos como cambiados después del tratamiento. Y el método HA se volvió a reafirmar como el más conservador (exige mayor magnitud de cambio para clasificarlo como tal). Los autores alcanzan unas conclusiones similares al estudio de McGlinchey, Atkins y Jacobson (2002).

En vez de utilizar datos empíricos Atkins , Bedics , McGlinchey y Beauchaine (2005) generan un estudio con datos simulados para poder explorar de forma sistemática el comportamiento de los métodos ante la variación de diferentes parámetros relevantes (tales como distintos tamaños del efecto o niveles de fiabilidad). Se valora el desempeño de los métodos: ICF<sub>JT</sub>, GLN, ES y HA. Los resultados mostraron un alto nivel de acuerdo entre todos los métodos, especialmente en las situaciones en la que la medida tuviese una elevada fiabilidad. Aunque las similitudes entre métodos superan sus diferencias en la mayoría de las situaciones los autores vuelven a corroborar las tendencias de los estudios anteriores: clasificaciones idénticas en ICF<sub>JT</sub> y GLN, más facilidad para ser clasificado como cambiado bajo método EN (estimación de un menor error estándar) y más dificultad para serlo bajo el método HA. Adicionalmente se observó que el criterio de una  $R_{adj}$  mínima sugerida en el método HA no se cumple en muchas de las situaciones simuladas. Esto puede problematizar la aplicación práctica de dicho método pese a su precisión.

En un estudio más reciente, Ronk, Hooke y Page (2012) utilizaron los datos de 2676 pacientes diagnosticados de trastornos del ánimo y tratados en un centro privado de psiquiatría. Se compararon las tasas de clasificación de los métodos: ICF<sub>JT</sub>, NK, GLN, ES y HA. Y además se realizó en base tres medidas de resultado diferentes de auto informe: sintomatología depresiva, nivel de salud mental general, y calidad de vida. Los métodos estadísticos mostraron un nivel muy alto de acuerdo en su clasificación, siendo HA el que más difería con respecto a los demás. Sin embargo, fueron las diferentes medidas de resultado las que generaron más discrepancia en la clasificación del cambio del paciente.

Los autores concluyen resaltando la importancia de detallar la medida de resultado y el constructo que se pretende medir en los estudios de resultados, y como esto tiene más relevancia para valorar CCS que el método utilizado. Este resultado es congruente con estudios anteriores como el de Beckstead y colaboradores (2003) que compararon las tasas de clasificación en una misma muestra de pacientes, utilizando exclusivamente el método ICF<sub>JT</sub>, pero con respecto a cinco medidas de resultado diferentes (instrumentos de auto reporte) alcanzando solo un 65% de acuerdo en la clasificación entre medidas.

En relación al CPM 50%, Hiller, Schindler y Lambert (2012) realizan un estudio con los resultados de 395 pacientes depresivos, cuyo cambio fue evaluado a través de un instrumento psicométrico de intensidad de sintomatología depresiva. Resultó una situación de mejoría general muy grande (gran tamaño del efecto grupal). Compararon el comportamiento de clasificación del método ICF<sub>JT</sub> con el CPM 50%. Se observó que el método CPM 50%, clasificó mayor tasa de pacientes con cambios significativos que el ICF<sub>JT</sub>, caracterizándose como más permisivo que este último. Pese a ello, la diferencia de tasas de clasificación no superó el 7%. Dado la diferente naturaleza de los indicadores, los autores concluyen que ambos métodos son complementarios, y que más que una alternativa el CPM 50 % representa un interesante extensión al ICF<sub>JT</sub>.

De nuevo, a través de simulación de datos por ordenador, Pardo y Ferrer (2013) evalúan el comportamiento de los métodos: TEI, DIT DTW, ICF<sub>JT</sub>, GLN, HA, EN, IPP e IPI. En este estudio sólo se valoró la capacidad de los métodos para discriminar falsos positivos respecto al cambio fiable. Se tuvieron en cuenta tres tamaños muestrales diferentes generando 1000 muestras ficticias para cada uno (n= 25, 50 y 100). El escenario simulado fue una situación donde no existe cambio sistemático entre el momento pre y el momento post. Calculan una puntuación post tratamiento mediante una combinación de la puntuación pre y una variable aleatoria de la distribución normal, asumiendo un nivel de correlación entre ambos momentos es de 0,80. En este contexto, lo esperable es que los diferentes métodos señalaran que no se había producido cambio real. Una tasa de falsos positivos mayores de lo esperable por azar indicaría que el estadístico utilizado tiende a identificar más de lo que debería como significativos cambios que son en realidad variaciones aleatorias. Sobre los resultados los autores señalan que los métodos TEI, EN e IPP presentan tasas de falsos positivos inaceptablemente altas. Los métodos ICF<sub>JT</sub> y GLN presentan tasas menores aunque más altas de lo esperable para cada tamaño muestral (alrededor de un 16%). La DIT ofrece

resultados también inadmisibles si se utilizan los criterios sugeridos para este método (los umbrales de Cohen). Sin embargo, los autores de este estudio aplican el mismo criterio que en el  $ICF_{JT}$  ( $\geq |1,96|$ ) y observan que con esta consideración el método es uno de los que menos tasas de falsos positivos tiene. El método DTW presenta tasas demasiado altas en sus diferentes criterios, exceptuando bajo su criterio más estricto ( $\geq |2,77|$ ) que se asemeja a las del  $ICF_{JT}$ . El método HA genera tasas de falsos positivos muy bajas, pero lo hace a costa de descartar alrededor de un cuarto de las muestras ficticias (debido a su condición de  $R_{ad}$  mínima). El método IPI también muestra tasas muy bajas destacándose como uno de los procedimientos que mejor funcionan en este aspecto; sin embargo, los autores subrayan una propiedad no deseable para el propósito de valorar el cambio individual: el cálculo de los límites del intervalo depende del tamaño muestral.

En base a estos estudios realizados con las diferentes alternativas al  $ICF_{JT}$ , varios académicos han llegado a una serie de conclusiones acerca de la idoneidad de este procedimiento para evaluar CCS (Wise 2004, Lambert, Hansen y Bauer 2008; Lambert y Ogles 2009; Gómez y Roussos, 2012):

- El efecto de regresión a la media no parece tener mucho impacto a la hora de clasificar al paciente en los estudios realizados.
- El método  $ICF_{JT}$  se ha mostrado como moderado en exigencia a la hora de clasificar el cambio en comparación a otros métodos alternativos.
- Los métodos que se han mostrado más precisos que el  $ICF_{JT}$  además de una mayor complejidad computacional, necesitan de más datos de referencia y/o de unas condiciones de aplicación que no siempre se podrán dar en el contexto asistencia clínica habitual.
- En varios estudios el  $ICF_{JT}$  ha mostrado un comportamiento similar y tasas de clasificación semejantes a otros métodos alternativos pero con una mayor facilidad de aplicación y cálculo
- El  $ICF_{JT}$  es el procedimiento para valorar CCS con mayor difusión y aceptación de los considerados en los estudios anteriores.
- Por todo lo anterior, se puede postular el  $ICF_{JT}$  como el método unificado de referencia en la evaluación del CCS.

- Se señala necesario la realización de estudios empíricos que en lugar de buscar el mejor método de valoración de forma genérica, se centre en determinar la utilidad de cada método para estudiar la CCS en las diferentes problemáticas o áreas de salud y ante diferentes medidas de resultado.
- Son las diferentes medidas de resultado las que generan mayor discrepancia en la valoración del CCS. Por lo que estudiar, desarrollar y detallar esas medidas debe ser prioritario para la investigación de resultados (son necesarias herramientas válidas y fiables).



## Capítulo 6

### Problemáticas de salud implicadas en el desarrollo empírico

---

En este último capítulo de la fundamentación teórica, se presentan las problemáticas de salud concretas sobre las que se ha intervenido en la parte empírica del trabajo, afrontando su conceptualización, epidemiología y diagnóstico. Inicialmente se explicita el modelo de salud general que se ha asumido en el desarrollo de la tesis y se señalan sus implicaciones básicas. A continuación se presentan las dos condiciones psicopatológicas, implicadas en los tres primeros estudios: ansiedad y depresión. Para finalmente definir el síndrome físico del Asma, (el trastorno crónico eje del cuarto estudio) junto a los factores psicológicos implicados en el curso y tratamiento del mismo.

#### 6.1. Definición de Salud y modelo biopsicosocial

En el modelo de salud biomédico tradicional se conceptualizó la salud en sentido negativo, es decir, como un estado de ausencia de enfermedad. La enfermedad se definió como un desorden del equilibrio natural del organismo (homeostasis) con una naturaleza cualitativamente diferente al estado normal del mismo, y con una causalidad biológica, lineal y unívoca. Esta concepción dicotómica de salud-enfermedad ha sido útil en el pasado para el control de multitud de problemas de salud (en especial los de tipo agudo o infeccioso) pero se ha mostrado obsoleta como modelo general ante la realidad contemporánea (como ya se explicitó en el capítulo 4). Desde la fisiopatología se señala que la diferenciación entre procesos de salud y enfermedad es una cuestión cuantitativa en vez de cualitativa. Es decir que pertenecen a un mismo continuo donde es difícil ubicar la frontera entre una y otra. Y desde una perspectiva sistémica también se advierte de la naturaleza múltiple e interactiva de los factores determinantes en el continuo salud-enfermedad, no solo en la génesis y mantenimiento de posibles problemáticas sino también en la propia interpretación social y personal de los estados de salud y enfermedad. Por lo que son necesarios enfoques de salud integradores y más complejos que el propuesto por el modelo tradicional (Morrison y Bennet,

2008; Brannon y Feist, 2009; Varela, 2010).

La Organización Mundial de la Salud, desde su carta fundacional, asume una definición de salud que explícitamente hace mención a la superación del modelo clásico de salud en sentido negativo (OMS, 1948 p.1): “*La salud es un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades*”. Se han planteado diversas críticas a esta definición tales como: el hecho de tratar la salud como un concepto estático (cuando es un fenómeno dinámico), hacer referencia al bienestar como un término universal y objetivo (y sin además afrontar la definición del mismo) o matizar de este último bajo el adjetivo “completo” resultando ser demasiado restrictivo y utópico (Navarro, 1998; Huber et al., 2011). A pesar de ello, esta definición ha tenido un gran impacto y aceptación en el ámbito sanitario, manteniendo vigencia y repercusión internacional en las políticas de salud de los diferentes países. Se reconoce como un hito por su apuesta por la visión de la salud como un fenómeno multidimensional, que implica tener en cuenta más factores además de los inmediatamente biológicos (Alcantara, 2008).

En psicología se han asumido diversos modelos de salud, con el fin de contextualizar las diferentes intervenciones en el ámbito sanitario, los cuales enfatizan aspectos concretos tales como: resiliencia (Fiorentino, 2008), inteligencia emocional (Augusto y Montes 2009), bienestar percibido (Kahn y Juter, 2002), recursos ambientales de afrontamiento (Antanovsky, 1991), holismo humanista (Ventegodt, Andersen y Merrick, 2003). Los distintos modelos difieren en los elementos que se acentúan y pueden ser más o menos adecuados para cada situación particular. Sin embargo, todos coinciden en señalar la necesidad de integración de múltiple factores para abordar el fenómeno de la salud, e identifican a esta con la presencia de unas características positivas de funcionalidad o capacidad que permitan al individuo adaptarse al medio ambiente y satisfacer sus necesidades vitales (Juárez, 2011).

En la presente tesis se ha optado por interpretar la salud desde el modelo denominado Biopsicosocial propuesto originariamente por Engel (1977). Este autor, a su vez, fundamenta su modelo en la Teoría General de los Sistemas del biólogo Von Bertalanffy (1976). Esta meta-teoría (aplicable a cualquier campo de investigación) subraya la necesidad de estudiar no sólo las partes concretas o procesos aislados de los fenómenos complejos (como el propio ser humano), sino también como estos se organizan e interactúan. Ya que

el comportamiento de los elementos que componen un fenómeno será diferente si se toma de forma aislada que dentro de un todo. Según este planteamiento todos los fenómenos naturales son un conjunto de sistemas interrelacionados y deben estudiarse como tales.

El modelo Biopsicosocial recoge esta consideración de complejidad interactiva para abordar el continuo salud-enfermedad, tratando de superar las explicaciones lineales de causa efecto en salud que solo atienden a unos condicionantes biológicos concretos (Engel, 1997). Además se traslada el foco de estudio de la enfermedad a la persona y a su contexto. Proponiendo un paradigma de intervención donde la persona no es un sujeto pasivo ante la enfermedad, sino un agente activo y determinante de su propio estado de salud y señalando la relevancia de iniciativas y conductas promotoras de la salud y preventivas de la enfermedad (Engel, 1980; León-Rubio, 2004). El modelo hace mención a tres grupos de factores principales que interactúan entre sí en la determinación del estado de salud del individuo: factores psicológicos (personalidad, afrontamiento, hábitos de vida, afectividad etc...), factores sociales (interacción familiar, contexto laboral, cultura, contexto macrosocial etc..) y factores biológicos (genética, fisiología, anatomía, condiciones bioquímicas, agentes infecciosos etc..). Se entiende el efecto de estos factores no de una forma aditiva sino interactiva y bidireccional, es decir el desequilibrio en uno de los niveles puede desequilibrar a los demás (Morrison y Bennett, 2008). El organismo se concibe como un sistema cuyas propiedades emergen de la interacción entre sus partes y de este con el entorno, por lo que se niega la conceptualización dualista de mente y cuerpo, ubicando al individuo como un todo integrado (Mira-Engo, 1990). Y el estado de salud se considera como un fenómeno dinámico (cambiante en el tiempo), resultado del proceso de adaptación del individuo al ecosistema con la interacción de los ya mencionados factores biopsicosociales (Santacreu, 1991).

## **6.2. Ansiedad**

Se denomina ansiedad a un estado de activación del sistema nervioso central, en el cual el organismo se prepara para enfrentarse a situaciones de peligro o incertidumbre. Se considera una respuesta emocional normal ante la percepción subjetiva de amenaza que permite al individuo adaptarse al entorno y prepararse para generar un comportamiento de afrontamiento eficaz. El estímulo que provoca esta respuesta ansiosa puede ser tanto externo (presente el medioambiente) o interno, como puede ser un pensamiento o una imagen

mental. Sin embargo, si esta respuesta es desproporcionada (demasiado intensa, frecuente o duradera) con respecto las situaciones a la que tiene que enfrentar, se considera patológica y desadaptativa. (Conde, Orozco, Báez y Dallos, 2009; Martínez, Sáiz y García-Portilla, 2008).

### **6.2.1. Aclaraciones conceptuales sobre la ansiedad**

Debido a las distintas traducciones en cada idioma y a las implicaciones históricas y teóricas de la ansiedad, se la ha asociado a una serie de términos relacionados, llegando a generar confusión y ambigüedad en el concepto, siendo los siguientes los más relevantes: angustia, miedo y estrés (Sierra, Ortega y Zubeidat, 2003; Sandín y Chorot, 2008).

La palabra alemana *Angst* (malestar y constricción) fue originalmente utilizada en el marco psicoanalista para referirse al afecto negativo y activación fisiológica desagradable presente en los trastornos neuróticos. Su traducción, tanto al español como al francés, generó dos términos distintos: ansiedad y angustia. El primer termino hacía referencia al mal estar o inquietud psíquica, es decir una predominancia de síntomas psicológicos, mientras que el segundo se referiría a la predominancia de síntomas fisiológicos. En nuestro país se ha terminado por integrar ambos términos, hablando de diferentes componentes dentro ansiedad en vez de términos separados (Sandín y Chorot, 2008).

También se ha asociado la ansiedad con el concepto de miedo, diferenciándose teóricamente en que este último consiste en una perturbación debida a la presencia de estímulos presentes inmediatos y a reacciones automáticas y de carácter más biológico. Mientras que la ansiedad se produciría más bien por una anticipación menos focalizada y difusa de peligro o amenaza futura y con una vivencia de aprensión psicológica. A nivel práctico la frontera entre ansiedad y miedo no es fácil de ubicar, y para sus implicaciones clínicas se utilizan de forma equivalente, de hecho las fobias (miedo patológico) se consideran dentro de los trastornos de ansiedad (Cruz-Morales; González-Reyes, Gómez-Romero y Arriaga, 2003; Sierra, Ortega y Zubeidat, 2003).

Por otro lado, el termino estrés que se utiliza con frecuencia como sinónimo de ansiedad, es en realidad un concepto más global que ésta. Se entiende estrés como un proceso de adaptación del organismo antes unas determinadas demandas ambientales (activándose para movilizar recursos), el protagonismo recae sobre todo en la situación y la valoración de

la persona para afrontarla. Aunque la ansiedad puede ser una respuesta posible al estrés, no es la única posible. Además mientras que el estrés se entiende como una respuesta a la situación que se vive, la ansiedad no necesita de estímulos externos para desencadenarse cobrando más peso la anticipación y la aprensión psicológica no focalizada que el entorno en sí mismo (Lazarus y Folkman, 1986, Ortega y Zubeidat, 2003; Sandin, 2008).

Otra matización conceptual respecto a la ansiedad, es la propuesta por Spielberg (1972) que distingue entre ansiedad-rasgo y ansiedad-estado. La primera se refiere a al tipo de tendencia estable en la persona hacia las respuestas ansiosas (estando más o menos predispuesto), mientras que la segunda identifica al estado emocional transitorio de ansiedad en un momento dado; y por tanto variable y modificable en el tiempo. El autor planteó una independencia conceptual entre ambos términos pero también señaló la evidente interacción existente entre ellos. Un alto nivel de ansiedad-rasgo genera que la persona tienda a percibir y evaluar un rango mayor de situaciones como amenazantes, por lo que tendrá una predisposición mayor a padecer ansiedad-estado elevada con mayor frecuencia. La ansiedad como rasgo estable de personalidad se infiere, por tanto, de la frecuencia con que una persona vivencia aumentos en su estado de ansiedad. Con posterioridad, se desarrolló un instrumento psicométrico con el fin de permitir medir ambas dimensiones denominado State-Trait Anxiety Inventory “STAI” (Spielberg, Gorsuch y Lushene, 1982). La validez de estos constructos y la relación demarcada por Spielberg ha recibido apoyo empírico y actualmente esta noción ampliamente utilizada en el estudio de la ansiedad (Ries, Castañeda, Campos y Del Castillo, 2012).

Por último, en lo que concierne a la respuesta ansiosa existe consenso en señalar que no se trata de un fenómeno unitario y que presenta 3 subsistemas o componentes que pese a estar relacionados mantienen cierta autonomía entre sí (Pallares, 2002; Conde; Sandín y Chorot, 2008; Orozco, Báez y Dallos, 2009; García-Fernández, Inglés, Martínez-Monteaquedo, Marzo y Estévez 2011). Es lo que se conoce como modelo tridimensional de la ansiedad propuesto originalmente por Lang (1968), que identifica las siguientes dimensiones de respuesta:

- Cognitiva: hace referencia a la experiencia subjetiva (percepción y valoración) de los estímulos y estados relacionados con la ansiedad, ubicándose en ella las experiencias de preocupación, hipervigilancia, temor, malestar interior, irritabilidad, sensación de

pérdida de control, pensamientos intrusivos y pensamientos de anticipación de peligro; junto a déficits cognitivos asociados a esta tensión psicológica tales como: dificultades en concentración, bloqueo o obstaculización del pensamiento y la toma de decisiones, abundantes sesgos perceptivos y atribuciones disfuncionales.

- Fisiológica: consiste en la respuesta biológica de la ansiedad y se traduce en un aumento de la actividad del sistema nervioso autónomo, que prepara al organismo para afrontar un peligro. Se puede apreciar una serie de signos fisiológicos tales como: incremento de tensión muscular, dilatación de las pupilas, sudoración, temblores, aceleración cardíaca y respiratoria, palpitaciones, disminución de la salivación, presión torácica, dolores tensionales, molestias digestivas, náuseas y vómitos, disfunciones sexuales... Y con posterioridad a la aceleración generalizada: bajada de tensión y disminución de tasa cardíaca muy acentuadas que pueden producir mareos y desmayos.
- Motora: se compone de los elementos observables en la conducta de la persona en la respuesta ansiosa. Sus manifestaciones más relevantes son las conductas instrumentales de evitación tanto directas (escape o huida de un estímulo o situación) como indirectas (consumo de drogas inhibitorias o realización de conductas compulsivas). También se pueden dar reacciones de paralización motora que en rara ocasión serán totales, pero sí con manifestaciones parciales como dificultades en la expresión verbal o disminución en la destreza motora. Así como las consecuencias directas sobre el ámbito motor de las otras dos dimensiones (expresión facial, posturas corporales, hiperactividad...).

Estos tres subsistemas se influyen mutuamente, pudiendo representar una respuesta de uno como estímulo para el otro (Antony y Swinson 2000), pese a ello se ha constatado experimentalmente que estas dimensiones pueden no covariar entre sí, presentándose esa posible disociación entre tipos de respuesta tal y como se enunciaba teóricamente (Wilhelm y Roth, 2001; Ollendick et al. 2011). De hecho, se ha llegado a profundizar en los diferentes perfiles de respuesta ansiosa, dependiendo de qué dimensiones predominan en cada caso y la relevancia de estos para la práctica clínica (Espada, Torres, Cano-Vindel y Ochoa 1994). También se han validado instrumentos psicométricos que atienden de forma específica a los tres sistemas de respuesta, como es el caso del Inventario de Situaciones y Respuestas de Ansiedad ISRA (Cano-Vindel y Miguel-Tobal, 1999). Actualmente se reivindica la

importancia de esta noción de la ansiedad para su evaluación e intervención, recibiendo abundante apoyo empírico desde diferentes ámbitos de la psicología (García-Fernández, Inglés, Martínez-Monteagudo, Marzo y Estévez, 2011; Martínez-Monteagudo, Inglés, Cano-Vindel y García-Fernández, 2012).

### **6.2.2. Categorización y epidemiología de la ansiedad**

La diferencia entre la ansiedad patológica y la respuesta ansiosa adaptativa es una cuestión más cuantitativa que cualitativa, ya que fenomenológicamente comparten las mismas manifestaciones. Su consideración clínica se delimita a partir de niveles elevados de ocurrencia, intensidad y duración, así como de una significativa interferencia en la vida cotidiana de la persona. También es relevante como criterio clínico dar una respuesta ansiosa intensa e incapacitante a estímulos inofensivos o en ausencia de peligro real, es decir con un foco inadecuado (Eysenck, 1997; Piqueras, Ramos, Martínez y Oblitas, 2010).

La ansiedad puede presentarse como un síntoma en diferentes trastornos mentales y no cobra entidad nosológica por sí sola a menos que represente el componente principal de la disrupción psicológica. En este último sentido, se han estudiado patrones diferenciales en la ansiedad patológica. La conceptualización y categorización de estos distintos síndromes clínicos se ha ido modificando a lo largo del tiempo hacia una cada vez mayor especificación. Las categorizaciones que mayor aceptación han poseído en el ámbito clínico son las reportadas por los sistemas de clasificación: *Diagnostic and Statistical Manual of Mental Disorders* (DSM) de la American Psychiatric Association (APA<sub>2</sub> en las siguientes menciones) y el capítulo 5 (Trastornos mentales y del comportamiento) de la *Clasificación Internacional de Enfermedades* (CIE) de la Organización Mundial para la Salud (Pallares, 2002; Lemos-Giráldez, 2010).

Debido a la gran polémica académica (Domingo, 2013) que ha despertado el DSM-V (APA<sub>2</sub>, 2013) se ha decidido hacer referencia en esta exposición también a su anterior versión DSM IV-TR (APA<sub>2</sub>, 2002) que se ha mantenido vigente durante más de una década, matizando los cambios entre ambas ediciones. Se ha tomado esta determinación por el rechazo total o parcial de la nueva versión por parte de importantes instituciones de la salud mental como el National Institute of Mental Health de Estados Unidos (NIMH e Insel, 2013) o la American Psychological Association (Clay, 2012).

Tanto el DSM V como el DSM IV-TR, presentan una sección concreta de “Trastornos de Ansiedad” para agrupar los síndromes cuyo elemento definitorio sea la ansiedad. Por otro lado, CIE-10 describe los trastornos de ansiedad dentro la sección denominada “Trastornos neuróticos, por estrés y somatomorfos”. Se constata concordancia a la hora de señalar las diferentes categorías diagnosticas, existiendo equivalencia diagnostica en la mayoría los síndromes señalados. De hecho, en las clasificaciones DSM se recogen los códigos numéricos correspondientes de cada categoría en la CIE-10. Sí existen ciertas discrepancias en algunos criterios diagnósticos concretos, sobre todo a nivel cuantitativo. Pero la diferencia más importante entre las clasificaciones es la forma de agrupar algunos síndromes. Mientras que en el DSM IV-TR el trastorno obsesivo compulsivo y el trastorno postraumático se presentan como una categoría más en los trastornos de ansiedad, la CIE-10 y el DSM-V TR les dan una entidad separada con respecto a los trastornos de ansiedad presentándolos en su correspondiente sección propia. Adicionalmente en la CIE-10 se dividen los trastornos de ansiedad en dos subcategorías: ansiedad fóbica y otros trastornos de ansiedad. La categorías del DSM trastorno de ansiedad debido a enfermedad medica, inducido por sustancia o no especificado no tienen su equivalente en la CIE-10, y la categoría de la CIE-10 trastorno mixto ansiedad-depresión no aparece la clasificación DSM. Por último, en el DSM-V no existe la sección “Trastornos de inicio en la infancia, la niñez y la adolescencia”, que si se considera tanto el DSM-V TR como en la CIE-10, por lo que incluye en el capítulo dedicado a los trastornos de ansiedad dos trastornos que las otras dos clasificaciones ubican en dicha sección: ansiedad por separación y mutismo selectivo (APA<sub>2</sub>, 2002, 2013; OMS, 2003; Lemos-Giráldez, 2010).

En la Tabla 14 se recoge una definición breve de cada trastorno y su código de clasificación en la CIE-10 y en las DSM-IV TR/DSM-V. Es necesario matizar que el código resaltado por los sistemas DSM en sus últimas versiones no es un código propio sino que se corresponde con la numeración de trastornos de otra versión de la CIE promovido por el gobierno estadounidense para su uso en su sistema sanitario “ICD-9 Clinical Modification” (National Center for Health Statistics NCHS y Centers for Medicare and Medicaid Services CMS, 2011). Por ello, existen síndromes descritos en ambas versiones del DSM que no tienen ningún código asignado ya que no existen en la ICD-9 MC (APA<sub>2</sub>, 2013).

**Tabla 14***Clasificación categorial de los trastornos de ansiedad*

<b>Síndromes ansiosos</b>	<b>Definición resumida</b>	<b>DSM(ICD-9MC)/CIE-10</b>
<i>Trastorno angustia o de pánico.</i>	Aparición súbita e inesperada de síntomas ansiosos intensos.	300.01 y 300.21/F41.0
<i>Agorafobia.</i>	Aparición de ansiedad en situaciones donde se perciba que es difícil escapar o encontrar apoyo. Comportamientos de evitación.	300.22/ F40.0
<i>Fobia específica.</i>	Ansiedad clínicamente significativa como respuesta a exposición de estímulos específicos temidos. Comportamientos evitación	300.22/ F40.2
<i>Fobia social.</i>	Respuesta ansiosa similar a la anterior pero el estímulo fóbico son las situaciones sociales.	300.23/F40.1
<i>Trastorno obsesivo compulsivo.</i>	Presencia de obsesiones generadoras de ansiedad y/o comportamientos compulsivos neutralizadores de ansiedad.	300.3/F42
<i>Trastorno por estrés postraumático.</i>	Re experimentación de acontecimientos traumáticos pasados que generan gran activación de organismo y comportamientos de evitación de estímulos relacionados.	309.81/F43.1
<i>Trastorno o reacción por estrés agudo.</i>	Similar al anterior pero síntomas aparecen inmediatamente después del suceso traumático.	308.3/F43.1
<i>Trastorno de ansiedad generalizada.</i>	Estado de ansiedad y preocupaciones excesivas que persiste un periodo prolongado de tiempo.	300.02/F41.1
<i>Trastorno de ansiedad debido enfermedad medica (DSM).</i>	Presencia de síntomas de ansiedad secundaria a efectos fisiológicos directos de enfermedad.	293.84/----
<i>Trastorno de Ansiedad de ansiedad debido a sustancias (DSM).</i>	Presencia de síntomas de ansiedad secundaria a efectos fisiológicos directos de consumo de droga, fármaco o toxico.	----/----
<i>Trastorno de ansiedad no especificado (DSM).</i>	Respuesta ansiosa clínicamente significativa que no se ajuste a los demás síndromes descritos.	300.00/----
<i>Trastorno mixto ansiedad-depresión (CIE).</i>	Síntomas de ansiedad y depresión donde no existe predominancia sintomática suficiente para diagnostico específico de un tipo u otro.	----/F41.2
<i>Ansiedad de separación (DSM-V).</i>	Ansiedad excesiva para su momento evolutivo con respecto a la separación de personas o lugares a los que se está vinculada.	309.21/F93.0
<i>Mutismo Selectivo (DSM-V)</i>	Incapacidad persistente para hablar en situaciones sociales específicas.	313.23/F94.0

Nota: Tabla generada a partir del contenido de APA<sub>2</sub>, (2002; 2013) y OMS (2003).

De forma genérica, se señalan a los trastornos de ansiedad como los más frecuentes entre los grupos de trastornos tipificados en los sistemas de clasificación psiquiátricos. Poseyendo además una alta comorbilidad con otras enfermedades (medicas o psiquiátricas). Estos trastornos provocan un elevado impacto social y económico, que convierten su abordaje en una prioridad para los sistemas de salud de cualquier país desarrollado (Oliva, López, Montejo, Osuna y Duque 2009; Segarra, Farriols y Palma, 2011). Estudios epidemiológicos estadounidenses sitúan a los trastornos de ansiedad a la cabeza de los trastornos mentales con una prevalencia anual del 18,1%, estimando además que alrededor de un 28% de la población general ha tenido algún trastorno de ansiedad a lo largo de su vida (Kessler, Chiu, Dernier y Walters, 2005; Kessler, Berglund, Demler, Jin y Walters, 2005). Otros estudios con población internacional identifican tasas de prevalencia entre 5,6%-18,1%. Y dentro de estos trastornos señalan una mayor prevalencia de la fobia específica, seguido por la ansiedad generalizada y la fobia social. Se observa además una presencia mucho mayor de los trastornos de ansiedad en mujeres que en hombres (Somers, Godner, Waraich Hsu, 2006; Baumeister y Härter, 2007).

Según el European study of the epidemiology of mental disorders (ESEMeD), en España, los trastornos de ansiedad presentan tasas de prevalencia anual del 6,2%. La prevalencia de la mayoría de los trastornos de ansiedad se muestra significativamente superior en mujeres que en hombres, exceptuando la fobia social donde no existe tal diferencia. Se han encontrado más prevalentes los trastornos de ansiedad en la población con los siguientes estatus de ocupación laboral: desempleados, baja por enfermedad o baja maternal. El trastorno de ansiedad con prevalencia más alta es el de fobia específica, figurando además como uno de los trastornos mentales específicos más frecuentes de todas las categorías (Haro et al. 2006).

Por otro lado, en estudios con población clínica (los que acuden a servicios de salud mental), entre los trastornos de ansiedad más prevalentes figuran el trastorno de angustia o pánico, el trastorno de ansiedad generalizado y la fobia social (Wagner et al., 2005). Un reciente estudio clínico nacional identifica al trastorno de ansiedad no especificado y a la fobia social como los diagnósticos más frecuentes dentro de los trastornos de ansiedad. También se confirma la presencia significativamente mayor de mujeres en todos los trastornos de ansiedad, excepto en el trastorno obsesivo-compulsivo donde la proporción se

iguala a la de los hombres (Bernaldo-de-Quirós et al 2012).

### **6.2.3. Modelos explicativos de los trastornos de ansiedad**

Desde un punto de vista estrictamente biológico, la etiopatogenia de los trastornos de ansiedad se trata de explicar a través de alteraciones neuroquímicas o disfunciones estructurales del sistema nervioso. Con respecto a los sistemas de neurotransmisión, se ha observado que la estimulación de los receptores beta adrenérgicos provoca un aumento significativo de la ansiedad, aunque no hay acuerdo la hora de los mecanismos exactos de cómo esto sucede. Se señala también al sistema serotoninérgico, bajo la teoría de la ansiedad patológica se deba a un déficit en la producción de serotonina (5-HT) y una hiposensibilidad en los receptores postsinápticos a la misma. Se adjudica a la serotonina un efecto modulador en estructuras neuronales implicadas en la respuesta ansiosa, ya que ha constatado que los fármacos inhibidores selectivos de la recaptación de 5-HT (ISRS) han mostrado eficacia en la reducción sintomática de varios trastornos de ansiedad. Bajo esta misma lógica se señala al complejo receptor ácido gamma-aminobutírico (GABA)-benzodiazepina, teorizándose una sensibilidad alterada del receptor o una síntesis anómala de un ligando endógeno, lo que justificaría la eficacia de las benzodiazepinas de alta potencia para suprimir los síntomas ansiosos. Por otro lado, a nivel anatómico se ha observado que las siguientes estructuras neuronales tienen un especial protagonismo en la respuesta ansiosa: córtex prefrontal, locus caeruleus, amígdala y áreas del sistema límbico en general (Barlow 2002, Hollander y Simeon 2004; Martínez, Sáiz y García-Portilla, 2008).

Dentro del ámbito psicoanalista, el análisis de la ansiedad patológica ha poseído un especial protagonismo para el corpus teórico general del modelo. La ansiedad se consideró la base principal de la neurosis, etiqueta diagnóstica bajo la que se incluyeron los trastornos de ansiedad en los primeros sistemas de clasificación psiquiátrica, y que hacía referencia no solo a la vivencia misma de malestar y angustia, sino también a los mecanismos de defensa psíquicos para evitarla (Pallares, 2002; Sandín y Chorot, 2008). Inicialmente Freud (1895a; 1895b) teorizó que la ansiedad procedía de la energía sexual no descargada adecuadamente (en el caso de la angustia sin mediación psicológica) o de traumas psíquicos de la niñez (en el caso de lo que llama psiconeurosis como las fobias o las obsesiones). En una segunda teorización Freud (1926) identificó tres fuentes diferentes de ansiedad en coherencia con su modelo estructural de personalidad: realista (el miedo en presencia de amenazas reales),

neurótica (es el miedo a que los impulsos animales del Ello, no puedan ser controlados por el yo) y moral (la vergüenza o culpa generada por presión del Superyo al Yo por los impulsos o conductas no aceptables por el primero). A la ansiedad generada por las dos últimas fuentes las denomina ansiedad libre flotante por no depender de ningún estímulo externo para generarse. Con esta concepción se da un especial protagonismo al conflicto intrapsíquico como generador de la ansiedad, y se otorga más importancia a las experiencias de la niñez tempranas, como explicación de esos conflictos que a la naturaleza objetiva de los estímulos circundantes actuales, ya que estos últimos solo los reactivan por asociación simbólica (Coderech, 1982). A pesar de que esta noción intrapsíquica es la predominante en el ámbito psicodinámico, algunos autores del modelo la han desafiado, señalado la importancia del ambiente en la ansiedad, concretamente del enfrentamiento del individuo frente al mundo que le rodea (Horney, 1964; Sullivan, 1968).

En el marco de las teorías del modelo conductista en la explicación de la ansiedad patológica, cobran protagonismo los procesos de aprendizaje, entendiendo los síntomas ansiosos como conductas inadaptadas aprendidas mediante condicionamiento. La propuesta inicial fue la disquisición desde el condicionamiento clásico del desarrollo de las fobias, según la cual la ansiedad patológica sería una respuesta condicionada ante estímulos que antes eran neutros pero que en algún momento del pasado se habían asociado con estímulos aversivos. Pese a que esta explicación sirvió para dar sentido a ciertos fenómenos específicos, se mostró insuficiente para explicar la mayoría de las situaciones clínicas reales, ni la no extinción de la respuesta condicionada en ausencia de reforzamiento (Eysenck, 1981; Pailhez y Bulbena, 2009). Para superar estas limitaciones se plantearon otras aproximaciones conductistas, pero desde los principios del condicionamiento operante. Mowrer (1940) con su modelo bifactorial añade al planteamiento clásico, el rol de las conductas de evitación a estímulos temidos como reforzadores de ese condicionamiento. De Silva, Rachman y Seligman (1977) señalan que pueda existir una predisposición evolutiva a adquirir determinados miedos, estando las personas filogénicamente predisuestas, lo que explicaría la adquisición rápida y difícil de distinguir de ciertas fobias muy frecuentes. Eysenck (1981) continúa desarrollando el modelo de Mowrer pero añadiendo matices como la incubación del miedo (estímulos condicionados fuertes pueden tomar las características del estímulo incondicionado, siendo resistentes a la extinción) o la existencia de determinados sesgos cognitivos interpretativos y de memoria en las personas con elevada ansiedad-rasgo. Bandura

(1977) añadiría que el desarrollo de reacciones de ansiedad puede realizarse también observando a otra persona mediante aprendizaje vicario, generándose expectativas de asociación entre dos estímulos sin necesidad de experiencia directa previa. Como se puede observar, se fueron incorporando cada vez más componentes cognitivos y variables intermedias a las explicaciones sobre la ansiedad (Estevez, 2008).

En el modelo cognitivo se subraya precisamente el peso específico de los procesos cognitivos en el desarrollo y mantenimiento de los trastornos de ansiedad. Entendiéndose cognición como toda tarea mental destinada a procesar información (atención, percepción, memoria, comparación...) (Piqueras, Ramos, Martínez y Oblitas 2010). En este sentido, Beck y Clark (1988; 1997) restan importancia a los estímulos externos objetivos y fijan la determinación de la vivencia ansiosa en los denominados “esquemas cognitivos” (construcciones subjetivas de la experiencia personal bastante estables que incluyen profundas creencias o supuestos sobre la realidad que rodea a la persona). Los autores señalan a la activación de esquemas cognitivos de daño como los responsables de que se sobreestime el peligro y devalúe la capacidad para afrontarlo. Esta activación provoca además una serie de pensamientos automáticos catastrofistas y de errores de procesamiento de la información (sesgos cognitivos), que mantienen las conductas de ansiedad y provocan conductas de evitación reforzantes. Además de evidenciarse la vinculación de estos procesos a la respuesta ansiosa, se ha constatado empíricamente que existe especificidad en el contenido de las cogniciones y tipos de sesgos para los diferentes trastornos de ansiedad tipificados (Breitholtz Johansson y Ost, 1999; Lerner et al., 1999; Mogg, Mathew y Weinman 1999; Hollander y Simeon, 2004). Por ello, los actuales modelos cognitivos para la ansiedad se centran tanto en el contenido de las cogniciones como el proceso por las cuales éstas influyen en la conducta (Ramos-Cejudo y Cano-Vindel, 2008).

### **6.3. Depresión**

A nivel psiquiátrico se entiende la depresión como un trastorno afectivo en el cual se sufre un estado de gran abatimiento y donde predominan los síntomas emocionales de: tristeza intensa, decaimiento, irritabilidad, impotencia y sensación subjetiva de malestar. Adicionalmente también se presentarán síntomas coherentes con este estado de tipo cognitivo, volitivo y somático (Lizondo y Llanque, 2008). El estado emocional de abatimiento y tristeza del ánimo depresivo puede tener funciones adaptativas para el organismo (como modo de comunicación para conseguir atención y cuidado de los que le rodean, y estado de ahorro y conservación de energía para futuros procesos de acomodación y supervivencia), sin embargo si su intensidad, duración y frecuencia, interfieren gravemente en la propia capacidad adaptativa se considerará necesariamente un estado patológico (Vázquez y Sanz, 2008). Las personas patológicamente deprimidas caen en un estado de apatía, desesperanza y desmotivación, en el cual generan creencias de que nada de lo que hagan podrá mejorar su situación de sufrimiento, pudiendo llegar a generar ideaciones suicidas e incluso llevarlas a cabo (Seligman, 1981).

#### **6.3.1. Aclaraciones conceptuales sobre la depresión**

El término depresión es utilizado en la actualidad con tres significados. En primer lugar, se puede entender como un síntoma que se manifiesta sin conllevar una entidad patológica por sí sola, aunque puede acompañar a otros trastornos. Este sentido se refiere al ánimo depresivo (sentirse triste) muy frecuente, en la mayoría de los cuadros psicopatológicos o de forma circunstancial en la población general. En segundo lugar, se puede referir a un síndrome, es decir, un conjunto de síntomas y signos relacionados que se dan de forma covariante. En el caso de la depresión se agrupan síntomas de muy diversa naturaleza (insomnio, desmotivación, tristeza, pérdida de peso ...) sobre los que se profundizará más adelante en este mismo apartado. Y el tercer significado es considerar la depresión como una enfermedad con entidad nosológica equivalente a un trastorno biológico, siendo posible delimitar una etiología determinada, un curso, un pronóstico y un tratamiento específico (Castilla-del-Pino, 2002; Vallejo, 2011).

Entendida como síndrome, la depresión implica una gran variedad de síntomas y signos que pueden diferir en intensidad y frecuencia entre personas o a lo largo del tiempo. Para

Vázquez y Sanz (2008) todos ellos se pueden concretar en cinco núcleos de manifestación:

- *Síntomas anímicos*: sentimientos de tristeza, infelicidad, anhedonia, irritabilidad, sensación de vacío nerviosismo, gran disminución de emociones positivas, e incluso en casos graves la sensación de no poder sentir nada.
- *Síntomas motivacionales y conductuales*: estado general de inhibición comportamental, apatía, desmotivación generalizada, conductas frecuentes de llanto (o en situaciones extremas no poderlo ejecutar pese a quererlo), y enlentecimiento motor y del habla (pudiendo llegar a situaciones de mutismo y parálisis casi totales).
- *Síntomas cognitivos*: disminución generalizada del rendimiento cognitivo de la persona (atención, concentración, memoria...), pensamientos suicidas, fijación cognitiva en las vivencias dolorosas, transformación negativa de la autoimagen, abundantes sesgos cognitivos, pensamientos rumiativos e ideación de culpa, ruina, hipocondría y suicidio.
- *Síntomas y signos físicos*: problemas de sueño (siendo el más frecuentemente el insomnio), fatiga, pérdida de apetito, disminución del deseo sexual (problemas de erección y lubricación) y molestias corporales difusas.
- *Síntomas interpersonales*: tendencia al aislamiento, disminución de interés por la gente, rechazo e incompreensión por parte de los que le rodean y deterioro generalizado de todas las relaciones personales.

A la hora de definir la depresión como trastorno, ha existido una tendencia histórica a entenderla de forma dicotómica, representando las subdivisiones endógena/ reactiva y psicótica/neurótica como las más influyentes y controvertidas. El supuesto base de la primera es etiológico. Se adjudica a la depresión tipo endógeno un origen biológico posiblemente hereditario sin necesidad de desencadenantes ambientales, y a la depresión reactiva se le atribuyeron factores precipitantes de tipo psicosocial. Sin embargo, la diferenciación práctica que se ha realizado con esta dicotomía ha sido en realidad fenomenológica identificando un patrón sintomatológico diferencial entre una y otra: predominio de síntomas de tipo vegetativo y mayor gravedad general en la endógena, y de menor gravedad y con un componente más ansioso y de irritabilidad en la reactiva. La otra distinción, a pesar de haber sido bastante utilizada en el ámbito psiquiátrico, ha resultado ser más ambigua en su definición que la anterior debido a los distintos significados que se le

ha conferido al término neurótico. De forma genérica, se ha asumido que la condición neurótica implica una intensidad sintomática moderada o ligera y que la depresión psicótica es de una intensidad mucho más grave, siendo por tanto cuestión más dimensional que categorial. Otra interpretación de esta taxonomía, ha sido simplemente constatar la presencia o no de síntomas psicóticos tal y como se entienden en la actualidad (delirios y alucinaciones). Ambas dicotomías no han recibido mucho apoyo empírico, y exceptuando la condición endógena (que sí ha mostrado cierta entidad empírica) no se consideran criterios diagnósticos fiables hoy en día (Vallejo y Gastó, 2000; Chinchilla 2008; Lemos-Giráldez, 2010).

La palabra melancolía también se encuentra estrechamente relacionada con la depresión clínica. Aunque su significado se ha ido modificando con respecto a la acepción que originariamente se le dio. Hipócrates designó melancolía a un estado clínico donde predominaba la inhibición y la tristeza. Lo achacó a un exceso de bilis negra, “melaina chole” en griego (humor responsable de la sensación de abatimiento y la somnolencia). Se le asociaron una serie de signos que combinaban las pasiones de tristeza y miedo (desesperación, intranquilidad, aversión a la comida...). Pese a que esta etiqueta diagnóstica incluiría sin duda lo que hoy en día se consideraría un trastorno depresivo, también abarcaría muchos otros estados psicopatológicos. Esta conceptualización predominó varios siglos con algunas variaciones de criterios (como la inclusión o no del riesgo suicida) y diversas subcategorizaciones (como la melancolía erótica, delirante o hipocondriaca). Sin embargo, con el nacimiento de la psicopatología moderna en el siglo XIX, el término dejó de ser utilizado como categoría general (sustituido por la conceptualización Krapeliana de la enfermedad maníaco-depresiva). Su consideración contemporánea ha dejado de ser sinónimo de trastorno depresivo para señalar de forma específica a las manifestaciones más intensas, graves y con menor reactividad al entorno (Cobo, 2005; Gérard 2012).

Otro concepto asociado que se ha asociado a la depresión, ha sido la clásicamente denominada locura maniaco-depresiva, o en una consideración más reciente trastorno bipolar. Durante el siglo XIX, la psiquiatría francesa y alemana (con autores como Baillarger, Khalbaum o Kraepelin) vincularon la depresión patológica con su polo opuesto la manía (excesiva euforia y excitación), debido a que en su observación clínica, constataron la existencia de cursos circulares o ciclotímicos entre ambos polos. Se generó la categoría maniaco-depresivo para diagnosticar los casos de “locura” caracterizados por excesos en

la afectividad, lo que incluía de manera indiferenciada todos los cuadros clínicos de lo que hoy se conoce como trastornos del ánimo. Posteriormente, ya en el siglo XX, se realizó una diferenciación básica entre aquellos cuadros que efectivamente mostraban esta oscilación entre la manía y la depresión, y los que solo mostraban uno de los dos polos. De esta forma la depresión (mono polar) obtuvo entidad diagnóstica propia y una definición más concreta que en épocas anteriores, siendo la no presencia de episodios maníacos o hipomaníacos (episodios maníacos de menor intensidad) el criterio fundamental para diferenciarla de los trastornos afectivos bipolares (Leonhard, 1999; González-Pinto, López-Peña y Zorrilla-Martínez, 2009).

### **6.3.2. Categorización y epidemiología de la depresión**

A la hora de ubicar los criterios para el diagnóstico de la depresión clínica, ha existido una gran controversia teórica sobre su concepción respecto a la tristeza o ánimo deprimido no patológicos (Aguelo, Spielberg y Buéla-Casal, 2007; Vázquez y Sanz, 2008). Por un lado, se ha expuesto que los patrones de manifestación de depresión patológica son cualitativamente diferentes al estado depresivo no clínico, por lo que cobran sentido las taxonomías de diagnóstico categoriales (Feighner et al. 1972; Andreasen, 1985). En oposición a esta postura se ubica la perspectiva dimensional, que defiende que el estado depresivo es un continuo donde solo varía la intensidad y gravedad de los síntomas, es decir, que su consideración clínica es una cuestión cuantitativa y no cualitativa (Golberg y Huxley 1992; Lewinsohn, Solomon, Seely y Zeiss, 2000).

En los últimos años se ha ido acumulando la evidencia empírica a favor de la interpretación dimensional, con diversos estudios que constatan la existencia del continuo normativo-clínico y problematizan la noción categorial (Ruscio y Ruscio, 2000; Slade y Watson, 2005; Hamkin, Fraley, Lahey y Waldman, 2005; Sakashita, Slade y Andrews, 2007). Además se ha observado mayor poder predictivo, con respecto al curso de la depresión, utilizando el enfoque dimensional que el categorial (Bjelland et al., 2009). Por otro lado, aunque este continuo es fácilmente constatable respecto a los síntomas de tipo afectivo (malestar emocional y psicológico), en algunas investigaciones se han encontrado dificultades para lo mismo con respecto a los síntomas somáticos (Ambrosini, Bennet, Cleland y Haslam, 2002; Beach y Amir, 2003).

Aunque el enfoque categorial puro se ha mostrado inadecuado o insuficiente para dar respuesta a la clínica real, la alternativa dimensional muestra también algunas limitaciones en la actualidad. Existe el problema de cómo articular patrones complejos de síntomas manifiestos para estudiar los mecanismos subyacentes. También se hace presente la dificultad de ubicar los límites entre dimensiones o entre normalidad y patología. Por otro lado, se pierde la facilidad de comunicación entre profesionales que poseía el uso de categorías. Por todo ello, hay autores dentro de la psiquiatría que plantean el uso combinado y complementario de ambas perspectivas por parte del profesional (Andrews et al., 2007; Widakowich, 2012; Stein, 2012). De hecho, algunos estudios taxométricos presentan argumentos a favor de la entidad de la categoría diagnóstica de trastorno depresivo mayor cuando se matiza con medidas de tipo dimensional (Ruscio, Brown y Ruscio, 2009). Esta sensibilidad mixta ha tratado de reflejarse en la última versión del DSM (Regier, Narrow, Kuhl y Kupfer, 2009; APA<sub>2</sub>, 2013), que sin abandonar la clasificación categorial ha incluido en su tercera sección medidas dimensionales de severidad y frecuencia de trece dominios sintomatológicos (entre los que se encuentran tanto los de tipo ansioso como depresivo) para utilizar en combinación con todas las categorías diagnósticas. Pese a ello, se ha criticado que esta implementación no es realmente una integración de las dos perspectivas, sino un añadido cosmético a un sistema cuya articulación continua siendo realmente categorial (Echeburúa, Salaberría y Cruz-Sáez, 2014).

Los planteamientos categoriales que mayor impacto han tenido en la práctica clínica en las últimas décadas han sido los reportados por los sistemas de clasificación: DSM de la Asociación Americana de Psiquiatría y CIE de la Organización Mundial para la Salud, cuyos criterios principales son mayormente congruentes y compatibles entre sí (Lemos-Giráldez, 2010). En esta exposición se ha decidido tomar la precaución de tener en cuenta no solo la última versión del DSM sino también la anterior por las razones expuestas en el apartado dedicado a la ansiedad.

Respecto a su tipificación tanto el DSM IV-TR como el CIE-10 incluyen los trastornos depresivos dentro de un capítulo más general dedicado a las alteraciones patológicas del estado del ánimo. De esta forma comparten sección con los trastornos bipolares. El DSM-V dedica una sección independiente a los trastornos depresivos, desvinculándolo de los trastornos bipolares. En el DSM IV-TR se señalan categorías tres diagnósticos diferenciales dentro de los trastornos depresivos: depresivo mayor, distímico y depresivo no

especificado, a los que se añaden los trastornos del ánimo inducido por sustancias o por enfermedad. En el CIE-10 se observan categorías equivalentes aunque el trastorno distímico se ubica bajo el epígrafe “otros trastornos del humor persistentes” junto a la ciclotimia. En el DSM V, también se recogen las categorías de trastornos de depresión del DSM IV-TR aunque se añaden dos nuevos cuadros que no tenían consideración de trastornos depresivos hasta esta versión: el trastorno disfórico premenstrual y el trastorno del estado de ánimo disruptivo y no regulado. (APA<sub>2</sub>, 2002; 2013; OMS 2003; Echeburúa, Salaberría y Cruz-Sáez, 2014). En la Tabla 15 se recoge una definición breve de cada trastorno y su código de clasificación en la CIE-10 y en las DSM-IV TR/DSM-V.

Tanto el CIE-10 como el DSM-IV TR resaltan el concepto de “episodio depresivo mayor”, no como categoría concreta sino como entidad central para el diagnóstico y especificación de los trastornos depresivos. En el DSM-V presenta los mismos criterios de episodio depresivo del DSM-IV TR, pero directamente subscritos al diagnóstico de trastorno depresivo mayor (APA<sub>2</sub> 2002, 2013; OMS 2003).

El primer criterio, para identificar un episodio depresivo mayor, según el DSM IV TR, es presentar un cuadro clínico con al menos cinco de los siguientes síntomas (incluyendo siempre necesariamente a los dos primeros), durante por lo menos dos semanas, de manera prácticamente diaria (APA<sub>2</sub>, 2002):

- Estado de ánimo depresivo (tristeza o vacío o en menores irritabilidad).
- Disminución del interés o de capacidad para el placer.
- Pérdida o aumento de peso significativo; o pérdida de apetito.
- Insomnio o Hipersomnia.
- Agitación o enlentecimiento psicomotor.
- Fatiga o pérdida de energía.
- Sentimientos de inutilidad o culpa excesivos o inapropiados.
- Disminución de la capacidad para pensar, concentrarse o tomar decisiones.
- Pensamientos recurrentes de muerte; ideación frecuente o tentativa de suicidio.

Además de la presencia de parte de estos síntomas, se debe constatar que los mismos interfieren de forma significativa con el funcionamiento cotidiano de la persona. Por otro lado, los síntomas no deben achacarse a efectos fisiológicos provocados por alguna sustancia o condición médica. Tampoco debe encajar con los criterios para identificar un episodio mixto maniaco depresivo. Por último, no se debe considerar un episodio depresivo si los síntomas se pueden explicar por un duelo reciente (menos de dos meses) a menos que el deterioro del funcionamiento sea muy grave. En el DSM-V esta última consideración se ha omitido totalmente, no representando el duelo ya una excepción a la identificación del episodio. En ambas versiones es necesario especificar la gravedad del episodio según el número de síntomas, su interferencia en la vida del paciente y la presencia o no de síntomas psicóticos (APA<sub>2</sub>, 2002; 2013).

El CIE-10 recoge criterios similares a los presentados en el DSM-IV para la identificación del episodio, sin embargo añaden un síntoma más a la lista (pérdida de confianza y autoestima) y elevan la importancia del síntoma “perdida de energía o fatiga” al poder sustituir a cualquiera de los dos primeros como síntoma nuclear básico. Además el número de síntomas exigidos para su constatación se fija en función de la gravedad que se pretenda diagnosticar. Por último, realiza una subtipificación adicional de los episodios según la presencia o ausencia de síntomas somáticos (OMS, 2003).

De esta forma, el diagnóstico de trastorno depresivo mayor vendrá definido (tanto en el DSM o como el CIE) por la presencia de uno o varios de estos “episodios depresivos mayores” en ausencia de episodios previos de manía o hipomanía y descartando presencia de trastornos psicóticos. El trastorno distímico se caracteriza por presencia de un estado depresivo continuo de larga duración (por lo menos dos años), que no llega a cumplir criterios de episodio depresivo mayor (aunque en el DSM-V no es ya excluyente dando más importancia a la cronicidad de los síntomas). Y el trastorno depresivo no especificado se reservaría para las situaciones donde se manifiesten síntomas depresivos disruptivos que no lleguen a cumplir las condiciones de episodio depresivo mayor, ni se pueda diagnosticar un trastorno distímico o un trastorno adaptativo con ánimo depresivo (APA<sub>2</sub>, 2002; 2013; OMS, 2003).

**Tabla 15***Clasificación categorial de los trastornos depresivos*

<b>Trastornos depresivos</b>	<b>Definición y subtipos</b>	<b>DSM(ICD-9MC)/ CIE-10</b>
Trastorno depresivo mayor episodio único.	Un episodio depresivo mayor: Leve, Moderado, Grave (con o sin síntomas psicóticos), En remisión (total o parcial), No especificado. Especificaciones adicionales: catatónicos, melancólicos, atípicos y postparto.	296.2X(0-9) /F32.X(0-8) Ultimo digito X determinado según gravedad.
Trastorno depresivo mayor, recidivante o recurrente.	Varios episodios depresivos mayores. Según último episodio: Leve, Moderado, Grave (con o sin síntomas psicóticos), En remisión (total o parcial) o No especificado. Especificaciones adicionales: catatónicos, melancólicos, atípicos, postparto, con o sin recuperación inter-episódica o con patrón estacional.	296.3X(0-9) /F33.X(0-7) Ultimo digito X determinado según gravedad.
Trastorno distímico o trastorno depresivo persistente	Dos años más con estado de ánimo depresivo habitual, junto con otros síntomas depresivos que no llegan a cumplir criterios de episodio depresivo mayor. Especificaciones: inicio antes o después de los 21 años y si existen síntomas atípicos.	300.4/ F34.1
Trastorno depresivo no especificado	Cuadro de síntomas depresivos que no cumplen criterios de ningún otro diagnóstico.	311/ y F32.9
Trastorno del estado del ánimo (depresivo) inducidos por sustancias.	Presencia de síntomas depresivos secundarios a efectos fisiológicos directos de consumo de droga, fármaco o toxico.	29X.XX/F1X.8 X especificación de sustancia
Trastorno del estado del ánimo (depresivo) a condición médica.	Presencia de síntomas depresivos secundarios a efectos fisiológicos de condición médica.	293.83/F06.3x
Trastorno del estado del ánimo disruptivo y no regulado. Solo DSM-V.	Presencia de irritabilidad severa crónica que lleva a explosiones significantes y dificultades para controlar sus emociones. Menores de 18 años.	296.99/ F34.8
Trastorno disfórico premenstrual. Solo DSM-V.	Presencia recurrente de síntomas anímicos (destacándose los de tipo depresivo) en la última semana de la fase lútea del ciclo menstrual.	625.4/N94.3

Nota: Tabla generada a partir del contenido de APA<sub>2</sub>, (2002; 2013) y OMS (2003)

La Organización Mundial de la Salud (OMS, 2001) identifica la depresión como la cuarta problemática de salud causante de discapacidad en el mundo, estimándose que en el año 2020 se convertirá en la segunda enfermedad que mayor carga y coste social genere. Adicionalmente señala que aunque la prevalencia de los trastornos depresivos puede variar de una población a otra, siempre existe una afectación mayor de mujeres que de hombres. También han constatado que un diagnóstico clínico de depresión multiplica por 22 el riesgo de suicidio respecto a la población general, solo estando por debajo (en riesgo) de los trastornos de alimentación y las toxicomanías (OMS, 2005). En otro informe reciente esta organización (OMS, 2014) ha señalado que la depresión es el factor predominante de discapacidad a nivel mundial en menores, de 10 a 18 años, representando también la tercera causa de muerte en esa misma franja de edad.

Según estudios epidemiológicos en la población general se observan tasas de prevalencia de entre 13-20% de síntomas depresivos, y una tasa de 3-8% que cumplen criterios para diagnóstico clínico. Evidenciándose, además, una prevalencia vital del 5% del episodio depresivo mayor en la población adulta (Myers et al., 1983; Gerard, 2012). Aunque los trastornos de ansiedad se han señalado como los más prevalentes como grupo diagnóstico general, se ha observado tanto en Europa como en Estados Unidos que el episodio depresivo mayor es la condición diagnóstica específica más prevalente entre los trastornos mentales, solo igualada por el diagnóstico de abuso de alcohol (Campo-Arias y Cassiani, 2008). Según la revisión Vazquez y Sanz (2008) se ha constatado un mayor riesgo de padecer trastornos depresivos en el grupo de edad de adultos jóvenes (de 25 a 44 años) y en población activa desocupada laboralmente. Estos autores también señalan un aumento gradual de la prevalencia de los trastornos depresivos en las últimas décadas, especialmente en los varones, reduciendo poco a poco la gran diferencia entre ambos sexos en tasas de prevalencia. A pesar de ello, la preponderancia de las mujeres en los trastornos depresivos es uno de los eventos epidemiológicos más frecuentes y estables en todos los estudios, figurando frecuentemente con una tasa que duplica a la de los varones en el diagnóstico de depresión mayor (Estévez, 2008).

Según el ESEMeD, en España, el episodio depresivo mayor tiene una prevalencia-año de 3,96 % y prevalencia vida 10,55 %, representando el diagnóstico de salud mental más frecuente en el país. Por su parte, el trastorno distímico presenta una prevalencia-año de 1,49% y una prevalencia-vida de 3,65%. Se confirma también en el país la superioridad

de prevalencia femenina en el diagnóstico de trastornos depresivos, con tasas de prevalencia año de 6,25% en depresión mayor y 2,38 en distimia frente a tasas masculinas de 2,15% y 0,52 respectivamente. Con respecto a otros factores asociados se observa mayor prevalencia de trastornos depresivos en personas que viven en grandes ciudades respecto a los que lo hacen en zonas rurales. Y se observa que la prevalencia de estos trastornos tiene una relación inversa con los años de educación formal (Haro et al., 2006).

### **6.3.3. Modelos explicativos de la depresión**

Desde la biología, se ha estudiado la depresión como un fenómeno orgánico complejo y multifactorial. Aunque aún no ha sido posible aclarar totalmente su fisiopatología, se han investigado una serie de condicionantes biológicos relevantes que implican: anomalías neuroquímicas, factores genéticos, alteraciones anatómicas y cambios endocrinos (Silva, 2002). Las principales hipótesis neuroquímicas se han enfocado sobre los sistemas: monoaminérgico (como déficit de la noradrenalina NA en la sinapsis), serotoninérgico ( como déficit de serotonina 5-HTT en diferentes áreas cerebrales) y colinérgico-adrenérgico ( descompensación permanente de equilibrio hacia colinérgico y acetilcolina, ACh) (Casado, 2006). Respecto a la influencia genética: los estudios con gemelos otorgan a este factor entre 40 % y 60% de la variabilidad fenotípica en la depresión mayor, los estudios familiares y de adopción encuentran una mayor prevalencia de trastornos depresivos en las personas con parientes de primer grado que los hayan sufrido, y en los estudios de ligamiento y asociación de marcadores genéticos se ha resaltado las regiones cromosómicas 11p15, 17q11.1-72q y 17q11.2, y de forma específica al gen SLC6A4 como importantes variables para el desarrollo e intensidad de las patologías depresivas por su influencia en la síntesis y funcionamiento de ciertos neurotransmisores (Mitjans y Arias, 2012). A nivel neuroanatómico se hipotetiza la alteración de dos circuitos neuronales reguladores del estado del ánimo: límbico-talámico-cortical (amígdala, núcleos dorsomediales del tálamo y corteza prefrontal) y límbico-estriado-pálido-talámico-cortical. Además mediante técnicas de neuroimagen se ha observado la activación diferencial de la corteza medial prefrontal izquierda y el giro cingulado derecho en pacientes deprimidos frente a personas no deprimidas. También se han constatado anomalías estructurales (como un menor tamaño) en los ganglios basales y el lóbulo frontal (Guadarrama, Escobar y Zhang, 2006). Respecto al sistema endocrino, se ha relacionado la depresión con una anomalía en el funcionamiento del eje hipotálamo-hipófisis-adrenal que implica un trastorno en la secreción de cortisol.

Esto a su vez altera los receptores glucocorticoides en ciertas zonas cerebrales como el hipocampo, lo que contribuye a generar sintomatología depresiva (Silva, 2002).

En el contexto psicoanalítico, la conceptualización clásica de la depresión patológica (que denominan melancolía) la relaciona con deseos destructivos hostiles hacia personas u objetos queridos y perdidos (real o simbólicamente). Esta hostilidad es introyectada hacia uno mismo de forma inconsciente, por la imposibilidad de desligarse emocionalmente de lo perdido, que para recuperarlo el paciente lo introduce simbólicamente dentro de su yo, y gestiona esa energía pulsional investida en ello. Lo que produce síntomas de culpa, baja autoestima y sentimientos de pérdida. Esta es la noción nuclear de la depresión en el modelo (Ávila, 1990) planteada originalmente por Freud (1917) que lo vinculó a la oralidad y a mecanismos obsesivos. Autores posteriores han mantenido la idea de pérdida y auto agresividad pero han diferido a la hora de dar protagonismo a determinadas estructuras o dinámicas tales como: características relación madre-hijo (Klein, 1940), narcisismo patológico (Gero, 1936; Kernberg, 1976) o autoestima y auto representaciones (Sandler y Joffe 1965; Jacobson, 1971). En las últimas décadas se ha dado un acercamiento desde el psicoanálisis a modelos cognitivos con conceptualizaciones que se enfocan en la interacción de estructura de personalidad (generada de forma experiencial en la infancia y adolescencia) con la percepción de incapacidad y frustración ante determinados hechos o eventos vitales (Ávila 1990; Villalba 2010).

Desde la perspectiva conductista, la sintomatología depresiva se debe a una modificación en los patrones de reforzamiento, que provoca una reducción significativa de conductas y una alteración general en la motivación. Las principales aportaciones teóricas del modelo han matizado esta noción base (Lemos-Giráldez, 2010). Ferster (1973) realizó un análisis funcional de las conductas de los pacientes depresivos, y concluyó que se caracterizaban por dos dinámicas: disminución de la frecuencia de conductas destinadas a controlar el medio que son reforzadas positivamente, y exceso de conductas de evitación ante estímulos aversivos. Debido a la evitación y a la falta de reforzamiento, se entra en un círculo vicioso de gran inhibición comportamental. Siguiendo esta lógica, Lewinsohn (1974) identifica la depresión como una respuesta de la persona ante la falta de reforzadores positivos contingentes a la conducta en las principales áreas de su vida. Considera el resto de la sintomatología depresiva como una consecuencia secundaria. Esta falta de reforzadores se puede deber a que : el ambiente no se los otorga a un nivel satisfactorio, la persona no

cuenta con las habilidades sociales suficientes para obtenerlos o existe una elevada ansiedad social que entorpece su consecución y disfrute. Debido a esto, se reduce el nivel general de actividad, lo que a su vez somete a la persona a un programa de extinción de conductas anteriormente reforzantes. Se mantiene esta respuesta debido a la reacción de las personas que rodean al paciente, que terminan evitándole reduciendo aún más los refuerzos positivos. Por otro lado, Costello (1972) quita protagonismo al número de refuerzos positivos y señala que la depresión es consecuencia más bien de la pérdida de la eficacia de las actividades reforzadoras llevadas a cabo por la persona. Esto se puede deber a cambios biológicos endógenos de la persona o a la pérdida de estímulos discriminativos en las cadenas conductuales que llevan al reforzador. En desarrollos posteriores se ha reformulado este enfoque clásico, centrado en el refuerzo, para incluir la mediación de factores cognitivos (Reiss y Bootzin, 1985).

En el modelo cognitivo para explicar la sintomatología clínica, se da más relevancia al cómo las personas construyen psicológicamente las situaciones que los acontecimientos externos en sí mismos. Por ello, se destacan los procesos cognitivos como fundamentales en el desarrollo, mantenimiento y remisión de los estados depresivos (Estévez, 2007). Beck (1970) considera la depresión como resultado de una distorsión sistemática (sesgo) en el procesamiento de la información de la persona. Esta distorsión se caracteriza por lo que denomina triada cognitiva de la depresión: una visión negativa de uno mismo, del mundo y del futuro. El resto de los síntomas depresivos son consecuencia directa de esto. Como explicación de la génesis de este procesamiento desajustado, se señala a la activación de una serie de esquemas disfuncionales depresógenos (actitudes o creencias estables profundamente arraigadas en la persona, posiblemente adquiridas en la infancia, que son poco realistas e inflexibles) ante un estresor o acontecimiento específico.

Por su parte Seligman (1975) conceptualiza un modelo de depresión denominado “indefensión aprendida”, por el cual la sintomatología se generaría por la percepción de la persona de incontabilidad respecto a los eventos aversivos, es decir, da igual lo que uno haga ya que el resultado no va depender de la propia conducta. Posteriormente este enfoque se matiza (Abramson, Seligman y Teasdale, 1978) con las teorías atribucionales de la psicología social. A esa situación de incontabilidad percibida se le suma un determinado estilo atributivo en el paciente depresivo. Este estilo se caracteriza por interpretar las causas de los sucesos negativos como factores internos (a uno mismo, provocando baja

autoestima), estables (manteniendo la respuesta depresiva en el tiempo) y globales (generalizable a muchas áreas de su vida).

Por otro lado, Rehm (1977) propone un marco explicativo que integra elementos cognitivos, conductuales y de aprendizaje social. Para este autor los episodios depresivos se desencadenan por la interacción entre dos elementos: una situación de falta o ausencia de reforzadores ambientales (estímulo desencadenante) y un déficit en las habilidades de autocontrol (vulnerabilidad psicológica). Sobre estas últimas se plantea la siguiente subdivisión: autoobservación (atención selectiva a sucesos negativos), autoevaluación (criterios excesivamente altos y estilo de atribución descrita en indefensión aprendida) y autoreforzos (se administra pocos premios y muchos castigos a sí mismo).

En la revisión de Vázquez, Hervás, Hernangómez y Romero (2010) se señala que todos estos planteamientos teóricos han ido acumulando evidencia empírica a lo largo de los años, pero también muestra como las últimas líneas de investigación han buscado una mayor concreción en el paradigma cognitivo añadiendo elementos tales como: importancia de las experiencias tempranas, sesgos cognitivos específicos según patrón de síntomas, estilo rumiativo de pensamiento o el papel de la vivencia del estrés en la sintomatología depresiva.

#### **6.4. Asma**

Sobre esta síndrome respiratorio no existe un acuerdo para generar una definición unívoca y precisa. Esto se debe a que este diagnóstico integra diferentes fenotipos que a pesar de compartir manifestaciones clínicas similares, pueden tener distinta etología. (Plaza et al. y Comité Ejecutivo de la Guía Española para Manejo del Asma, GEMA, 2003). Por ello, su caracterización se ha realizado a través sus características clínicas y fisiopatológicas, estableciéndose un consenso a la hora de destacar los siguientes elementos comunes en sus diversas manifestaciones (Osakidetza Servicio Vasco de Salud, 2005; National Health, Lung and Blood Institute NHLBI, 2007; Sociedad Española de Neumología y Cirugía Torácica SEPAR et al., 2009; Scottish Intercollegiate Guidelines Network SIGN y British Thoracic Society BTS, 2012, Global Initiative for Asthma GINA, 2014):

- Se presenta un desorden inflamatorio crónico de las vías respiratorias.
- Existe una hiper-reactividad bronquial ante diversos estímulos (entendida como

respuesta broncoconstrictora excesiva ante estímulos físicos químicos o biológicos).

- Los síntomas se asocian a la obstrucción variable del flujo aéreo parcialmente reversible (de forma espontánea o con tratamiento). Su exacerbación se desencadena por factores como: elementos alérgenos, cambios climatológicos, contaminantes atmosféricos, infecciones víricas, tabaco, humedad etc...
- Existe evidencia de la limitación de flujo aéreo a través de instrumentos objetivos que midan función pulmonar antes de iniciar tratamiento.
- Se dan episodios de sibilancias, disnea, opresión torácica y tos. Estos varían en presencia, frecuencia e intensidad a lo largo del tiempo.
- Los síntomas empeoran a menudo de noche o en primeras horas del día.
- En la condición inflamatoria se destaca el papel de ciertas células inflamatorias: mastocitos, eosinófilos, linfocitos  $T_h 2$ , neutrófilos y macrófagos.
- Pueden darse cambios estructurales en las vías respiratorias tales como: epitelio bronquial dañado, fibrosis subepitelial, activación de nervios colinérgicos, hiperplasia de las glándulas mucosas, remodelación de vías aéreas secundarias, y estimulación de fibroblastos y miofibroblastos.
- En el asma infantil, especialmente en los menores de 7 años, es la presencia de sibilancias recurrentes el síntoma central, al no poder constatar otros criterios del síndrome general como la reducción de la función pulmonar.

#### **6.4.1. Epidemiología, clínica, diagnóstico y control del asma**

El asma se ha mostrado como una de las causas de discapacidad más comunes a nivel mundial. Con unos costes de servicios de salud considerables y una importante pérdida de productividad de los afectados, así como una reducción significativa de su calidad de vida. Figura como la enfermedad crónica más común entre los niños y adultos jóvenes. Tanto la prevalencia global como la carga económica han ido en aumento en las últimas décadas. En la actualidad la prevalencia mundial en adultos se sitúa alrededor del 4,5% pero existe cierta heterogeneidad de esta tasa dependiendo del territorio, presentando la más alta Australia con un 21,5%. Un cuarto de la población con el síndrome asmático es también fumadora, y un 20% no ha recibido nunca tratamiento. No parece existir diferencias en prevalencia entre los

habitantes de zonas urbanas y zonas rurales (Sidney y Braman, 2006; To et al., 2012).

Según los dos estudios “The European Community Respiratory Health Survey” ECRHS, la prevalencia en adultos del asma en España se puede considerar media baja en comparación del resto de Europa (los países anglosajones son las que mostraron una tasa más elevada). Se encontraron grandes diferencias entre las ciudades españolas participantes en los estudios respecto a la prevalencia de sintomatología asmática, con la menor tasa en Galdakano y la mayor en Albacete. Entre ambos estudios (nueve años transcurridos entre uno y otro) se ha constatado un aumento del diagnóstico y tratamiento del asma, con el paso de una prevalencia de diagnóstico de 4,5% a una del 7,3% en el grupo no fumador (Grupo Español del Estudio Europeo del Asma, 1996; Urrutia et al. y Grupo Español ECRHS, 2007).

Por otro lado, se ha observado que la mortandad del asma se ha ido reduciendo de forma importante en las últimas décadas en España, estabilizándose en una tasa de mortalidad de 2,2 en esta década (Sánchez-Bahillo, García-Marcos, Pérez-Fernández, Martínez-Torres y Sánchez-Solís 2009). Respecto a la prevalencia en menores, en el “International Study of Asthma and Allergies Childhood” ISAAC se identificó, en España, una prevalencia de sintomatología asmática reciente (últimos 12 meses) de entre 7,1 % y 15,3% en las franjas de edad analizadas (6-7 años y 13-14 años), con una tasa significativamente mayor en las ciudades de la zona atlántica del país (Carvajal-Ureña et al., 2005).

Sobre la clínica del trastorno, se destaca el proceso inflamatorio persistente en las vías respiratorias junto a la hiper-reatividad bronquial. A pesar de configurar el patrón habitual del síndrome, no se ha clarificado del todo la relación entre ambos fenómenos, ni todos sus mecanismos subyacentes. Aunque se ha evidenciado la activación e interacción de una serie de células inflamatorias también presentes en las reacciones alérgicas convencionales (funcionando como una reacción inmunológica exagerada). La hiper-reatividad bronquial se ve condicionada por la inflamación pero se ha observado que existen también otros elementos que la determinan, tales como disfunciones neuroreguladoras, cambios estructurales, o factores hereditarios. Las dos anteriores condiciones conducen el evento fisiopatológico más característico del asma: el estrechamiento de la vías respiratorias reversible (con la consecuente reducción del flujo aéreo). Este evento recurrente genera una exacerbación sintomatológica que se evidencia con episodios de significativas dificultades respiratorias. Este estrechamiento puede generarse mediante diversos mecanismos, siendo

el más común la contracción del músculo liso bronquial como reacción defensiva ante estímulos ambientales. Otros mecanismos que provocan este estrechamiento y limitación del flujo aéreo son: edema de vía aérea (acumulación de líquido), hipersecreción de moco (generación de tapones), y cambios estructurales más estables y menos reversibles (como la remodelación vías respiratorias) (National Health, Lung and Blood Institute, NHLBI 2007; Sociedad Española de Neumología y Cirugía Torácica SEPAR et al., 2009).

Sobre el diagnóstico inicial, la Global Initiative for Asthma (GINA, 2014) indica que se debe seguir un procedimiento de tres fases, antes de prescribir ningún tratamiento (ya que los síntomas podrían verse encubiertos fácilmente por los fármacos):

- 1º- Identificar un patrón característico de síntomas respiratorios: suele observarse más de uno de los siguientes síntomas: sibilancias, disnea, tos y opresión torácica. Estos deben ser variables en intensidad y frecuencia a lo largo del tiempo, y a menudo parecen estar desencadenados por estímulos o condiciones ambientales.
- 2º- Explora el historial clínico previo y las condiciones físicas: antecedentes del paciente de rinitis alérgica, atopía o eccema son significativos, así como inicio de síntomas respiratorios en infancia. Presencia de asma o alergia en familiares también son relevantes a la hora de adjudicar el diagnóstico. La exploración física no debe reportar anomalías evidentes, exceptuando posibles sibilancias espiratorias al realizar la auscultación.
- 3º- Realizar pruebas de función pulmonar: se debe comprobar con instrumentos objetivos bien calibrados que existe una limitación del flujo aéreo y que esta es variable en el tiempo. Los indicadores más relevantes para ello son: la capacidad vital forzada (FVC), el volumen espiratorio forzado en 1 segundo ( $FEV_1$ ) y el flujo espiratorio máximo o Peak Flow (PEF) sobre los que se profundizará en el siguiente apartado. También se debe comprobar la reversibilidad de la limitación de la función pulmonar suministrando fármacos y volviendo a medir (broncodilatadores o corticosteroides inhalados).

Se organiza la evaluación mediante estas fases para poder realizar un correcto diagnóstico diferencial con otros trastornos, ya que los síntomas y signos del asma no son exclusivos del síndrome por sí solos, y es necesario documentar evidencias del patrón completo del asma (SEPAR et al., 2009; GINA, 2014). En la Tabla 16 se recogen las

principales patologías a tener en cuenta para el diagnóstico diferencial, según la Guía Vasca de Práctica Clínica sobre el Asma (Osakidetza Servicio Vasco de Salud, 2005). Respecto a esto, en los adultos, se destaca la Enfermedad Pulmonar Obstructiva Crónica EPOC por poseer una manifestaciones clínicas bastante similares al asma, y a menudo ser confundidos en contextos de atención primaria. El EPOC se diferencia del asma en: una relación directa y consistente con el tabaquismo, inicio más tardío (y mayor prevalencia cuanto mayor sea la edad), poca variabilidad de intensidad y presencia de síntomas, menor reversibilidad de la obstrucción y peor respuesta a corticoesteroides. Pese a ello en el ámbito clínico se ha detectado solapamiento de ambos síndromes en un pequeño porcentaje de los pacientes (Naberan, 2002; Cosio, Fiorentino y Scrimini, 2010).

**Tabla 16**

*Diagnostico diferencial del asma*

<b>Diagnostico diferencial en adultos</b>	<b>Diagnostico diferencial en menores</b>
- Enfermedad pulmonar obstructiva crónica EPOC	- Fibrosis quística
- Enfermedades cardíacas	- Disquinesias ciliar
- Enfermedad pulmonar intersticial	- Malformaciones cardíacas o pulmonares
- Embolismo pulmonar	- Problemas laríngeos
- Reflujo gastroesofágico	- Bronquiectasias
- Aspiración o obstrucciones localizadas de la vía aérea (tumores, cuerpos extraños ...)	- Tuberculosis
	- Infecciones respiratorias de las vías respiratorias bajas recurrentes
	- Reflujo gastroesofágico

Nota: Tabla generada a partir del contenido de Osakidetza (2005).

Actualmente no se cuenta con un tratamiento curativo del asma, por lo que la intervención terapéutica tiene como objetivo el control sintomatológico de la enfermedad. Desde hace tiempo, existen medios médicos para que en la mayoría de los casos se consiga un buen control del asma. Sin embargo, se ha evidenciado que esto no se consigue con demasiada frecuencia, estimándose que un 70% de los pacientes lo tienen mal controlado. Como responsables de esto se señalan a factores como: poca adhesión al tratamiento,

falta de competencias de autogestión o comorbilidad con otros trastornos médicos o psicológicos (Demoly et al., 2009; Martínez, 2010; Cano-De la Cuerda, Useros-Olmo y Muñoz-Hellín, 2010). Por ello, guías clínicas como la British Guideline on the Management of Asthma (SIGN y BTS, 2012) subrayan la relevancia de conseguir un consenso y alianza paciente-sanitario a la hora de gestionar la enfermedad, adaptando y ajustando el tratamiento a las condiciones particulares de la persona en cada momento. En esta misma guía se establece que se puede aspirar a un control completo de la enfermedad, siendo la meta del proceso terapéutico alcanzar y mantener este estado:

- No aparecen síntomas diurnos, ni se despierta de noche debido al asma.
- No se dan exacerbaciones ( crisis asmáticas).
- No hay necesidad de medicación de refuerzo.
- No hay limitaciones de actividad, incluido ejercicio físico.
- Mantiene una función pulmonar en franja de normalidad.
- Se sufren pocos efectos secundarios de la medicación.

La prevención secundaria, la terapia farmacológica, la inmunoterapia y la educación terapéutica, representan las principales estrategias de intervención para abordar la patología. La prevención secundaria consiste en controlar y evitar los factores de riesgo que provocan o agravan la sintomatología (elementos alérgenos ambientales, contaminación, tabaco, obesidad, bacteria neumococo, tipo de dieta etc...). Con respecto a la terapia farmacológica se diferencian dos grupos de medicamentos. Por un lado, los destinados a paliar exacerbaciones o crisis sintomatológicas (de uso puntual) llamados de rescate o de alivio, que se caracterizan por su acción broncodilatadora rápida (tales como los agonistas  $\beta_2$  adrenérgicos de acción corta o los anticolinérgicos). Y por otro, los que son prescritos para su uso sostenido para prevenir futuras exacerbaciones y generar un control sintomatológico a largo plazo, denominados de mantenimiento con efectos anti-inflamatorios o broncodilatadores de acción lenta (como los glucocorticoides o los agonistas  $\beta_2$  adrenérgicos de acción larga). Sobre la inmunoterapia (vacunas controladas de alérgenos) se señala como recomendable en los fenotipos de asma alérgica con rinitis alérgica o dermatitis atópica en la que se haya conseguido un control medio de los síntomas, pero debe tenerse constancia de que efectivamente existe esa condición alérgica (necesario realizar pruebas de

sensibilización). Respecto a la educación terapéutica se profundizara en ello en el ultimo apartado del capítulo (Larenas-Linnemann, 2009; SEPAR et al., 2009).

A lo largo del tiempo la gravedad del síndrome asmático puede ir variando, y es necesario una supervisión y reajuste de los parámetros de tratamiento según el estado del paciente en cada momento. A pesar de que la medicación en dosis altas es muy eficaz para remitir la sintomatología, sus efectos secundarios hacen que su prescripción se intente ubicar la dosis mínima para mantener el control de la enfermedad. Son necesarias por tanto revisiones periódicas del estado del paciente en las cuales se atiende a dos criterios fundamentalmente: la presencia e intensidad de síntomas y la función pulmonar (NHLBI, 2007; GINA, 2014).

#### **6.4.2. Medidas de la función pulmonar y espirometría**

El método más utilizado para comprobar la función pulmonar es la medición de la velocidad de salida del aire por los bronquios, que se realiza a través de una técnica denominada espirometría forzada. Esta técnica debe ser realizada por personal sanitario formado específicamente en ella y con instrumentos adecuadamente calibrados. Consiste en solicitar al paciente que realice una inspiración completa, mantenga el aire unos segundos y después expulse, (el técnico deberá dar la orden de forma enérgica y tajante) violentamente y forzando al máximo, todo el aire de sus pulmones a través de una boquilla conectada a un espirómetro. Durante la maniobra el paciente deberá permanecer sentado, erecto, con la espalda apoyada en respaldo y sin flexionar el cuello. A nivel internacional se ha realizado una estandarización del procedimiento y se han establecido unos criterios técnicos de aceptabilidad y reproducibilidad de la espiración. Estos implican el análisis del curso de la maniobra y una realización de mínimo tres maniobras ( un máximo de 8 si no se consideran aceptables). Para considerar fiable el procedimiento no deben diferir demasiado en su flujo cada maniobra aceptable. Los mejores valores de cada maniobra se utilizarán para obtener los indicadores espirométricos (American Thoracic Society ATS, 1995; Miller et al., 2005; García-Río et al., 2013).

El espirómetro registra dos curvas a partir de flujo de aire espirado: la curva volumen-tiempo, que relaciona el volumen de aire expresado en litros con el tiempo en segundos empleado para la espiración, y la curva flujo-volumen, que relaciona el flujo de aire expresado en litros/segundo (velocidad) con el volumen en litros espirado en cada instante. Gracias a estos registros se pueden observar y calcular una serie de indicadores de la función pulmonar, entre los que se destacan por su utilidad clínica los siguientes (Miller et al., 2005; Osakidetza Servicio Vasco de Salud, 2005; Burgos, 2008):

- Capacidad Vital Forzada (FVC): Es un indicador directo de la capacidad pulmonar. Expresa en litros el volumen total de aire expulsado en la espiración forzada.
- Volumen Máximo espirado en el primer segundo ( $FEV_1$ ): es un índice de flujo de aire por lo que se expresa en litros por segundo. Identifica el volumen de aire inspirado en el primer segundo. Muy utilizado como indicador clínico de la función pulmonar al compararlo con los valores teóricos ( ver más adelante).
- Relación porcentual entre  $FEV_1$  y FVC ( $FEV_1/FVC\%$ ): es el coeficiente entre los dos anteriores indicadores. Es un parámetro útil para identificar el nivel de obstrucción de las vías respiratorias. Si es inferior al 70% se considera una condición obstructiva clínica.
- Relación porcentual entre  $FEV_1$  y  $FEV_6$  ( $FEV_1/FEV_6\%$ ): Representa una alternativa más estable que el anterior como indicador para valorar la obstrucción. Se relaciona el  $FEV_1$  con el volumen de aire expulsado en los 6 primeros segundos, en vez usar el volumen total espirado.
- Flujo espiratorio máximo (PEF): también denominado Peak Flow. Es el máximo flujo ( litros por segundo) de aire conseguido en la espiración. Debido a que se muestra como un indicador fiable de la función pulmonar en el momento de realizar la prueba, se le otorga bastante utilidad clínica tanto en el diagnóstico inicial del asma (variabilidad pre tratamiento de función pulmonar) como en su seguimiento (comparación con valores teóricos).

La interpretación clínica de estos indicadores se realiza comparando los valores obtenidos en la espirometría con los valores teóricos que le corresponderían una persona sana de sus mismas características, es decir con unos valores de referencia poblacional. Para ello se divide el valor observado entre el valor real y se expresa como un porcentaje de

uno sobre el otro. En la interpretación del FEV1 y el PEF, existe consenso para señalar como umbral entre normalidad y patología al 80% sobre el valor de referencia. Y dentro del estado clínico: al intervalo entre 79% y 50 % se considera de afección de leve a moderada, y a valores inferiores al 50% como graves e indicativos de mal control de la enfermedad. Aunque la función pulmonar puede no presentar congruencia con la sintomatología de ese momento, si es un importante factor predictivo de exacerbaciones futuras. Por ello, su registro es relevante no solo en el diagnóstico sino también en el seguimiento de la enfermedad (SEPAR et al., 2009; GINA, 2014). En el contexto Español, la SEPAR (2002) recomienda que se utilicen como valores teóricos obtenidos en población normal en el Estudio Multicéntrico Barcelona (Roca et al., 1986).

### **6.4.3 Factores psicológicos en el asma**

En la patología asmática se ha reconocido el significativo impacto de factores psicológicos, actuando como un elementos influyentes en la exacerbación de la sintomatología y el mal control de la enfermedad. Incluso se ha llegado a señalar su relevancia en la génesis del asma y su mortalidad. También se observa una relación circular de estos factores con la enfermedad, actuando tanto como causa como consecuencia a lo largo de la vida de paciente. Y aunque no se ha clarificado del todo, se ha destacado un correlato fisiológico de la conexión de ciertos aspectos psicológicos y el asma, a través de la influencia del sistema nervioso en el sistema endocrino y el sistema inmunológico (Benítez, Molina y Camps, 2005; Suárez, Guadalupe y Del Olmo, 2010).

Lo factores psicológicos que se han mostrado empíricamente más influyentes en el asma se pueden agrupar en las siguientes áreas: las emociones y el estrés, la comorbilidad de trastornos y sintomatología psicopatológica, la familia del paciente y la adhesión al tratamiento (Posada y De Zubiría, 2004).

Sobre el primer factor se ha constatado que un porcentaje significativo de pacientes asmáticos generan broncoconstricción ante estímulos emocionales aversivos o sugestivos de agentes nocivos (Ritz, Wilhem, Meuret, Gerlach y Roth, 2011), y que el nivel de estrés vivenciado influye de manera importante en la respuesta inmunológica celular (Th2) de la condición inflamatoria asmática (Murali, Hanson y Chen, 2006). Al respecto se ha señalado al nervio vago como posible mediador de la reactividad respiratoria y las emociones, y a las

vías respiratorias superiores (inervadas por neuronas colinérgicas) como más sensibles a los cambios emocionales que las vías pequeñas (Coffman, 2002).

Con respecto a la comorbilidad psicopatológica se observa una mayor prevalencia de los trastornos psicológicos (especialmente ansiedad y depresión) en pacientes asmáticos que en la población general (Goodwin, Jacobi y Thefeld, 2003), y también peores indicadores de salud mental general (de Miguel et al., 2012). De forma más específica, se ha evidenciado una significativa prevalencia de trastornos psicológicos en estados de asma grave de riesgo vital (Kolbe, Fergusson, Vamos y Garrett, 2000). Además se ha asociado la presencia de trastornos de ansiedad o depresión con un peor control de la sintomatología asmática y una menor calidad de vida (Lavoie et al., 2005; Deshmukh, Toelle, Usherwood, O'Grady y Jenkins, 2008; Thomas, Butron, Moffat y Cleland, 2010). También se ha observado un mayor número de exacerbaciones y visitas a urgencias en pacientes asmáticos con depresión (Ahmedani, Peterson, Wells y Williams, 2013) y una mayor utilización de recursos sanitarios en los pacientes que cursan el asma junto alguna condición psicopatológica (ten Brinke, Ouwerkerk, Zwinderman, Spinhoven, y Bel, 2001).

Los factores familiares se han mostrado de especial relevancia en el inicio y curso de la enfermedad en el asma infantil (Benítez, Molina y Camps, 2005). Se ha asociado la presencia de disrupciones familiares junto a problemas del comportamiento en edades tempranas con un desarrollo precoz del asma (Calam, et al. 2003). Además en estudios prospectivos se ha podido comprobar que la interacción de un estilo de paternidad problemática (no adecuada atención del niño) con eventos vitales estresantes aumenta el riesgo de desarrollo de patología asmática (Kilnert et al., 2001). También se ha señalado a la familia como un importante modulador de la intensidad sintomatológica, relacionando determinadas estructuras familiares disfuncionales con manifestaciones asmáticas persistentes moderadas y graves (Chen y Escarce, 2008). Y los patrones de ingresos hospitalarios recurrentes se han asociado a la presencia de conflictos o tensiones familiares (Chen, Bloomberg, Fisher y Strunk, 2003).

Sobre la adhesión al tratamiento, se ha estimado que actualmente el 50% de los pacientes asmáticos que se mantienen bajo tratamiento a largo plazo, no cumplen con las prescripciones medicas (Boulet, Verloet, Magar y Foster, 2012). Esto tiene una repercusión directa sobre el control de la enfermedad, incrementándose el grado de inflamación de las

vías áreas, generando importantes limitaciones funcionales y aumentando el consumo de recursos asistenciales (Gillensen, 2007). Se señala como una de las principales causas de mal control del asma, y uno de los primeros elementos a explorar por el profesional cuando no existe una correcta respuesta al tratamiento (López-Viña, 2006; Martínez, 2010). Además se destaca la relevancia de su abordaje específico para generar un plan terapéutico efectivo en situaciones de asma grave (Rodríguez-Trigo, Plaza, Picado y Sanchis, 2008). Se han identificado una serie de factores que contribuyen a este mal cumplimiento (López-Viña et al. 2005; Vázquez-Cordero, 2007; GINA, 2014), relacionados fundamentalmente con:

- *La medicación y administración*: dificultades de uso, interferencia en vida cotidiana, pauta de compleja, mala comprensión de instrucciones, preocupaciones por efectos secundarios.
- *Actitud del paciente frente al tratamiento*.
- *El personal sanitario*: insatisfacción con la asistencia, relación con el profesional, infravaloración de la importancia del cumplimiento, cuestiones culturales o de personalidad.
- *Circunstancias involuntarias*: coste del tratamiento, olvidos, ausencia de rutinas o falta de habilidades de automanejo del paciente.

Para minimizar el impacto de estos factores psicológicos sobre el curso de la enfermedad se han planteado diferentes estrategias de intervención entre las que se destacan: la psicoterapia, la educación terapéutica y los programas de mejora de la adhesión al tratamiento (Jáuregui y Tejedor, 2004).

Se ha hipotetizado sobre la mejora de los resultados del tratamiento cuando el plan de intervención incluye estrategias psicoterapéuticas, resaltándose de forma específica: las técnicas de relajación, la terapia cognitivo conductual, las técnicas de bioalimentación, las intervenciones grupales y la terapia familiar (Posada y de Zubiría, 2004). Según la última revisión metanalítica Cochrane al respecto (Yorke, Fleming y Shuldham, 2009), no se ha podido fundamentar una evidencia consistente de los beneficios de la psicoterapia en adultos en el curso del asma, aunque esto se achaca más a las limitaciones de los estudios realizados que a una constatación empírica firme. A pesar de ello los autores resaltan resultados prometedores de mejoría de indicadores específicos: calidad de vida con terapia cognitivo

conductual, flujo espiratorio con bioretroalimentación y reducción de dosis de fármacos con técnicas de relajación. Estudios posteriores han reforzado la evidencia de beneficios de la terapia cognitivo conductual en adultos (Parry et al., 2012). En otra revisión Cochrane (Yorke y Shuldán, 2005) se identificó un probable beneficio de la psicoterapia familiar en casos de asma infantil de control difícil, aunque se señaló que el número de estudios eran aun escasos y con pequeños tamaños muestrales. En un estudio reciente se ha reportado una gran mejoría de síntomas clínicos y función pulmonar en niños asmáticos con familias disfuncionales al recibir terapia familiar (Berbesi, García, Segura, Posada y Serna, 2013).

Según la OMS (1998) la educación terapéutica no solo implica transmisión de información al paciente sobre su enfermedad, sino que además se busca capacitarlos en habilidades de autogestión de la enfermedad. Se entiende como un proceso interactivo en el cual el paciente participa en el control de su enfermedad (Korta et al., 2007). Recibir información sobre la enfermedad (acerca de fármacos, valoración de síntomas, identificación de factores desencadenantes, signos de exacerbación y de medios de prevención) es relevante para el paciente aunque por sí sola no ha mostrado suficiente para lograr una mejoría significativa en el control del asma (Gibson et al., 2002a). Por otro lado, el adiestramiento en autocuidado (con énfasis en la auto vigilancia de función pulmonar y síntomas, y pautas de manejo ante exacerbaciones y administración de la medicación), junto a un plan escrito del tratamiento, y una supervisión del médico regular, mejora de forma evidente las distintas medidas de resultado de salud (Gibson et al., 2002b). Se han mostrado además beneficios adicionales de tipo comunitario de este tipo de programas con menor tasa de hospitalizaciones, días de ausencia laboral del paciente y visitas a urgencias (Boyd et al., 2009; Gibson et al., 2002b). Un estudio reciente en el marco de la atención primaria, ha mostrado cómo la educación terapéutica, que contempla la información y la capacitación en habilidades de automanejo, mejora significativamente indicadores de calidad de vida y control de la enfermedad (Cano et al., 2014).

Estrechamente relacionado con lo anterior, la mejora de la adhesión al tratamiento en el asma se ha asociado a la presencia de elementos como: comunicación efectiva sanitario-paciente, plan terapéutico acordado por escrito, educación terapéutica, revisiones periódicas programadas, concienciación del paciente de riesgos de no cumplimiento y de su capacidad para controlar la enfermedad (Lopez-Viña, 2005; Lora-Espinosa, 2005; Vázquez-Cordero, 2007). También se ha evidenciado que la toma de decisiones acerca del tratamiento

compartida entre paciente y profesional mejora la adhesión y por tanto los resultados clínicos (Wilson et al., 2010). Pese a la importancia del abordaje de la adhesión ha existido una escasa validación empírica de programas concretos para mejorarla en el asma, siendo los estudios realizados en este sentido poco numerosos y en general con problemas metodológicos para demostrar eficacia (debido al pequeño tamaño muestral), aunque en ellos se señala también a los factores anteriormente mencionados (Haynes, Aclko, Sahota, McDonald y Yao, 2008; GINA, 2014).

**SEGUNDA PARTE**

**DESARROLLO EMPÍRICO**

---



## Capítulo 7

### Introducción a estudios empíricos

---

#### 7.1. Objetivo general

*Contrastar* la metodología de *Índice de Cambio Fiable de Jacobson y Truax* (ICF<sub>JT</sub>) con la *evaluación de resultados tradicional* (basada en estadísticos de significación, a partir de promedios grupales) y con otras *alternativas de análisis individual* de resultados en *diferentes contextos* de la *práctica clínica real* y ante *diferentes medidas de resultado*. Este objetivo general se concreta en los objetivos específicos de cada uno de los estudios que se van a presentar en esta parte empírica del trabajo.

#### 7.2. Hipótesis generales

*1º Hipótesis:* La metodología ICF<sub>JT</sub> se mostrará *más idónea y útil* que la metodología de significación de promedios de análisis de resultados. De esta primera hipótesis se derivan tres sub-hipótesis específicas:

- El ICF<sub>JT</sub> se evidenciará más útil para la evaluación del cambio en diferentes *medidas de resultado* (referidas a intensidad sintomática, funcionalidad en vida cotidiana y a normalidad biométrica) que la significación de promedios.
- El ICF<sub>JT</sub> será más útil en la evaluación de resultados de intervenciones en *distintas problemáticas de salud* (ansiedad, depresión y asma) y *contextos terapéuticos diferentes* (Centro privado psicología clínica, Centro de Salud Mental público y Centro de Salud ambulatorio público) que el procedimiento habitual basado en análisis promedios grupales.
- El ICF<sub>JT</sub> resultará más conveniente para la determinación de la *efectividad* de las intervenciones ante *diferentes condiciones de diseño experimental* (pre experimental sin grupo control, cuasi experimental con grupo control no equivalente, y ensayo clínico pragmático con grupo control) que el análisis de promedios habitual.

2° Hipótesis: La metodología ICF<sub>JT</sub> resultará *más adecuada* para la evaluación del cambio y su significación clínica que *otras alternativas metodológicas de análisis de resultados individuales* (GLN, EN, HA, TEI, DIT, ESG, DTW, IPP, IPI y CPM 50%).

### **7.3. Hipótesis específicas referidas a cada estudio**

*Estudio 1: En el marco del desempeño habitual de un centro privado de psicología clínica, en un estudio de resultados sin presencia de grupo control.*

La metodología de ICF<sub>JT</sub> reportará más información útil, tanto para la determinación de la efectividad del protocolo de intervención para *el tratamiento de la ansiedad* (en sentido de reducción de *intensidad sintomática*), como para la toma de decisiones del profesional, que la metodología de evaluación basada en estadísticos de significación y el cálculo del tamaño del efecto grupal.

*Estudio 2: En un contexto evaluativo de la práctica habitual de un Centro de Salud Mental público, en una situación sin grupo control.*

La metodología de ICF<sub>JT</sub> se mostrará de mayor utilidad para la evaluación de resultados terapéuticos de una unidad asistencial, *desde diferentes dimensiones de resultados* (síntomas, relaciones interpersonales, adecuación a rol social e insatisfacción vital general), que la metodología basada en estadísticos de significación de promedios y el cálculo del tamaño del efecto grupal.

*Estudio 3: En el marco de desempeño habitual de un terapeuta en un Centro de Salud Mental público, y en situación de estudio con grupo control no equivalente.*

La metodología de ICF<sub>JT</sub>, se mostrará superior a *otras alternativas de análisis de resultados*, tanto desde el punto de la utilidad clínica como de ajuste a la realidad, en una situación real de desempeño psicoterapéutico *para evaluar mejoría sintomática de tipo depresiva*. Esto se especifica en dos sub-hipótesis:

- La metodología de ICF<sub>JT</sub> reportará más información útil, tanto para la determinación de la *efectividad de la intervención* para tratamiento de la depresión, como para la

toma de decisiones del profesional, que la metodología de evaluación basada *en estadísticos de significación e indicadores del tamaño del efecto*.

- El método de ICF<sub>JT</sub> se mostrará más adecuado para la valoración del cambio de intensidad sintomática individual que las *propuestas alternativas de estimación de significación clínica* (GLN, EN, HA, TEI, DIT, ESG, DTW, IPP, IPI y CPM 50%).

*Estudio 4: En el contexto de un ensayo clínico aleatorizado con grupo control realizado en el ámbito ambulatorio de la Red Pública de Salud.*

La metodología de ICF<sub>JT</sub> se mostrará más sensible y reportará más información útil para la valoración de la efectividad de *una intervención educativa en Asma* (en el sentido de *mejoría espirométrica de capacidad pulmonar*), que la metodología habitual de análisis de promedios.

#### **7.4. Justificación y presentación de estudios empíricos**

Como se ha mostrado en los capítulos precedentes, en la actualidad se ha hecho manifiesta la necesidad de unir el ámbito del desempeño clínico habitual con la investigación acerca de resultados de intervenciones psicológicas. Haciendo posible una relación bidireccional que enriquezca la evidencia científica y mejore los medios terapéuticos con conclusiones centradas en la relevancia clínica en contextos reales (APA, 2006; Kazdin 2008; APA 2011; Henton, 2012). Además de esta necesidad de orientación hacia el estudio de la efectividad, se ha constatado empíricamente que la evaluación de resultados y su reporte supone un beneficio significativo para el propio proceso terapéutico de cada individuo (Lambert, 2010).

Por otro lado, las estrategias de análisis de resultados habituales en el ámbito de los ensayos clínicos (la significación estadística de las pruebas de contraste de hipótesis nula) han sido señaladas como insuficientes para generar conclusiones y evidencias prácticas para el desempeño clínico (Loannidis, 2005; Lambdin, 2012; Nuzzo, 2014). Por ello, diferentes instituciones han venido recomendado que en cualquier investigación se incluyan estadísticos acerca de la importancia del efecto encontrado, yendo más allá de su significación estadística (APA-Task Force on Statistical Inference, 1999; National Center for Education Statistics, 2002; International Committee of Medical Journal Editors 2013). Además de lo anterior,

en el entorno clínico habitual es difícil contar con las condiciones de los estudios clínicos controlados que serían las óptimas para el uso de estadísticos de significación de promedios grupales, tales como: tamaño muestral determinado, grupo control equivalente no tratado o cumplimiento de supuestos de distribución (Pascual, Frías, García 1996; Iraurgi, 2010).

Por todo lo anterior, se hace necesario el planteamiento de otras metodologías de análisis de resultados que sean compatibles con investigación en contextos terapéuticos naturales y que reporten información relevante acerca de la importancia del efecto observado. Y que además resulten útiles tanto al propio terapeuta en su toma de decisiones particular de cada paciente como al investigador en sus conclusiones acerca de la efectividad de las condiciones de tratamiento. Para dar respuesta a esta necesidad se vertebra la metodología análisis de cambio individual de Jacobson y Truax (1991) con su doble criterio de fiabilidad estadística y significación clínica. Esta propuesta ha sido la alternativa a los estadísticos clásicos más recomendada y con mayor aceptación en el ámbito de la investigación de resultados psicoterapéuticos (Ogles, Lunnen y Bonesteel, 2001; Wise, 2004; Lambert, Hansen y Bauer 2008). Sin embargo, a pesar de esta aceptación académica y su relativa antigüedad, sigue siendo una opción minoritaria frente los estadísticos de significación de promedios clásicos (Faulkner, Fidler, y Cumming, 2008; Hiller, Shindler y Lambert, 2014).

Pese a la existencia de literatura científica centrada en la valoración de la idoneidad del método ICF<sub>JT</sub>, sigue siendo necesaria la realización de un mayor número de estudios empíricos que desde diferentes contextos terapéuticos y con distintas medidas de resultado permitan vislumbrar una perspectiva más amplia del comportamiento de este tipo de metodología. Todo ello, con el fin de poder alcanzar un consenso para la investigación en resultados de la práctica clínica habitual, y estandarizar medios comunes para analizar la significación clínica (Jacobson, Robers, Berns y McGlinchey, 1999; Lambert y Ogles, 2009; Roussos, 2012). El presente trabajo se alinea con este objetivo y lo hace a través de las situaciones de desempeño clínico real de los cuatro estudios que lo componen.

El *estudio 1* se plantea desde una orientación bastante más inclinada al extremo pragmático que explicativo. Es decir, se centra en la valoración los resultados de la intervención terapéutica en condiciones de práctica real, en decremento del control experimental de la variables intervinientes (Thorpe et al., 2009). A pesar de ello, se tienen en cuenta ciertos criterios de selección muestral (como el diagnóstico), y de

homogeneización de condiciones de tratamiento (mismo centro, mismo terapeuta y mismo protocolo de intervención). Se trata de comprobar las posibles ventajas del uso del ICF<sub>JT</sub> (en contraste con el procedimiento habitual de significación de diferencia de medias), principalmente para la toma de decisiones del clínico acerca de su propio procedimiento y posteriormente considerar si además esta información es útil para generar evidencia general acerca del protocolo de tratamiento. Este estudio se realiza sin grupo control, en un centro privado, con una intervención psicoterapéutica de tipo cognitivo conductual, sobre pacientes con un diagnóstico del grupo de los trastornos de ansiedad y tomando como medidas de resultado cuatro escalas sintomáticas relacionadas con este diagnóstico.

El estudio 2 sigue manteniendo esa cercanía al polo pragmático en su planteamiento pero su foco es diferente al anterior estudio. En el ámbito de la evaluación de resultados, en asistencia sanitaria general, se ha señalado que para la determinación real de la efectividad y calidad de un determinado servicio asistencial es necesario recoger indicadores de resultado relevantes para el paciente que vayan más allá de lo inminentemente clínico y sintomático. Es decir, evaluar también indicadores más globales sobre el impacto de la intervención en sus condiciones vitales generales (Roland y Togerson, 1998; Badia y Bigorra, 2002). De forma más específica, lo anterior también se subraya como deseable para la determinación de la verdadera significación clínica de las intervenciones psicológicas (Kadzin, 2001). Por ello, este estudio incluye medidas de resultado que además de considerar la dimensión sintomática, añaden indicadores acerca de la adecuación social y funcionamiento interpersonal de los pacientes. Se pretende estudiar el comportamiento del ICF<sub>JT</sub> con este tipo de medidas de resultado, de nuevo frente al análisis clásico de significación de promedios. Se vuelve a plantear como un estudio sin grupo control en condiciones naturales pero en vez de evaluar un único procedimiento y a un solo terapeuta, se plantea como una evaluación de resultados de desempeño asistencial del Centro de Salud Mental (varios terapeutas e intervenciones) durante un periodo limitado de tiempo y en un extracto de pacientes determinado (predominio de síntomas ansioso depresivos).

El *estudio 3* tiene el mismo propósito que el primer estudio, y además añade el contraste del ICF<sub>JT</sub> con otras alternativas de análisis de resultados individuales. Aunque con anterioridad ya se habían realizado varios estudios comparando el funcionamiento de diferentes estrategias de análisis de cambio individual, estas investigaciones o se han limitado a contrastar únicamente las alternativas más cercanas al ICF<sub>JT</sub>, siendo estas las basadas en

el error estándar de medida (McGlinchey, Atkins y Jacobson 2002; Ronk, Hooke y Page, 2012), o se han generado con datos simulados por ordenador emulando una determinada configuración de condiciones de forma artificial (Atkins, Bedics, McGlichey y Beauchaine 2005; Pardo y Ferrer, 2013). Por ello, se ha decidido incluir en el contraste las principales propuestas metodológicas (en este sentido) resaltadas en la literatura psicológica actual (Iraurgi, 2010; Hiller, Schindler y Lambert, 2012; Pardo y Ferrer, 2013). Por otro lado, para este estudio se ha contado con un grupo control con tres propósitos: generar un situación experimental diferente a las anteriores (para el contraste con la metodología de análisis habitual), poder calcular algunos alternativas de análisis individual que necesitan de la existencia de un grupo control sin cambio sistemático, y contrastar los métodos en dos contextos de cambio muy diferentes. La investigación se realiza en el marco de un Centro de Salud Mental público, evaluando el cambio clínico producido en pacientes depresivos después de una intervención cognitivo conductual realizada por único un psicoterapeuta. La medida resultado elegida es un indicador de la intensidad sintomatológica depresiva. El grupo control se recluta en el mismo ámbito geográfico en población no clínica, pareado los participantes por sexo y edad con la muestra clínica.

El *estudio 4* comparte el objetivo con los anteriores estudios de contrastar el ICF<sub>JT</sub> con el análisis de resultados clásico. Sin embargo, se diferencia en gran medida de éstos por la naturaleza de la intervención, el tipo de diseño experimental, la problemática de salud implicada y la medida de resultado elegida. El ICF<sub>JT</sub> se planteó desde una perspectiva de cambio clínico en trastornos de tipo agudo, donde la exigencia de cercanía a la funcionalidad es alta. Por ello, la idoneidad de esta perspectiva en patologías que impliquen un curso crónico es necesario estudiarla y analizarla (Wise, 2004). Por otro lado, en el ámbito de la psicología de la salud también se ha subrayado la necesidad de aumentar la evidencia acerca de las intervenciones realizadas sobre trastornos físicos en condiciones clínicas reales (Taylor, 2004; Garzón y Caro, 2014). En el contexto concreto del asma, la educación terapéutica y las intervenciones destinadas a aumentar la adhesión al tratamiento se han mostrado relevantes para el buen manejo de la enfermedad (Korta et al., 2007; De Miguel, Barcina, Muñoz y Leal, 2008). Y aunque existe cierta evidencia que relacionan un mejor curso de la enfermedad con programas de intervención en este sentido (Levy et al., 2000; Gibson et al., 2002b; Pérez, Fernández, Fernández y Amigo, 2003), en general las diferentes propuestas realizadas al respecto han encontrado dificultades metodológicas (tales

como reducidos tamaños muestrales en los estudios) para demostrar efectividad mediante los análisis de resultados convencionales (Haynes, Ackloo, Sahota, McDonald y Yao, 2008). En este ensayo se trata de estudiar la viabilidad del ICF<sub>JT</sub> en un contexto diferente para el que fue desarrollado inicialmente, ya que se considera una alternativa útil (a los análisis habituales) que puede responder a la necesidades de este área de investigación. Se plantea un ensayo clínico con grupo experimental y grupo control aleatorizados de pacientes asmáticos, que están siendo supervisados en ambulatorios de la red pública de salud del mismo municipio. El grupo experimental habrá es atendido por profesionales sanitarios que han realizado un programa de formación focalizado en aumentar la adhesión al tratamiento y la educación para el manejo de la enfermedad. La medida de resultado será un índice biométrico que registra la capacidad pulmonar en una prueba de espirometria forzada.



## Capítulo 8

### Estudio 1. Evaluación de resultados en intervención sobre trastornos de ansiedad en ámbito privado

---

#### 8.1. Objetivo

El propósito del presente estudio es contrastar la metodología de evaluación de resultados tradicional (basada en el uso de estadísticos de significación a partir de promedios grupales) con la metodología de Índice de Cambio Fiable (explicitada en el capítulo 5 de la justificación teórica) en el contexto de la práctica real (condiciones naturales) sobre trastornos de ansiedad de un gabinete privado de psicología clínica.

#### 8.2. Hipótesis

La metodología de ICF<sub>JT</sub> reportará más información útil, tanto para la determinación de la efectividad del protocolo de intervención para *el tratamiento de la ansiedad* (en sentido de reducción de *intensidad sintomática*) como para la toma de decisiones del profesional, que la metodología de evaluación basada en estadísticos de significación y el cálculo del tamaño del efecto grupal.

#### 8.3. Metodología

##### 8.3.1. Participantes

La selección de los participantes del estudio se realizó entre los sujetos que acudieron al servicio privado solicitando atención psicológica para su malestar. En todos los casos fueron tratados por el mismo terapeuta, cumplían con los criterios para el diagnóstico de un trastorno de la sección de trastornos de ansiedad del DSM-IV TR y se siguió un protocolo de intervención equivalente en todos los casos.

La muestra quedó formada por 34 personas (22 mujeres), de entre los 18 y 71 años de edad, y una media de 45,05 años (DE= 15,99). Los diagnósticos de trastornos de ansiedad se ajustaron a una de las siguientes tres categorías: Fobia específica, Ataque de pánico o Trastorno de ansiedad generalizada.

### 8.3.2. Diseño

Se trata de un diseño cuasi-experimental pre-post facto de un solo grupo. También denominado diseño pre-experimental por no contar con grupo control. En la Tabla 17 se ofrece una representación gráfica del diseño (Cammbell y Stanley, 1966; Alaminos y Castejon, 2006; Montero y León, 2007).

**Tabla 17**

*Representación gráfica del diseño del estudio I*

Grupos	Medidas pre-programa	Programa experimental	Medidas post-programa
<i>Intervención</i>	O <sub>1</sub>	X	O <sub>2</sub>

A pesar de las amenazas a la validez interna que implica este diseño, se ha asumido para adecuarse a la situación natural del desempeño clínico del gabinete privado donde no es posible generar un grupo control. Se tendrá en cuenta en las conclusiones de estudio.

### 8.3.3. Variables dependientes

Como variables dependientes del estudio se han considerado una serie de dimensiones sintomáticas que atañen a los trastornos de ansiedad, según lo explicado en el capítulo 6. Para ello se han utilizado las puntuaciones de las siguientes escalas específicas del instrumento de evaluación aplicado, además de una escala general de gravedad sintomática:

- *Ansiedad*: Signos generales de ansiedad tales como nerviosismo, tensión o ataques de pánico.

- *Ansiedad Fóbica*: Respuesta persistentes de miedo irracional y desproporcionado ante estímulos o situaciones específicas.
- *Obsesión-Compulsión*: Pensamientos, acciones o impulsos no deseados que son vivenciados como imposibles de evitar.
- *Somatización*: alteraciones neurovegetativas en sistemas cardiovascular, respiratorio, gastrointestinal y muscular.
- *Gravedad sintomática general*: una medida generalizada de la intensidad del sufrimiento psicológico generada partir de la consideración de todas las dimensiones sintomáticas del instrumento.

### **8.3.4. Variable Independiente**

La variable independiente del estudio es un protocolo de intervención psicoterapéutica específico para el tratamiento de trastornos de ansiedad. El contexto de la intervención es un gabinete privado de psicología clínica de la Comunidad Autónoma Vasca. El terapeuta que la ejecuta es un varón, de mediana edad, nativo de dicho territorio, con 25 años de experiencia clínica y orientación cognitivo-conductual. La intervención tiene una duración de 8 sesiones, a las que se añaden dos o tres previas de entrevista y evaluación diagnóstica. Las sesiones tienen una duración de 50 minutos.

El protocolo psicoterapéutico sigue el esquema habitual de las intervenciones cognitivo-conductuales en ansiedad, consistente en: entrenamiento en técnicas de relajación, uso de auto-registros de reacciones de ansiedad (en unidades subjetivas de ansiedad de 0 a 100), afrontamiento de sesgos cognitivos, exposición interoceptiva, modelado de respuesta, exposición progresiva y generalización (Caballo, 2007). Además se destaca como eje de la intervención el entrenamiento autógeno de Schultz, tal y como es expuesto en el manual de Lindemann (1990), como habilidad de afrontamiento basada en la autosugestión y relajación (mediante visualización y auto mensajes) .

Una vez finalizadas las sesiones de evaluación y diagnóstico, el tratamiento en si mismo se sucede en tres fases:

- 1.-Entrenamiento: Se enseña al paciente relajación mediante respiración diafragmática, y se realiza el entrenamiento autógeno. Se seleccionan auto mensajes específicos

para el paciente (según pensamientos distorsionados), se vinculan estados de relajación a imágenes mentales mediante visualización y se realiza una modificación de pensamientos disfuncionales y cogniciones estresantes.

2.-Ensayos programados en consulta: Puesta en práctica de habilidades adquiridas en fase anterior con exposición gradual a situaciones o estímulos ansiógenos dentro del contexto de la consulta. Especial énfasis en la automonitorización, el autocontrol y el modelado de respuesta. Diferirá el nivel de especificación de esta fase dependiendo del tipo diagnóstico.

3.-Aplicación en vida cotidiana: afrontamiento en vivo de situaciones de la vida cotidiana “disparadoras” de la ansiedad (depende de diagnóstico) y generalización de estrategias de afrontamiento al día a día del paciente.

### **8.3.5. Instrumento**

Para medir las variables dependientes antes señaladas se ha recurrido a la adaptación española del Symptom Checklist 90 Revised, SCL90-R (González, De las Cuevas, Rodríguez y Rodríguez, 2002). Se trata de un instrumento de auto informe clínico configurado por un listado de 90 síntomas psiquiátricos que pretende medir el grado de malestar psicológico de una persona en un momento dado. El usuario deberá responder en que medida le ha perturbado o molestado durante los últimos siete días cada síntoma listado. Estos signos y síntomas se valorarán en una escala de intensidad de tipo Likert de cinco niveles (0= Nada, 1= Muy poco, 2= Poco, 3= Bastante y 4= Mucho).

A partir de estos 90 ítems se derivan 9 dimensiones primarias de sintomatología: Somatizaciones, Obsesiones y Compulsiones, Sensibilidad Interpersonal, Depresión, Ansiedad, Hostilidad, Ansiedad Fóbica, Ideación Paranoide y Psicoticismo. Sin embargo, 7 de los 90 ítems no se engloban en las anteriores dimensiones (falta de apetito, comer demasiado, dificultades para conciliar el sueño, sueño inquieto, despertarse de madrugada, ideas sobre la muerte o el hecho de morir y sentimientos de culpabilidad) aunque sí se tienen en cuenta en el cálculo de los índices generales. Con el cómputo total de todas las respuestas se pueden calcular tres índices globales: Índice Global de severidad, Índice Positivo de Malestar y Total de Síntomas Positivos. El primero es una medida general de la intensidad sintomática tomando en cuenta los valores asignados a cada ítem (de 0 a 4). El segundo es el

número de ítems con un valor asignado distinto de “0”, lo que viene a identificar la amplitud y diversidad sintomatológica que sufre la persona. Y el último es un indicador de la intensidad sintomática media y la tendencia de respuesta del sujeto, que se expresa como el cociente entre el total de los valores asignados a todos los ítems y el índice positivo de mal estar.

Aunque se aplicó el cuestionario de forma íntegra a cada paciente, como indicadores de resultado para este estudio se han seleccionado sólo las escalas que responden a las dimensiones explicitadas en el apartado de variables dependientes. En la Tabla 18 se recoge los estadísticos descriptivos de estas escalas según la validación española (tanto con población general como con población clínica) y su correspondiente coeficiente de fiabilidad. La puntuación directa de las escalas es resultado de dividir la suma de valores asignados a cada ítem entre el número de ítems que componen la escala.

**Tabla 18**

*Datos normativos nacionales y fiabilidad de escalas del SCL90-R utilizadas en el estudio 1*

Nombre de Escala (Número Ítems)	Población General	Población Clínica	Fiabilidad
	M (DE)	M (DE)	$\alpha$
<i>Ansiedad (10)</i>	0,52 (0,49)	1,64 (0,97)	0,90
<i>Ansiedad Fóbica (7)</i>	0,25 (0,36)	1,04 (1,05)	0,86
<i>Somatización (12)</i>	0,55 (0,55)	1,67(1,03)	0,87
<i>Obsesión-Compulsión (10)</i>	0,60 (0,51)	1,42 (0,99)	0,88
<i>Índice Global de Gravedad (90)</i>	0,51 (0,36)	1,49 (0,84)	0,96*

Notas: \* Dato de fiabilidad no presente en manual de adaptación española del instrumento pero sí en otros estudios. M= Media; DE = Desviación Estándar;  $\alpha$  = Coeficiente alpha (Consistencia Interna).

En el manual se refleja el valor de fiabilidad de cada escala, pero no informa sobre este mismo valor respecto al Índice Global de Gravedad (IGG). Debido a que es un dato necesario para la aplicación de la metodología de ICF, se han revisado otros estudios nacionales sobre valores normativos del instrumento. En ellos se ha encontrado que el IGG presenta un elevado coeficiente de fiabilidad con valores comprendidos entre 0,96 y 0,97 (Robles, Andre

y Peña, 2002; Caparrós-Caparrós, Villas-Hoz, Juan-Ferrer y Viñas-Poch, 2007). Este dato coincide con los valores hallados en estudios equivalentes en otros países de habla hispana (Gempp y Avendaño, 2008; Sánchez y Ledesma 2009). De forma cautelosa se ha seleccionado el valor más bajo de los informados en estos estudios (0,96) para su consideración en este trabajo.

### **8.3.6. Procedimiento**

Los participantes en el estudio dieron su consentimiento informado del uso limitado de sus datos (edad, género, diagnóstico general y medidas de resultado del SCL90-R) para fines de investigación. No se modificó el procedimiento habitual del centro donde la cumplimentación del SCL 90-R era ya un protocolo establecido para todos los pacientes que acudían al servicio.

Se generaron dos momentos de medición con el instrumento. Uno durante las sesiones iniciales de evaluación y diagnóstico, y otra una vez terminada la intervención psicoterapéutica. En ningún caso se superó las 12 sesiones (incluyendo evaluación, tratamiento y finalización) para el proceso clínico-asistencial completo, por lo que la temporización entre las medidas post tratamiento y pre tratamiento fue bastante homogénea en todos los casos. Las medidas de resultado fueron registradas en una base de datos con la identificación del paciente mediante su número de historia clínica, solo decodificable por su terapeuta, por lo que se guardó la confidencialidad y anonimato.

### **8.3.7. Análisis de datos**

En el análisis de datos se ha utilizado la aplicación software Statistical Package for Social Sciences SPSS de IBM en su versión 21.0. Para los cálculos y estadísticos no reflejados en las opciones estándar del programa, se ha recurrido a su configuración manual a través de la ventana de sintaxis.

Inicialmente, se realizó un análisis descriptivo y exploratorio de las variables dependientes tanto en el momento post tratamiento como en el pre tratamiento. Se estudió también la distribución de las diferencias entre ambos momentos en cada variable, comprobando su ajuste a la ley normal mediante la prueba *Kolmogorov-Smirnov*. No se

realizó ninguna prueba de homocedasticidad ya que el estudio solo implica a un solo grupo.

Para la metodología basada en significación estadística de promedios grupales se ha recurrido a la prueba paramétrica *t de Student* para muestras relacionadas, con el fin de confirmar si las diferencias de las medias entre el momento pre tratamiento y el post tratamiento en cada variable dependiente son significativas.. En todos los contrastes anteriores se asumió como umbral de significación  $p < 0,05$  que se corresponde con un nivel de confianza de 95%. Con este mismo nivel de confianza se calculó el Intervalo de Confianza de la diferencia de medias en cada variable.

Posteriormente se calculó el tamaño del efecto del cambio mediante de *d de Cohen* para cada variable dependiente. Se han asumido las indicaciones para situaciones con medidas intrasujetos pre y post intervención donde no exista grupo control. Para lo cual se utilizó como denominador de la diferencia entre las medias de los dos momentos la desviación típica de la medida pre tratamiento (Morris, 2008).

Para la metodología de  $ICF_{JT}$  (Jacobson y Truax 1991) se usaron los datos normativos de la validación del instrumento de medida en España (ver apartado Instrumento). En primer lugar, se calculó el umbral de significación clínica de la medida post tratamiento para cada escala bajo el criterio “C”, ya que se asumió solapamiento entre la distribución funcional y disfuncional en las variables analizadas. Siendo necesario el uso de los estadísticos descriptivos de dos muestras de referencia (clínica y general). En un segundo lugar se calculó el  $ICF_{JT}$  para cada paciente y en cada variable dependiente, teniendo en cuenta la fiabilidad informada en la validación del instrumento y la desviación estándar de la muestra clínica de referencia. Se consideró como cambio estadísticamente significativo aquellos valores de  $ICF_{JT}$  que en términos absolutos se alejasen del valor 1,96 (valor z de un nivel de confianza de 95% en la curva normal). Finalmente, atendiendo a los dos criterios generados se clasificó el cambio terapéutico acontecido. Para mayor detalle de este procedimiento consultar la última sección del capítulo 5.

## 8.4. Resultados

### 8.4.1. Análisis mediante contrastes de significación clásicos de promedios

En la Tabla 8.3 se recogen los principales datos descriptivos en lo referente a las variables dependientes, así como el análisis de la normalidad de la distribución de sus diferencias. De forma general, se observa una reducción de todas puntuaciones promedio post-tratamiento respecto a las de pre-tratamiento y se constata que las medias de todas las variables no consiguen superar la unidad después del tratamiento. La media post tratamiento más alta se observa en la variable **Ansiedad**. En el momento pre tratamiento la variable **Ansiedad Fóbica** presenta la mayor dispersión de puntuaciones, mientras que en el post tratamiento la muestra la variable **Obsesión-Compulsión**. Ninguna de las pruebas de normalidad de las distribuciones de diferencias ha resultado estadísticamente significativa, por lo que se mantiene la hipótesis nula que asume la distribución normal.

**Tabla 19**

Estadísticos descriptivos de la muestra y pruebas de normalidad de distribución del estudio 1

Variable resultado	Pre-test Min. / Max.	Pre-test M (DE)	Post- test Min. / Max.	Post- test M (DE)	Normalidad K.S. (p)
Ansiedad	0,40 / 3,80	2,16 (0,85 )	0 / 2,30	0,95 (0,71)	0,148 (0,057)
Ansiedad Fóbica	0 / 3,68	1,29 (1,02)	0 / 3,14	0,77 (0,84)	0,121 (0,200)
Somatización	0,25/ 3,59	1,46 (0,77)	0 / 2,25	0,94 (0,64)	0,126 (0,189)
Obsesión- Compulsión	0,20 / 2,99	1,67 (0,76)	0 / 3,69	0,97(0,87)	0,134 (0,126)
Índice Global de Gravedad	0,64 / 2,91	1,48(0,58)	0,14 / 2,34	0,86(0,54)	0,122 (0,200)

Notas: Min.= Puntuación Mínima; Max.= Puntuación Máxima; M = Media; DE = Desviación Estándar; K.S.= Kolmogorov-Smirnov (Prueba de normalidad de distribución diferencias); p = valor de significación del estadístico.

En la Tabla 20 se recogen los valores del estadístico de contraste realizado junto a su significación estadística. En la misma Tabla se recoge la diferencia de medias en cada variable y su Intervalo de Confianza del 95%. Así mismo, se presentan los valores de magnitud del efecto ( $d$  de Cohen). En todas las variables resultado se obtuvieron diferencias significativas entre las medias pre-test y post-test, al realizar la prueba *t de Student*. Siendo el sentido de todas estas diferencias significativas hacia el disminución sintomática. La mayor diferencia de medias se localiza en la variable **Ansiedad**, y la menor en la variable **Somatización**. Respecto a la magnitud del efecto de las diferencias, y siguiendo los criterios sugeridos por Cohen (1988) y Rosenthal (1996), se observa: un efecto moderado ( $d \geq 0,50$ ) en las variables **Ansiedad Fóbica** y **Somatización**, un efecto grande ( $d \geq 0,80$ ) en las variables **Obsesión-Compulsión** e **Índice Global de Gravedad**, y un efecto muy grande ( $d \geq 1,30$ ) en la variable **Ansiedad**.

**Tabla 20**

*Resultados de contraste de significación de diferencias de medias y magnitudes del efecto en estudio I*

Variable resultado	Pre-Post (IC 95%)	Estadístico: <i>t</i>	p.	Magnitud del efecto: <i>d</i>
Ansiedad	1,207 (0,89 ; 1,53)	7,628	0,000	1,42
Ansiedad Fóbica	0,523 (0,29 ; 0,77)	4,348	0,000	0,51
Somatización	0,521 (0,30 ; 0,74)	4,843	0,000	0,68
Obsesión-Compulsión	0,700 (0,46 ; 0,95)	5,876	0,000	0,92
Índice Global de Gravedad	0,624 (0,47 ; 0,78)	8,036	0,000	1,08

Notas: Pre-Post= diferencia de medias entre los dos momentos; IC= Intervalo de confianza de la diferencia;  $t = t$  de Student (Prueba diferencias de medias para muestra relacionadas); p = valor de significación del estadístico;  $d = d$  de Cohen.

#### 8.4.2. Análisis de cambio individual mediante Índice de Cambio Fiable

En las Tablas 21 a 25 se presentan las puntuaciones individuales y valores necesarios para clasificar el cambio acontecido según la metodología de Índice de Cambio Fiable en las variables resultado (una tabla para cada una). Las primeras columnas identifican demográficamente cada caso con la edad y el género. A continuación se muestran las puntuaciones pre y post tratamiento y su diferencia. En la columna “*Criterio por punto de corte Post  $\leq$* ” se especifica el primer criterio de clasificación, ubicando a cada caso a través de su puntuación post tratamiento respecto al punto de corte de funcionalidad (el cual se ha calculado previamente para cada variable resultado). Las dos siguientes columnas se refieren al segundo criterio que se corresponde al propio Índice de Cambio Fiable (ICF<sub>JT</sub>) y a sí este es de suficiente magnitud para igualar o superar el umbral de +/- 1,96 necesario para considerar que ha existido un cambio no debido al azar. La columna final aúna los dos criterios anteriores para clasificar el tipo de cambio clínico producido en el paciente.

En la variable **Ansiedad** (Tabla 21) se obtuvo el punto de corte de funcionalidad de 0,90 y el Error Estándar de la Diferencia (el denominador del ICF<sub>JT</sub>) en el valor 0,43. En 32 casos (94% de la muestra) se observa que la puntuación se reduce en el momento post tratamiento lo que señala a una mejora sintomática. Sin embargo, son 20 (59% de la muestra) los casos que han experimentado un cambio estadísticamente fiable según el ICF<sub>JT</sub>. De los anteriores, 15 casos (44% de la muestra) han conseguido situar su puntuación post tratamiento en la zona de funcionalidad, (*Recuperado*), y los otros cinco (15% de la muestra) no alcanzan el punto de corte de funcionalidad pese a constatar una reducción fiable de la puntuación (*Mejoría dentro de la disfuncionalidad*). Los restantes 14 casos (41 % de la muestra) no han experimentado un cambio de suficiente magnitud para ser considerados estadísticamente fiables. De estos, seis casos (18% de la muestra) hallan en a zona de funcionalidad (*Sin cambio dentro de la funcionalidad*) en el post tratamiento (tres ya lo estaban en el pre tratamiento), y ocho casos (23% de la muestra) no han salido de la zona de disfuncionalidad (*Sin cambio dentro de la disfuncionalidad*). En tres casos (9% de la muestra) se observa una puntuación pre tratamiento inferior al punto de corte de funcionalidad calculado, lo que señala que para estos casos el indicador se encontraba en valores no clínicos al inicio del estudio.

En la variable **Ansiedad Fóbica** (Tabla 22) se obtuvo el punto de corte de funcionalidad de 0,45 y el Error Estándar de la Diferencia en el valor 0,56. En 24 casos (71% de la muestra) se observa una reducción de la puntuación entre los dos momentos de medida. Pero solamente ocho casos (24% de la muestra) experimentan un cambio estadísticamente fiable, de los cuales cinco (15% de la muestra) se clasifican como *Recuperado*, y tres (9% de la muestra) como *Mejoría dentro de la disfuncionalidad*. Por otro lado, en 26 casos (76 % de la muestra) su evolución se ha considerado *Sin Cambio* fiable, ubicándose 13 de estos (38% de la muestra) *dentro de la funcionalidad* y los otros 13 (38% de la muestra) *dentro de la disfuncionalidad* en el momento post tratamiento. Aunque en cinco casos aumenta ligeramente la puntuación en el momento post tratamiento, el cambio no es de suficiente magnitud para ser considerado fiable, por lo que no hay casos que *Empeoran*. Es también destacable que 14 casos (41 % de la muestra) muestren una puntuación pre tratamiento inferior al punto de corte de funcionalidad calculado.

En la variable **Somatización** (Tabla 23) se obtuvo el punto de corte de funcionalidad de 0,94 y el Error Estándar de la Diferencia en el valor 0,53. En 26 casos se observa una reducción de la puntuación entre los dos momentos de medida (76% de la muestra). Pero solamente siete casos (21% de la muestra) experimentan un cambio estadísticamente fiable, de los cuales seis (18 % de la muestra) se clasifican como *Recuperado*, y uno (3% de la muestra) como *Mejoría dentro de la disfuncionalidad*. Por otro lado, en 27 casos (79 % de la muestra) su evolución se ha considerado *Sin Cambio* fiable, ubicándose 13 de estos (38 % de la muestra) *dentro de la funcionalidad* y los otros 14 (41 % de la muestra) *dentro de la disfuncionalidad*, en el momento post tratamiento. Aunque en siete casos aumenta ligeramente la puntuación en el momento post tratamiento, el cambio no es de suficiente magnitud para ser considerado fiable, por lo que no hay casos que *Empeoran*. En cinco casos (15% de la muestra) se observa una puntuación pre tratamiento inferior al punto de corte de funcionalidad calculado.

En la variable **Obsesión-compulsión** (Tabla 24) se obtuvo el punto de corte de funcionalidad de 0,89 y el Error Estándar de la Diferencia en el valor 0,48. En 29 casos se observa una reducción de la puntuación entre los dos momentos de medida (85 % de la muestra). De los anteriores son 12 (35% de la muestra) los casos que experimentan un cambio estadísticamente fiable, de los cuales 9 (26% de la muestra) se clasifican como *Recuperado*, y tres (9% de la muestra) como *Mejoría dentro de la disfuncionalidad*. Por

otro lado, en 22 casos (65% de la muestra) su evolución se ha considerado *Sin Cambio* fiable, ubicándose 10 de estos (30% de la muestra) *dentro de la funcionalidad* y los otros 12 (35% de la muestra) *dentro de la disfuncionalidad*, en el momento post tratamiento. Aunque en tres casos aumenta ligeramente la puntuación en el momento post tratamiento, el cambio no es de suficiente magnitud para ser considerado fiable, por lo que no hay casos que *Empeoran*. En seis casos (18% de la muestra) se observa una puntuación pre tratamiento inferior al punto de corte de funcionalidad calculado.

En la variable **Índice Global de Gravedad** (Tabla 25) se obtuvo el punto de corte de funcionalidad en la puntuación 0,80 de y el Error Estándar de la Diferencia en el valor de 0,24. En 33 casos se observa una reducción de la puntuación entre los dos momentos de medida (97% de la muestra). Pero son 20 (59% de la muestra) los casos que experimentan un cambio estadísticamente fiable, de los cuales 13 (38% de la muestra) se clasifican como *Recuperado*, y siete (21% de la muestra) como *Mejoría dentro de la disfuncionalidad*. Por otro lado, en 14 casos (41% de la muestra) su evolución se ha considerado *Sin Cambio* fiable; ubicándose seis de estos (18% de la muestra) *dentro de la funcionalidad* y los otros ocho (23% de la muestra) *dentro de la disfuncionalidad*, el momento post tratamiento. En dos casos aumenta ligeramente la puntuación en el momento post tratamiento, el cambio no es de suficiente magnitud para ser considerado fiable, por lo que no se pueden clasificar como *Empeorado*. En tres casos (9% de la muestra) se observa una puntuación pre tratamiento inferior al punto de corte de funcionalidad calculado.

**Tabla 21**

*Resultados individuales en variable Ansiedad según método de ICF<sub>JT</sub> en estudio 1*

Caso	Edad	Género	Puntuación Pre /Post	Sentido del cambio Post-Pre	Criterio por punto de corte Post ≤ 0,90	ICF <sub>JT</sub>	Tipo de Cambio según ICF <sub>JT</sub> ≥ 1,96	Clasificación de situación final Punto de Corte + ICF <sub>JT</sub>
1	54	Mujer	2,10 / 0,52	-1,58 Reduce	Funcionalidad	-3,64	Cambio	<b>Recuperado</b>
2	52	Mujer	2,50 / 1,60	-0,90 Reduce	Disfuncionalidad	-2,07	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
3	56	Mujer	2,10 / 2,10	0,00 Mantiene	Disfuncionalidad	0,00	No cambio	Sin cambio dentro de la disfuncionalidad
4	34	Mujer	3,80 / 1,23	-2,57 Reduce	Disfuncionalidad	-5,92	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
5	39	Hombre	2,82 / 0,30	-2,52 Reduce	Funcionalidad	-5,81	Cambio	<b>Recuperado</b>
6	61	Mujer	1,23 / 0,30	-0,93 Reduce	Funcionalidad	-2,14	Cambio	<b>Recuperado</b>
7	39	Mujer	2,50 / 1,00	-1,50 Reduce	Disfuncionalidad	-3,46	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
8	68	Mujer	0,70 / 0,60	-0,10 Reduce	Funcionalidad	-0,23	No cambio	Sin cambio dentro de la funcionalidad
9	64	Mujer	3,43 / 0,00	-3,43 Reduce	Funcionalidad	-7,91	Cambio	<b>Recuperado</b>
10	38	Mujer	2,50 / 2,10	-0,40 Reduce	Disfuncionalidad	-0,92	No cambio	Sin cambio dentro de la disfuncionalidad
11	41	Hombre	1,00 / 0,28	-0,72 Reduce	Funcionalidad	-1,66	No cambio	Sin cambio dentro de la funcionalidad
12	44	Hombre	2,82 / 0,50	-2,32 Reduce	Funcionalidad	-5,35	Cambio	<b>Recuperado</b>
13	46	Hombre	2,30 / 1,90	-0,40 Reduce	Disfuncionalidad	-0,92	No cambio	Sin cambio dentro de la disfuncionalidad
14	50	Mujer	1,40 / 0,50	-0,90 Reduce	Funcionalidad	-2,07	Cambio	<b>Recuperado</b>
15	24	Hombre	2,82 / 0,10	-2,72 Reduce	Funcionalidad	-6,27	Cambio	<b>Recuperado</b>
16	19	Hombre	2,30 / 0,28	-2,02 Reduce	Funcionalidad	-4,66	Cambio	<b>Recuperado</b>
17	56	Mujer	2,10 / 2,10	0,00 Mantiene	Disfuncionalidad	0,00	No cambio	Sin cambio dentro de la disfuncionalidad
18	71	Hombre	1,50 / 0,90	-0,60 Reduce	Funcionalidad	-1,38	No cambio	Sin cambio dentro de la funcionalidad
19	18	Hombre	0,40 / 0,20	-0,20 Reduce	Funcionalidad	-0,46	No cambio	Sin cambio dentro de la funcionalidad
20	63	Mujer	2,94 / 1,30	-1,64 Reduce	Disfuncionalidad	-3,78	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
21	23	Mujer	0,60 / 0,20	-0,40 Reduce	Funcionalidad	-0,92	No cambio	Sin cambio dentro de la funcionalidad
22	51	Mujer	1,30 / 0,10	-1,20 Reduce	Funcionalidad	-2,77	Cambio	<b>Recuperado</b>
23	27	Mujer	2,10 / 1,30	-0,80 Reduce	Disfuncionalidad	-1,84	No cambio	Sin cambio dentro de la disfuncionalidad
24	26	Hombre	1,50 / 0,50	-1,00 Reduce	Funcionalidad	-2,31	Cambio	<b>Recuperado</b>
25	68	Mujer	2,50 / 2,10	-0,40 Reduce	Disfuncionalidad	-0,92	No cambio	Sin cambio dentro de la disfuncionalidad
26	64	Hombre	2,82 / 0,90	-1,92 Reduce	Funcionalidad	-4,43	Cambio	<b>Recuperado</b>
27	39	Mujer	2,70 / 2,30	-0,40 Reduce	Disfuncionalidad	-0,92	No cambio	Sin cambio dentro de la disfuncionalidad
28	22	Mujer	2,10 / 1,70	-0,40 Reduce	Disfuncionalidad	-0,92	No cambio	Sin cambio dentro de la disfuncionalidad
29	45	Mujer	2,70 / 1,60	-1,10 Reduce	Disfuncionalidad	-2,54	Cambio	<b>Recuperado</b>
30	25	Mujer	3,43 / 0,60	-2,83 Reduce	Funcionalidad	-6,22	Cambio	<b>Recuperado</b>
31	55	Hombre	1,90 / 0,50	-1,40 Reduce	Funcionalidad	-3,23	Cambio	<b>Recuperado</b>
32	68	Mujer	3,43 / 1,30	-2,13 Reduce	Disfuncionalidad	-4,91	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
33	35	Hombre	1,50 / 0,50	-1,00 Reduce	Funcionalidad	-2,31	Cambio	<b>Recuperado</b>
34	47	Mujer	1,50 / 0,90	-0,60 Reduce	Funcionalidad	-1,38	No cambio	No cambio fiable en funcionalidad

**Tabla 22**

*Resultados individuales en variable Ansiedad Fóbica según método ICF<sub>JT</sub> en estudio 1*

Caso	Edad	Género	Puntuación Pre /Post	Sentido del cambio Post-Pre	Criterio por punto de corte Post ≤ 0,45	ICF <sub>JT</sub>	Tipo de Cambio según ICF <sub>JT</sub> ≥ 1,96	Clasificación de situación final Punto de Corte + ICF <sub>JT</sub>
1	54	Mujer	2,71 / 0,86	-1,85 Reduce	Disfuncionalidad	-3,33	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
2	52	Mujer	3,68 / 3,14	-0,54 Reduce	Disfuncionalidad	-0,97	No cambio	Sin cambio dentro de la disfuncionalidad
3	56	Mujer	0,00 / 0,57	+0,57 Aumenta	Disfuncionalidad	+1,03	No cambio	Sin cambio dentro de la disfuncionalidad
4	34	Mujer	2,09 / 1,00	-1,09 Reduce	Disfuncionalidad	-1,96	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
5	39	Hombre	1,46 / 0,00	-1,46 Reduce	Funcionalidad	-2,63	Cambio	<b>Recuperado</b>
6	61	Mujer	0,29 / 0,29	0,00 Mantiene	Funcionalidad	0,00	No cambio	Sin cambio dentro de la funcionalidad
7	39	Mujer	0,69 / 0,57	-0,12 Reduce	Disfuncionalidad	-0,22	No cambio	Sin cambio dentro de la disfuncionalidad
8	68	Mujer	0,29 / 0,14	-0,15 Reduce	Funcionalidad	-0,27	No cambio	Sin cambio dentro de la funcionalidad
9	64	Mujer	2,09 / 0,14	-1,95 Reduce	Funcionalidad	-3,51	Cambio	<b>Recuperado</b>
10	38	Mujer	2,71 / 2,09	-0,62 Reduce	Disfuncionalidad	-1,12	No cambio	Sin cambio dentro de la disfuncionalidad
11	41	Hombre	2,57 / 1,30	-1,27 Reduce	Disfuncionalidad	-2,29	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
12	44	Hombre	1,46 / 0,75	-0,71 Reduce	Disfuncionalidad	-1,28	No cambio	Sin cambio dentro de la disfuncionalidad
13	46	Hombre	2,00 / 2,43	+0,43 Aumenta	Disfuncionalidad	+0,77	No cambio	Sin cambio dentro de la disfuncionalidad
14	50	Mujer	0,43 / 0,50	+0,07 Aumenta	Disfuncionalidad	+0,13	No cambio	Sin cambio dentro de la disfuncionalidad
15	24	Hombre	1,46 / 0,00	-1,46 Reduce	Funcionalidad	-2,63	Cambio	<b>Recuperado</b>
16	19	Hombre	0,43 / 0,00	-0,43 Reduce	Funcionalidad	-0,77	No cambio	Sin cambio dentro de la funcionalidad
17	56	Mujer	0,00 / 0,57	+0,57 Aumenta	Disfuncionalidad	+1,03	No cambio	Sin cambio dentro de la disfuncionalidad
18	71	Hombre	0,43 / 0,29	-0,14 Reduce	Funcionalidad	-0,25	No cambio	Sin cambio dentro de la funcionalidad
19	18	Hombre	0,29 / 0,00	-0,29 Reduce	Funcionalidad	-0,52	No cambio	Sin cambio dentro de la funcionalidad
20	63	Mujer	1,00 / 0,43	-0,57 Reduce	Funcionalidad	-1,03	No cambio	Sin cambio dentro de la funcionalidad
21	23	Mujer	0,29 / 0,14	-0,15 Reduce	Funcionalidad	-0,27	No cambio	Sin cambio dentro de la funcionalidad
22	51	Mujer	0,43 / 0,43	0,00 Mantiene	Funcionalidad	0,00	No cambio	Sin cambio dentro de la funcionalidad
23	27	Mujer	2,71 / 2,71	0,00 Mantiene	Disfuncionalidad	0,00	No cambio	Sin cambio dentro de la disfuncionalidad
24	26	Hombre	0,43 / 0,43	0,00 Mantiene	Funcionalidad	0,00	No cambio	Sin cambio dentro de la funcionalidad
25	68	Mujer	0,43 / 0,86	+0,43 Aumenta	Disfuncionalidad	+0,77	No cambio	Sin cambio dentro de la disfuncionalidad
26	64	Hombre	1,46 / 1,00	-0,46 Reduce	Disfuncionalidad	-0,83	No cambio	Sin cambio dentro de la disfuncionalidad
27	39	Mujer	2,71 / 1,71	-1,00 Reduce	Disfuncionalidad	-1,80	No cambio	Sin cambio dentro de la disfuncionalidad
28	22	Mujer	1,71 / 0,43	-1,28 Reduce	Funcionalidad	-2,30	Cambio	<b>Recuperado</b>
29	45	Mujer	0,57 / 0,29	-0,28 Reduce	Funcionalidad	-0,50	No cambio	Sin cambio dentro de la funcionalidad
30	25	Mujer	2,09 / 0,14	-1,95 Reduce	Funcionalidad	-3,51	Cambio	<b>Recuperado</b>
31	55	Hombre	1,46 / 0,43	-1,03 Reduce	Funcionalidad	-1,85	No cambio	Sin cambio dentro de la funcionalidad
32	68	Mujer	2,71 / 2,09	-0,62 Reduce	Disfuncionalidad	-1,12	No cambio	Sin cambio dentro de la disfuncionalidad
33	35	Hombre	0,43 / 0,00	-0,43 Reduce	Funcionalidad	-0,77	No cambio	Sin cambio dentro de la funcionalidad
34	47	Mujer	0,43 / 0,00	0,00 Mantiene	Funcionalidad	0,00	No cambio	Sin cambio dentro de la funcionalidad

**Tabla 23**

*Resultados individuales en variable Somatización según método ICF<sub>JT</sub> en estudio I*

Caso	Edad	Género	Puntuación Pre /Post	Sentido del cambio Post-Pre	Criterio por punto de corte Post ≤ 0,94	ICF <sub>JT</sub>	Tipo de Cambio según ICF <sub>JT</sub> ≥ 1,96	Clasificación de situación final Punto de Corte + ICF <sub>JT</sub>	
1	54	Mujer	2,58 / 0,67	-1,91	Reduce	Funcionalidad	-3,64	Cambio	<b>Recuperado</b>
2	52	Mujer	2,80 / 2,10	-0,70	Reduce	Disfuncionalidad	-1,33	No cambio	Sin cambio dentro de la disfuncionalidad
3	56	Mujer	2,00 / 1,25	-0,75	Reduce	Disfuncionalidad	-1,43	No cambio	Sin cambio dentro de la disfuncionalidad
4	34	Mujer	1,58 / 0,88	-0,70	Reduce	Funcionalidad	-1,33	No cambio	Sin cambio dentro de la funcionalidad
5	39	Hombre	1,27 / 0,25	-1,02	Reduce	Funcionalidad	-1,94	No cambio	Sin cambio dentro de la funcionalidad
6	61	Mujer	0,42 / 0,00	-0,42	Reduce	Funcionalidad	-0,80	No cambio	Sin cambio dentro de la funcionalidad
7	39	Mujer	1,75 / 1,83	+0,08	Aumenta	Disfuncionalidad	+0,15	No cambio	Sin cambio dentro de la disfuncionalidad
8	68	Mujer	1,08 / 0,67	-0,41	Reduce	Funcionalidad	-0,78	No cambio	Sin cambio dentro de la funcionalidad
9	64	Mujer	1,50 / 0,17	-1,33	Reduce	Funcionalidad	-2,53	Cambio	<b>Recuperado</b>
10	38	Mujer	1,83 / 1,50	-0,33	Reduce	Disfuncionalidad	-0,63	No cambio	Sin cambio dentro de la disfuncionalidad
11	41	Hombre	0,42 / 0,56	+0,14	Aumenta	Funcionalidad	+0,27	No cambio	Sin cambio dentro de la funcionalidad
12	44	Hombre	1,42 / 1,42	0,00	Mantiene	Disfuncionalidad	0,00	No cambio	Sin cambio dentro de la disfuncionalidad
13	46	Hombre	0,50 / 1,27	+0,77	Aumenta	Disfuncionalidad	+1,47	No cambio	Sin cambio dentro de la disfuncionalidad
14	50	Mujer	2,00 / 0,92	-1,08	Reduce	Funcionalidad	-2,06	Cambio	<b>Recuperado</b>
15	24	Hombre	1,50 / 0,08	-1,42	Reduce	Funcionalidad	-2,70	Cambio	<b>Recuperado</b>
16	19	Hombre	0,50 / 0,42	-0,08	Reduce	Funcionalidad	-0,15	No cambio	Sin cambio dentro de la funcionalidad
17	56	Mujer	2,00 / 1,25	-0,75	Reduce	Disfuncionalidad	-1,43	No cambio	Sin cambio dentro de la disfuncionalidad
18	71	Hombre	1,25 / 0,58	-0,67	Reduce	Funcionalidad	-1,28	No cambio	Sin cambio dentro de la funcionalidad
19	18	Hombre	1,00 / 0,42	-0,58	Reduce	Funcionalidad	-1,10	No cambio	Sin cambio dentro de la funcionalidad
20	63	Mujer	2,10 / 1,50	-0,60	Reduce	Disfuncionalidad	-1,14	No cambio	Sin cambio dentro de la disfuncionalidad
21	23	Mujer	1,00 / 0,33	-0,67	Reduce	Funcionalidad	-1,28	No cambio	Sin cambio dentro de la funcionalidad
22	51	Mujer	1,00 / 0,33	-0,67	Reduce	Funcionalidad	-1,28	No cambio	Sin cambio dentro de la funcionalidad
23	27	Mujer	2,95 / 1,58	-1,37	Reduce	Disfuncionalidad	-2,61	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
24	26	Hombre	1,52 / 0,17	-1,35	Reduce	Funcionalidad	-2,57	Cambo	<b>Recuperado</b>
25	68	Mujer	1,33 / 1,25	-0,08	Reduce	Disfuncionalidad	-0,15	No cambio	Sin cambio dentro de la disfuncionalidad
26	64	Hombre	1,92 / 0,88	-1,04	Reduce	Funcionalidad	-1,98	Cambio	<b>Recuperado</b>
27	39	Mujer	0,25 / 0,50	+0,25	Aumenta	Funcionalidad	+0,48	No cambio	Sin cambio dentro de la funcionalidad
28	22	Mujer	0,67 / 1,33	+0,66	Aumenta	Disfuncionalidad	+1,26	No cambio	Sin cambio dentro de la disfuncionalidad
29	45	Mujer	3,59 / 2,95	-0,64	Reduce	Disfuncionalidad	-1,22	No cambio	Sin cambio dentro de la disfuncionalidad
30	25	Mujer	1,58 / 0,75	-0,83	Reduce	Funcionalidad	-1,58	No cambio	Sin cambio dentro de la funcionalidad
31	55	Hombre	1,27 / 1,25	-0,02	Reduce	Disfuncionalidad	-0,04	No cambio	Sin cambio dentro de la disfuncionalidad
32	68	Mujer	1,33 / 0,50	-0,83	Reduce	Funcionalidad	-1,58	No cambio	Sin cambio dentro de la funcionalidad
33	35	Hombre	1,12 / 1,33	+0,21	Aumenta	Disfuncionalidad	+0,40	No cambio	Sin cambio dentro de la disfuncionalidad
34	47	Mujer	0,56 / 1,00	+0,44	Aumenta	Disfuncionalidad	+0,84	No cambio	Sin cambio dentro de la disfuncionalidad

**Tabla 24**

*Resultados individuales en variable Obsesión-compulsión según método ICF<sub>JT</sub> en estudio*

Caso	Edad	Género	Puntuación Pre /Post	Sentido del cambio Post-Pre	Reduccion/Aumento	Criterio por punto de corte Post ≤ 0,80	ICF <sub>JT</sub>	Tipo de Cambio según ICF <sub>JT</sub> ≥   1,96	Clasificación de situación final Punto de Corte + ICF <sub>JT</sub>
1	54	Mujer	2,58 / 1,60	-0,98	Reduce	Disfuncionalidad	-2,02	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
2	52	Mujer	2,99 / 1,12	-1,87	Reduce	Disfuncionalidad	-3,86	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
3	56	Mujer	1,90 / 0,50	-1,40	Reduce	Funcionalidad	-2,89	Cambio	<b>Recuperado</b>
4	34	Mujer	1,12 / 0,50	-0,62	Reduce	Funcionalidad	-1,28	No cambio	Sin cambio dentro de la funcionalidad
5	39	Hombre	2,40 / 0,20	-2,20	Reduce	Funcionalidad	-4,54	Cambio	<b>Recuperado</b>
6	61	Mujer	1,30 / 0,10	-1,20	Reduce	Funcionalidad	-2,47	Cambio	<b>Recuperado</b>
7	39	Mujer	1,90 / 2,30	+0,40	Aumenta	Disfuncionalidad	+0,82	No cambio	Sin cambio dentro de la disfuncionalidad
8	68	Mujer	0,50 / 0,10	-0,40	Reduce	Funcionalidad	-0,82	No cambio	Sin cambio dentro de la funcionalidad
9	64	Mujer	1,60 / 0,00	-1,60	Reduce	Funcionalidad	-3,30	Cambio	<b>Recuperado</b>
10	38	Mujer	1,90 / 1,90	0,00	Mantiene	Disfuncionalidad	0,00	No cambio	Sin cambio dentro de la disfuncionalidad
11	41	Hombre	1,30 / 0,90	-0,40	Reduce	Disfuncionalidad	-0,82	No cambio	Sin cambio dentro de la disfuncionalidad
12	44	Hombre	2,50 / 1,40	-1,10	Reduce	Disfuncionalidad	-2,27	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
13	46	Hombre	2,86 / 3,69	+0,83	Aumenta	Disfuncionalidad	+1,17	No cambio	Sin cambio dentro de la disfuncionalidad
14	50	Mujer	1,90 / 0,00	-1,90	Reduce	Funcionalidad	-3,92	Cambio	<b>Recuperado</b>
15	24	Hombre	1,70 / 0,40	-1,40	Reduce	Funcionalidad	-2,89	Cambio	<b>Recuperado</b>
16	19	Hombre	2,00 / 0,70	-1,30	Reduce	Funcionalidad	-2,68	Cambio	<b>Recuperado</b>
17	56	Mujer	1,90 / 0,50	-1,40	Reduce	Funcionalidad	-2,89	Cambio	<b>Recuperado</b>
18	71	Hombre	1,30 / 0,40	-0,90	Reduce	Funcionalidad	-1,86	No cambio	Sin cambio dentro de la funcionalidad
19	18	Hombre	0,80 / 0,10	-0,70	Reduce	Funcionalidad	-1,44	No cambio	Sin cambio dentro de la funcionalidad
20	63	Mujer	2,99 / 2,58	-0,41	Reduce	Disfuncionalidad	-0,85	No cambio	Sin cambio dentro de la disfuncionalidad
21	23	Mujer	0,70 / 0,40	-0,30	Reduce	Funcionalidad	-0,62	No cambio	Sin cambio dentro de la funcionalidad
22	51	Mujer	0,60 / 0,30	-0,30	Reduce	Funcionalidad	-0,62	No cambio	Sin cambio dentro de la funcionalidad
23	27	Mujer	2,58 / 2,30	-0,28	Reduce	Disfuncionalidad	-0,58	No cambio	Sin cambio dentro de la disfuncionalidad
24	26	Hombre	0,90 / 0,60	-0,30	Reduce	Funcionalidad	-0,62	No cambio	Sin cambio dentro de la funcionalidad
25	68	Mujer	2,80 / 2,30	-0,50	Reduce	Disfuncionalidad	-1,03	No cambio	Sin cambio dentro de la disfuncionalidad
26	64	Hombre	1,50 / 0,85	-0,65	Reduce	Funcionalidad	-1,34	No cambio	Sin cambio dentro de la funcionalidad
27	39	Mujer	1,21 / 0,94	-0,27	Reduce	Disfuncionalidad	-0,56	No cambio	Sin cambio dentro de la disfuncionalidad
28	22	Mujer	1,70 / 1,40	-0,30	Reduce	Disfuncionalidad	-0,62	No cambio	Sin cambio dentro de la disfuncionalidad
29	45	Mujer	0,66/ 0,50	-0,16	Reduce	Funcionalidad	-0,33	No cambio	Sin cambio dentro de la funcionalidad
30	25	Mujer	2,30 / 0,70	-1,60	Reduce	Funcionalidad	-3,30	Cambio	<b>Recuperado</b>
31	55	Hombre	0,20 / 0,30	+0,10	Aumenta	Funcionalidad	+0,21	No cambio	Sin cambio dentro de la funcionalidad
32	68	Mujer	1,90 / 1,12	-0,78	Reduce	Disfuncionalidad	-1,61	No cambio	Sin cambio dentro de la disfuncionalidad
33	35	Hombre	1,30 / 1,20	-0,10	Reduce	Disfuncionalidad	-0,21	No cambio	Sin cambio dentro de la disfuncionalidad
34	47	Mujer	1,10 / 1,10	0,00	Mantiene	Disfuncionalidad	0,00	No cambio	Sin cambio dentro de la disfuncionalidad

**Tabla 25**

*Resultados individuales en Índice Global de Gravedad según método ICF<sub>JT</sub> en estudio 1*

Caso	Edad	Género	Puntuación Pre /Post	Sentido del cambio Post-Pre	Criterio por punto de corte Post ≤ 0,90	ICF <sub>JT</sub>	Tipo de Cambio según ICF <sub>JT</sub> ≥ 1,96	Clasificación de situación final Punto de Corte + ICF <sub>JT</sub>	
1	54	Mujer	2,35 / 1,28	-1,07	Reduce	Disfuncionalidad	-4,50	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
2	52	Mujer	2,35 / 1,60	-0,75	Reduce	Disfuncionalidad	-3,16	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
3	56	Mujer	1,28 / 0,78	-0,50	Reduce	Funcionalidad	-2,10	Cambio	<b>Recuperado</b>
4	34	Mujer	1,76 / 0,76	-1,00	Reduce	Funcionalidad	-4,21	Cambio	<b>Recuperado</b>
5	39	Hombre	1,73 / 0,19	-1,54	Reduce	Funcionalidad	-6,48	Cambio	<b>Recuperado</b>
6	61	Mujer	0,67 / 0,33	-0,34	Reduce	Funcionalidad	-1,43	No cambio	Sin cambio dentro de la funcionalidad
7	39	Mujer	1,76 / 1,04	-0,72	Reduce	Disfuncionalidad	-3,03	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
8	68	Mujer	0,64 / 0,33	-0,31	Reduce	Funcionalidad	-1,30	No cambio	Sin cambio dentro de la funcionalidad
9	64	Mujer	1,60 / 0,14	-1,46	Reduce	Funcionalidad	-6,15	Cambio	<b>Recuperado</b>
10	38	Mujer	1,71 / 1,40	-0,31	Reduce	Disfuncionalidad	-1,30	No cambio	Sin cambio dentro de la disfuncionalidad
11	41	Hombre	0,89 / 0,48	-0,41	Reduce	Funcionalidad	-1,73	No cambio	Sin cambio dentro de la funcionalidad
12	44	Hombre	1,98 / 1,19	-0,79	Reduce	Disfuncionalidad	-3,33	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
13	46	Hombre	2,18 / 2,34	+0,16	Aumenta	Disfuncionalidad	+0,67	No cambio	Sin cambio dentro de la disfuncionalidad
14	50	Mujer	1,28 / 0,33	-0,95	Reduce	Funcionalidad	-4,00	Cambio	<b>Recuperado</b>
15	24	Hombre	1,40 / 0,25	-1,15	Reduce	Funcionalidad	-4,84	Cambio	<b>Recuperado</b>
16	19	Hombre	1,40 / 0,67	-0,73	Reduce	Funcionalidad	-3,07	Cambio	<b>Recuperado</b>
17	56	Mujer	1,28 / 0,78	-0,50	Reduce	Funcionalidad	-2,10	Cambio	<b>Recuperado</b>
18	71	Hombre	1,11 / 0,44	-0,67	Reduce	Funcionalidad	-2,82	Cambio	<b>Recuperado</b>
19	18	Hombre	0,67 / 0,19	-0,48	Reduce	Funcionalidad	-2,02	Cambio	<b>Recuperado</b>
20	63	Mujer	2,55 / 1,76	-0,79	Reduce	Disfuncionalidad	-3,33	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
21	23	Mujer	0,82 / 0,38	-0,44	Reduce	Funcionalidad	-1,85	No cambio	Sin cambio dentro de la funcionalidad
22	51	Mujer	0,79 / 0,46	-0,33	Reduce	Funcionalidad	-1,39	No cambio	Sin cambio dentro de la funcionalidad
23	27	Mujer	2,91 / 1,60	-1,31	Reduce	Disfuncionalidad	-5,51	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
24	26	Hombre	1,03 / 0,48	-0,55	Reduce	Funcionalidad	-2,31	Cambio	<b>Recuperado</b>
25	68	Mujer	1,40 / 1,50	+0,10	Aumenta	Disfuncionalidad	+0,42	No cambio	Sin cambio dentro de la disfuncionalidad
26	64	Hombre	1,40 / 0,79	-0,61	Reduce	Funcionalidad	-2,57	Cambio	<b>Recuperado</b>
27	39	Mujer	1,28 / 1,13	-0,15	Reduce	Disfuncionalidad	-0,63	No cambio	Sin cambio dentro de la disfuncionalidad
28	22	Mujer	1,60 / 1,40	-0,20	Reduce	Disfuncionalidad	-0,84	No cambio	Sin cambio dentro de la disfuncionalidad
29	45	Mujer	1,40 / 1,13	-0,27	Reduce	Disfuncionalidad	-1,14	No cambio	Sin cambio dentro de la disfuncionalidad
30	25	Mujer	2,12 / 0,38	-1,74	Reduce	Funcionalidad	-7,32	Cambio	<b>Recuperado</b>
31	55	Hombre	0,88 / 0,48	-0,40	Reduce	Funcionalidad	-1,68	No cambio	Sin cambio dentro de la funcionalidad
32	68	Mujer	2,00 / 1,28	-0,72	Reduce	Disfuncionalidad	-3,03	Cambio	<b>Mejoría dentro de la disfuncionalidad</b>
33	35	Hombre	1,11 / 1,03	-0,08	Reduce	Disfuncionalidad	-0,34	No cambio	Sin cambio dentro de la disfuncionalidad
34	47	Mujer	1,03 / 0,82	-0,21	Reduce	Disfuncionalidad	-0,88	No cambio	Sin cambio dentro de la disfuncionalidad

En relación a la evolución de cada individuo y teniendo en cuenta todos los índices sintomáticos, se observan varios datos relevantes. En siete pacientes (21%) no acontece un cambio fiable en ninguna de las dimensiones, por lo que no es posible afirmar que se haya producido, en estos, un beneficio significativo después del tratamiento. Se trata de los casos número: 8, 10, 13, 21, 25, 27, 34. Es necesario matizar que en los casos 8 y 21 se observan niveles subclínicos en todos los indicadores (respecto al punto de corte de funcionalidad calculado) tanto al inicio como al final del tratamiento. Por otro lado, en 27 pacientes (79 % de la muestra) si se han constatado cambios fiables en sentido de mejoría como mínimo en una de las dimensiones sintomáticas. En tres de estos casos (número: 9, 15 y 16) se observa un estado final de recuperación en todos los índices, constándose un efecto generalizado de mejoría sintomática. En otros 17 casos de los que experimentan cambios fiables se evidencia que en una o más variables de resultado su evolución se clasifica como *Recuperado* (número: 1, 3, 4, 5, 6, 12, 14, 17, 18, 19, 22, 24, 26, 28, 29, 31, 33). Finalmente, en siete casos, a pesar de evidenciarse cambios fiables en uno o más índices, estos no llegan a situarse en la zona de funcionalidad clasificándose como de *Mejoría dentro de la disfuncionalidad* (número: 2, 7, 11, 12, 20, 23, 32), lo que implica una reducción significativa de la intensidad de algunos de los síntomas monitorizados pero no lo suficiente como para dejar de considerarse de gravedad clínica.

El análisis individual se puede pormenorizar con la interpretación del cambio de las diferentes dimensiones sintomáticas en cada paciente particular, pero debido a que resultaría redundante con la información reflejada en las tablas anteriores (y en exceso extenso), a continuación se ofrecerán solo un par de casos paradigmáticos como ejemplo de las valoraciones que se pueden realizar.

El caso número 24 es un hombre de 26 años de edad. Presenta puntuaciones de gravedad clínica antes del tratamiento (según el punto de corte de funcionalidad calculado previamente), en todas las dimensiones sintomáticas excepto en la **Ansiedad Fóbica**. Después del tratamiento se observa una reducción de la puntuación en todas estas, excepto, de nuevo, en la **Ansiedad Fóbica** que sigue manteniendo la misma puntuación. El cambio en las dimensiones **Ansiedad**, **Somatización** e **Índice Global de Gravedad** es de suficiente magnitud como para ser

considerado fiable, y además se sitúan en valores de funcionalidad después del tratamiento, por lo que se clasifica como *Recuperado*. Por contra, en la dimensión **Obsesión-compulsión** no se observa un cambio de suficiente magnitud como para no poder achacarlo a la variabilidad generada por la aplicación del instrumento de medida, pese a que la reducción de su puntuación llega a situar al paciente por debajo del punto de corte de la funcionalidad. Tanto en **Obsesión-Compulsión** como en **Ansiedad fóbica** el cambio se clasifica como *Sin cambio dentro de la funcionalidad*. Por todo lo anterior, se puede valorar que los signos generales de la ansiedad (tales como los ataques de pánico o el nerviosismo), las implicaciones corporales de esa ansiedad (fisiológicas y neurovegetativas) y el sufrimiento psicológico global han visto reducida su intensidad de forma fiable y clínicamente significativa después del tratamiento. Mientras que los pensamientos y comportamientos obsesivo-compulsivos no han mostrado una mejoría fiable. Respecto a las respuestas fóbicas específicas, no se evidencia ningún tipo de cambio, pero al no presentar niveles disfuncionales en ninguno de los dos momentos de medida, no se considera una dimensión de relevancia clínica en este caso.

El caso número 7 es una mujer de 39 años de edad. Antes del tratamiento, presenta puntuaciones de gravedad clínica en todas las dimensiones sintomáticas registradas. Después del tratamiento se observa que las puntuaciones en **Ansiedad**, **Ansiedad Fóbica** e **Índice Global de Gravedad** se reducen, mientras que las de **Somatización** y **Obsesión-Compulsión** aumentan. El cambio en las dimensiones **Ansiedad** e **Índice Global de Gravedad** es de suficiente magnitud para ser considerado fiable, pero no alcanza valores de funcionalidad después del tratamiento por lo que se clasifica como *Mejoría dentro de la disfuncionalidad*. Por otro lado, la reducción de la puntuación en **Ansiedad Fóbica** es demasiado pequeña como para no poder atribuirle a la variabilidad generada por la aplicación del instrumento de medida, y además sigue manteniéndose por encima del corte de funcionalidad. El aumento en la puntuación en **Somatización** y **Obsesión-Compulsión** no se puede considerar un empeoramiento significativo porque su cambio tampoco es de la suficiente magnitud como para ser considerado fiable. Por lo que la evolución en estas tres variables se clasifican como *Sin cambio dentro de la disfuncionalidad*. Debido a todo lo anterior, se puede señalar una importante reducción de la intensidad de los signos generales de la ansiedad y del sufrimiento psicológico global de la paciente, aunque estos aún se mantengan con un nivel de afección clínica. Esto evidencia, que

efectivamente, existe una buena evolución en estas dimensiones pero que es necesario reducir aún más la intensidad sintomática para que llegue a considerarse recuperada. Se observa además una mayor magnitud de mejoría en los síntomas ansiosos generales (cambio de -1,50 puntos e  $ICF_{JT} = -3,46$ ) que en el sufrimiento psicológico global (cambio de -0,72 puntos e  $ICF_{JT} = -3,03$ ). Por otro lado, las molestias corporales, los síntomas fóbicos y los pensamientos y comportamientos obsesivo-compulsivos no reducen su intensidad de forma fiable, por lo que no se puede señalar una evolución beneficiosa después del tratamiento en estas áreas.

## 8.5. Discusión y conclusiones

El método clásico de análisis de promedios grupales reporta unos resultados muy positivos con respecto al cambio producido en los pacientes. No solo se observan que las diferencias de medias entre los dos momentos resultan ser estadísticamente significativas, sino que además el tamaño del efecto ( $d$  de Cohen) no es inferior al valor de 0,50 en ninguna de las dimensiones. Sin embargo, estos resultados se muestran sobredimensionados en comparación con el método de análisis individual.

Aunque se encuentra cierta proporcionalidad entre los tamaños del efecto hallados en el análisis de diferencias de medias y los resultados obtenidos por la metodología de Índice de Cambio Fiable, resultan en su interpretación finalmente divergentes. Efectivamente, las dimensiones sintomáticas más generales muestran los mejores resultados, **Ansiedad** e **Índice Global de Gravedad**, atendiendo tanto al porcentaje de casos mejorados o recuperados como a la magnitud del efecto grupal, y peores en el caso de **Ansiedad Fóbica** y **Somatización**. Sin embargo, incluso en el caso de estas dos últimas dimensiones la lectura desde los promedios grupales orienta a valorar que ha acontecido un cambio relevante. En cambio, al observar la proporción de casos que realizan un cambio no atribuible a la variabilidad de la medida (según  $ICF_{JT}$ ), apenas se alcanza una cuarta parte de los casos de la muestra en ambas dimensiones. Y a la hora de valorar su significación clínica el número de casos que pueden considerarse recuperados (reducida su sintomatología de forma fiable a niveles no clínicos) es aún menor: 15 % en **Ansiedad Fóbica** y 18% en **Somatización**. Ante estos datos es más difícil mantener una interpretación tan positiva de los resultados de forma grupal, aunque haya individuos concretos

que sí se hayan visto notablemente beneficiados. En la variable **Obsesión-compulsión** se observa también discrepancia en el hecho de resultar un tamaño del efecto grupal grande ( $d=0,92$ ), pero con una proporción de casos de cambio fiable más bien moderada-baja (35%). Por otro lado, respecto a la gran magnitud de cambio registrada en **Ansiedad** ( $d= 1,42$ ), en el método ICF<sub>JT</sub> se traduce en que tres de cada cinco casos han experimentado un cambio a mejor estadísticamente fiable, y que un 44% de la muestra se pueda considerar recuperada después del tratamiento (cambio similar al registrado en **Índice Global de Gravedad** aunque con una proporción un poco menor de recuperados). Resultados importantes en sentido de mejoría clínica aunque matizados por el hecho de que una parte de la muestra no evidencia de hecho cambio fiable, y que a nivel clínico más de un 1/3 de los pacientes se encuentre en una situación de gravedad clínica después del tratamiento.

Se observa por tanto, una lectura de resultados globales más conservadora en el método ICF<sub>JT</sub>, que en el de análisis de promedios grupales. En este último se hace evidente que la magnitud del cambio de alguno de los casos compensa la situación de poco cambio de otros casos, distorsionando el análisis por el uso de medias. Algo fácilmente observable en **Ansiedad Fóbica**, donde no sólo la mayoría de los cambios en dirección a la funcionalidad son considerados no fiables, sino que además en el 29% de los casos o no existe ningún cambio (misma puntuación en ambos momentos) o llega agravarse ligeramente. Esto contrasta con el hecho de que la diferencia de medias entre los dos momentos resulta ser estadísticamente significativa en el sentido de a la reducción sintomática.

Con respecto a la interpretación de la importancia clínica de los resultados, de nuevo la consideración de promedios resulta ser engañosa. Los tamaños del efecto en forma de  $d$  de Cohen, señalan a una magnitud relevante en todas las variables de resultado, resultando ser muy sobresaliente en **Ansiedad**. Sin embargo, atendiendo al criterio de punto de corte de funcionalidad del método ICF<sub>JT</sub>, en ninguna de las dimensiones se alcanza el 50% de casos clasificados como *Recuperado* (44% es el máximo en Ansiedad), siendo esta la situación de cambio la que se podría considerar de significación clínica. Lo que relativiza los tamaños del efecto grupales hallados. Esta discrepancia se debe a que aunque en ambos métodos se está atendiendo la magnitud del cambio producido, usan criterios de análisis diferentes. Como

señalan Preacher y Kelley (2012) no es lo mismo generar un indicador estandarizado del efecto ( $d$  de Cohen) que valorar la importancia sustantiva de ese efecto en su contexto (significación clínica). No es de extrañar que Kazdin (2001) considere insuficiente la lectura de estos indicadores grupales estandarizados, por no tener en cuenta el paciente individual, ni las particularidades de las variables clínicas analizadas. El ICF<sub>JT</sub> plantea un criterio de comparación normativa respecto a la variable resultado que si contextualiza el efecto, resultando ser, como se observa en el estudio, menos indiscriminado y arbitrario que la  $d$  de Cohen.

Después de realizar una interpretación colectiva de los resultados (proporciones según clasificación del cambio), con el método ICF<sub>JT</sub> se puede generar un nivel de análisis individual (como ha quedado ejemplarizado al final del apartado del resultados) imposible para los análisis de significación de promedios. Esto permite observar la evolución diferencial de cada dimensión en el paciente, y tener en cuenta cuales de las áreas sintomáticas tienen relevancia clínica antes y después del tratamiento. Sirviendo al clínico para una toma de decisiones inmediata que mediante la consideración análisis grupal no se vería facilitada. De esta forma se hacen evidentes circunstancias como la del caso número 21, que efectúa una mejoría fiable y clínicamente significativa de la dimensión **Obsesiva-Compulsiva**, pero que no se beneficia de ninguna cambio en el índice genérico de **Ansiedad** que permanece en gravedad clínica, evento que se diluye en la consideración grupal donde a dimensión **Ansiedad** parece ser la dimensión con mayor mejoría después del tratamiento.

Además al tratarse de un procedimiento que compara a cada individuo con valores de referencia normativos, cada caso es independiente en análisis del resto del grupo evaluado y, por tanto, susceptible de ser registrado y acumulado en bases de datos. De esta forma, se puede enriquecer el análisis de resultados del protocolo de intervención con futuros casos, e incluso realizar comparativas entre diferentes grupos de perfiles de pacientes (características sociodemográficas, predominio de diferentes dimensiones sintomáticas etc...) o incluirse esos datos en redes de investigación de práctica habitual como las impulsadas por Castaguay (2011) y poder estratificar análisis según un gran número de variables. Para esta posibilidad de integrar casos de diferentes momentos, los estadísticos de significación clásicos verán muy limitada su utilidad, ya que al estar muy condiciones por el número de sujetos que se analiza cada vez, una

mínima diferencia podrá seguir resultando significativa si se cuenta con el suficiente número de casos registrados (Manterola y Pineda, 2008; Everitt y Skronnal, 2010).

La no existencia de un grupo control y la confirmación de presencia de perfiles sintomáticos bastante heterogéneos (dentro de la problemática ansiosa) complican realizar conclusiones firmes sobre los efectos de la intervención (eficacia). Sin embargo, partiendo de la situación de estudio de extremo pragmático (Thorpe et al. 2009), la metodología de ICF<sub>JT</sub> reporta mucha más información útil y fiable para la determinación de la efectividad del tratamiento que la metodología basada en promedios grupales. Al no existir un grupo de comparación para controlar el cambio espontáneo de los pacientes (el que no se debe a la intervención), la prueba *t* de Student y la *d* de Cohen se muestran limitadas e indiscriminadas para valorar la significación del efecto, ya que el criterio matemático se apoya únicamente en el tamaño muestral y en la propia diferencia de medias. EL ICF<sub>JT</sub> presenta un doble criterio que genera un análisis más riguroso. Por un lado, la variabilidad del instrumento de medida se incluye en el error del denominador de la diferencia de la diferencia entre los dos momentos. Y por otro, se compara las puntuaciones de cada paciente con valores normativos para poder darle una interpretación clínica contextualizada.

En conclusión, la metodología de ICF<sub>JT</sub> ha mostrado un análisis de los resultados colectivos más ajustado a la realidad del cambio clínico producido que la metodología habitual de análisis de promedios. Además el ICF<sub>JT</sub> con el análisis particular de cada caso presenta dos ventajas frente al método clásico: permite una lectura individual de la importancia clínica de la evolución del paciente (ayudando al clínico en su toma de decisiones inmediata) y al ser la unidad de análisis individual e independiente, la posibilidad de registrar y acumular casos en una base de datos para ir enriqueciendo la evidencia acerca de ese protocolo de intervención en un futuro.



## Capítulo 9

### Estudio 2. Evaluación de resultados en un Centro de Salud Mental del ámbito publico

---

#### 9.1. Objetivo

Este estudio tiene como objetivo comparar la metodología de Índice de Cambio Fiable de Jacobson y Truax (1991) con la metodología habitual de análisis de promedios, tomando de como variables de resultado indicadores que van más allá de la reducción sintomática (aunque también se incluyen). Realizándose en el ámbito del desempeño terapéutico cotidiano en un Centro de Salud Mental publico.

#### 9.2. Hipótesis

La metodología de ICF<sub>JT</sub> se mostrara de mayor utilidad para la evaluación de resultados terapéuticos de una unidad asistencial, *desde diferentes dimensiones de resultados* (síntomas, relaciones interpersonales, adecuación a rol social e insatisfacción vital general), que la metodología basada en estadísticos de significación de promedios y el cálculo del tamaño del efecto grupal.

#### 9.3. Metodología

##### 9.3.1 Participantes

Los participantes el estudio fueron captados entre las personas que acudieron, con demanda de asistencia terapéutica, a un Centro Comunitario de Salud Mental. Inicialmente aceptaron participar en el estudio 139 personas, sin embargo, 39 abandonaron el tratamiento antes de poder realizar la medida post tratamiento. En la Tabla 26 se presentan las características sociodemográficas, diagnóstico y puntuaciones de OQ-45 pre tratamiento, tanto de la muestra de

estudio como el grupo d.e abandono. Como se puede observar no se encontraron diferencias significativas en las variables registradas entre ambos grupos de paciente.

**Tabla 26**

*Caracterización y comparación de grupo estudio y grupo de casos abandono del estudio 2*

		Grupo Estudio (n= 100)		Grupo Abandono (n= 39)		Prueba	g.l.	p
		n	%	n	%			
<b>Sexo</b>	<b>Hombre</b>	36	36	24	61	0,07	1	0,787
	<b>Mujer</b>	64	64	15	39			
<b>Estado Civil</b>	<b>Soltero</b>	46	46	23	59	2,90	3	0,407
	<b>Casado</b>	45	45	15	38,5			
	<b>Separados</b>	8	8	1	2,5			
	<b>Viudos</b>	1	1	0	0			
<b>Nivel de Estudios</b>	<b>Primarios</b>	4	4	0	0	1,92	3	0,589
	<b>Secundarios</b>	37	37	17	43,6			
	<b>Bachiller</b>	34	34	13	33,3			
	<b>Superiores</b>	25	25	9	23,1			
<b>Situación Laboral</b>	<b>Trabaja</b>	63	63	24	61,5	3,72	2	0,155
	<b>No trabaja</b>	19	19	12	30,8			
	<b>Otros</b>	18	18	3	7,7			
<b>Grupo Diagnostico / Trastornos de</b>	<b>Animo</b>	32	32	14	35,9	2,72	3	0,437
	<b>Ansiedad</b>	37	37	9	23,1			
	<b>Adaptativos</b>	17	17	8	20,5			
	<b>Control de impulsos</b>	14	14	8	20,5			
		M	DE	M	DE	t	g.l.	p
<b>Edad</b>	<b>Años</b>	33,46	8,40	32,31	7,69	1,03	137	0,306
<b>Puntuación OQ-45 Pre-tratamiento</b>	<b>Total</b>	77,52	28,46	81,64	35,24	-0,65	58,35	0,517
	<b>Síntomas</b>	43,78	19,95	46,79	24,35	-0,69	58,94	0,494
	<b>R. Interpersonal</b>	17,54	6,95	18,03	7,38	-0,36	137	0,717
	<b>Rol social</b>	16,20	5,17	16,82	6,28	-0,60	137	0,551

Notas: la prueba para el caso de variables nominales el Ji cuadrado y para variables continuas la t de Student con aplicación de las pruebas de corrección en caso de incumplimiento de requisitos. M – Media; DE – Desviación Estándar; g.l. – grados de libertad; p – valor de probabilidad

### 9.3.2. Diseño

Se trata de un diseño cuasi-experimental pre-post facto de un solo grupo. También denominado diseño pre-experimental por no contar con grupo control. En la Tabla 27 se ofrece una representación gráfica del diseño (Cammbell y Stanley, 1966; Alaminos y Castejon, 2006; Montero y León, 2007).

**Tabla 27**

*Representación gráfica del diseño del estudio 2*

Grupos	Medidas pre-programa	Programa experimental	Medidas post-programa
<i>Intervención</i>	O <sub>1</sub>	X	O <sub>2</sub>

A pesar de las amenazas a la validez interna que implica este diseño, se asume por ser el único posible en el contexto de estudio, y coherente con los objetivos de evaluar los resultados en situación de práctica cotidiana.

### 9.3.3. Variables dependientes

Las variables dependientes contempladas en este estudio son distintas dimensiones de resultados terapéuticos, que reportan no solo el posible cambio sintomático sino que también el impacto sobre la vida cotidiana del paciente. En los capítulos 3 y 4 se resaltó la necesidad de tener en cuenta diversos indicadores más allá de los signos directos de la patología a la hora de valorar la efectividad de las intervenciones. Por ello, se ha elegido un instrumento de medida que tiene una perspectiva de los resultados más amplia que otras herramientas. De las escalas de este instrumento se derivan las siguientes variables:

- *Síntomas*: consideración general del distress psicopatológico o malestar intrapsíquico, con especial protagonismo de los síntomas afectivos (ansioso-depresivos) y el abuso de sustancias.
- *Relaciones interpersonales*: medida de la insatisfacción y problemática con las relaciones con otras personas. Evaluando fricción, aislamiento, rechazo e inadecuación con respecto a amigos, familia y pareja.
- *Rol Social*: reflejo de falta de rendimiento social entendido como inadecuación o conflicto en el desempeño de tareas relacionadas con el empleo, el ocio y los roles familiares (nivel de afectación de la habilidad para trabajar, disfrutar o amar).
- *Cómputo Total*: índice global de la valoración subjetiva del estado de insatisfacción vital y del nivel de funcionalidad en el entorno. Para lo cual se tienen en cuenta las tres dimensiones anteriores.

#### **9.3.4. Variable Independiente**

En el presente estudio se establece como variable independiente la atención asistencial realizada por cuatro terapeutas, en el marco de un Centro de Salud Mental público de la provincia de Vizcaya. Los terapeutas son dos hombres y dos mujeres, nativos de dicho territorio con más de 15 años de experiencia clínica. Cada clínico ajustó la modalidad de tratamiento en función de la evaluación de cada paciente. Se recurrió en la mayoría de los casos a técnicas basadas en el modelo cognitivo-conductual, aunque dentro de las posibilidades que ofrece la intervención psicoterapéutica en el ámbito público.

El formato básico de intervención se establece en sesiones de 40 minutos de psicoterapia y 15 minutos de seguimiento farmacológico (si este último fuese necesario). Con una periodicidad quincenal de sesiones cada dos semanas, aunque con una distancia de una única semana entre las dos primeras sesiones dedicadas a evaluación. Tras realizar esta evaluación diagnóstica los pacientes reciben tratamiento psicoterapéutico, y farmacológico en caso de necesitarlo. Para el estudio se acoto los cuatro primeros meses de esta intervención.

### 9.3.5. Instrumento

Las dimensiones de resultado, antes explicitadas en el apartado de variables dependientes, se han medido mediante el Outcome Questionnaire 45 (OQ-45). Se trata de un cuestionario de auto informe desarrollado por Lambert y su equipo (Lambert et al., 1996) con el objetivo específico de medir resultados psicoterapéuticos, y ser aplicado en más de una ocasión en el curso de proceso terapéutico. Aunque muestra validez concurrente con instrumentos específicamente psicopatológicos, el mismo autor advierte que no debe ser utilizado con fines diagnósticos, ya que su meta es valorar el cambio terapéutico y no patrones sintomáticos. El cuestionario lo componen 45 ítems en el que el paciente debe valorar con qué frecuencia los han experimentado en los últimos siete días, la afirmación que contiene cada ítem. La respuesta se ve acotada por una escala de cinco puntos en orden creciente de frecuencia, bajo las etiquetas: nunca, casi nunca, a veces, con frecuencia y casi siempre. Se hace explícita también la valoración numérica en cada opción de respuesta (escala Likert 0-4) pero dependiendo del significado de la afirmación (negativo o positivo en sentido de vivencia psicológica) esta seguirá un orden concordante con la frecuencia señalada o será inverso a esta.

El cuestionario genera un índice general de resultado que tienen en cuenta la valoración numérica de todos los ítems: **OQ-45 Total**. Y de forma más específica se delimitan tres sub-escalas que explicitan tres áreas de resultados diferenciadas: **OQ-45 Sintomatología** (25 ítems), **OQ-45 Interpersonal** (11 ítems) y **OQ-45 Rol Social** (nueve ítems). En todas las escalas una mayor puntuación se corresponde con un peor estado auto percibido del paciente en esa dimensión, y una menor puntuación con un mejor estado.

En este trabajo se ha aplicado una traducción local al español del cuestionario original americano (adjunto en Anexo 2). Ésta se realizó en un estudio preliminar para la adaptación del instrumento al ámbito nacional (Iraurgi, Hormaechea, Jiménez-Lerma y Landabaso 2009). Sin embargo, la validación completa del instrumento en nuestro país aún está en proceso (el estudio citado contó con una muestra clínica muy reducida). Por ello, y dado que son necesarios datos normativos para la utilización de la metodología ICF<sub>JT</sub>, se tomó la determinación de utilizar los datos generados en la validación del cuestionario en diferentes países: Estados Unidos (Lambert et al. 1996; Lambert et al. 2004), Holanda (de Jong et al, 2007), Chile (de la Parra y von Bergen,

2006) y Nueva Zelanda (Fitzgerald, 2013). En la Tabla 28 se recogen las medias, desviaciones estándar y coeficientes de fiabilidad de las diferentes escalas en cada estudio (tanto de la muestras clínicas como no clínicas). A partir de estos datos, se han calculado unos datos normativos ponderados, y se ha generalizado los coeficientes de fiabilidad mediante una estrategia de cálculo meta-analítica (Sánchez-Meca, López-Pina y López, 2009). En la Tabla 29 se recogen los valores finales que se han tenido en cuenta para este estudio.

**Tabla 28**

*Datos de normativos de OQ-45 en diferentes estudios de validación*

		EE.UU. (Lambert et al., 2004)		Chile (de la Parra et al., 2006)		Holanda (de Jong et al, 2007)		Nueva Zelanda (Fitzgerald, 2013)
		General (n= 815)	Clínica (n= 342)	General (n= 129)	Clínica (n= 124)	General (n= 807)	Clínica (n= 1920)	Clínica (n= 245)
<b>OQ-45</b>	M (DE)	45,2 (19,0)	83,1 (22,0)	48,7 (19,3)	100,4 (21,7)	38,7 (16,0)	79,5 (25,0)	77,8 (21,1)
<b>Total</b>	$\alpha$	0,93*	0,93*	0,91	0,91	0,91	$\alpha = 0,93$	No informa.
<b>OQ-45</b>	M (DE)	25,4 (12,0)	49,4 (15,0)	26,0 (12,4)	61,9 (14,1)	22,2 (10,0)	48,9 (16,0)	47,0 (13,5)
<b>Síntomas</b>	$\alpha$	0,92*	0,91*	0,88	0,89	0,89	0,91	No informa.
<b>OQ-45</b>	M (DE)	10,2 (6,0)	19,7 (6,0)	12,0 (0,71)	20,8 (5,9)	8,4 (5,0)	16,8 (7,0)	17,8 (6,3)
<b>Interpersonal</b>	$\alpha$	0,74*	0,74*	0,71	0,67	0,77	0,80	No informa.
<b>OQ-45</b>	M (DE)	9,6 (4,0)	14,1 (5,0)	10,7 (0,65)	17,7 (5,4)	8,1 (3,0)	13,6 (6,0)	13,3(5,1)
<b>Rol Social</b>	$\alpha$	0,70*	0,71*	0,65	0,63	0,53	0,69	No informa.

Notas: \* Valores alpha del estudio de Lambert et al. (1996) de los mismos autores (Estudiantes n=157; Clínica n= 298). M= Media; DE = Desviación Estándar, n= Tamaño muestral y  $\alpha$  = Coeficiente alpha (Consistencia Interna).

**Tabla 29***Datos de normativos ponderados de OQ-45 utilizados en el estudio 2*

<b>Escala</b>	<b>Población General (n=1751) M (DE)</b>	<b>Población Clínica (n=2631) M (DE)</b>	<b>Coefficiente <math>\alpha</math> Generalizado</b>
OQ-45 Total	42,46 (17,70)	80,79 (24,14)	0,924
OQ-45 Síntomas	23,97 (11,16)	49,40 (15,57)	0,905
OQ-45 Interpersonal	9,50 (5,50)	17,46 (6,77)	0,779
OQ-45 Rol Social	8,99 (3,6)	13,83 (5,77)	0,658

Notas: M= Media; DE = Desviación Estándar, n= tamaño muestral;  $\alpha$  = Coeficiente alpha (Consistencia Interna).

### **9.3.6. Procedimiento**

Se respetó el procedimiento habitual de asistencia del centro (ver variable independiente). Únicamente se introdujo el cuestionario OQ-45 como parte de la evaluación diagnóstica y de seguimiento. Se solicitó su cumplimentación por parte del paciente, en su domicilio, antes del inicio de la segunda sesión (a modo de línea base), y de nuevo al de cuatro meses, en el mismo centro. Se generaron de esta forma dos momentos de medición fuera de las sesiones dedicadas a la intervención. El periodo de tiempo evaluado en todos los casos ha sido de cuatro meses (16 semanas), lo que implicó una media de 10 sesiones de atención en el centro. Los casos de la muestra fueron captados durante el primer semestre del año, dando continuidad al tratamiento más allá de los cuatro meses de duración del estudio si fuese necesario.

Antes de iniciar el tratamiento (a partir de la tercera sesión) se solicitó a todos los participantes el consentimiento informado con respecto al uso de sus datos para fines de investigación. Con el fin de mantener la confidencialidad de los pacientes, se codificaron en la

base de datos de forma anónima, generando únicamente una clave solo decodificable por los terapeutas del centro.

### **9.3.7. Análisis de datos**

En el análisis de datos se ha utilizado la aplicación software SPSS de IBM, en su versión 21.0. Para los cálculos y estadísticos no reflejados en las opciones estándar del programa, se ha recurrido a su configuración manual a través de la ventana de sintaxis.

Inicialmente se realizó un análisis descriptivo y exploratorio de las variables dependientes tanto en el momento post tratamiento como en el pre tratamiento. Se estudió también la distribución de las diferencias entre ambos momentos en cada variable, comprobando su ajuste a la ley normal mediante la prueba *Kolmogorov-Smirnov*. No se realizó ninguna prueba de homocedasticidad ya que el estudio solo implica a un sólo grupo

Para la metodología basada en significación estadística de promedios grupales se ha recurrido a la prueba paramétrica *t de Student* para muestras relacionadas, con el fin de confirmar si las diferencias de las medias entre el momento pre tratamiento y el post tratamiento en cada variable dependiente son significativas. Esta prueba se ha realizado con todas las variables, a pesar que en dos de ellas se observa el incumplimiento del supuesto de normalidad. Se ha tomado esta determinación debido a que el contraste paramétrico se considera más potente que la alternativa no paramétrica, y en muestras grandes (como la del estudio,  $n=100$ ) se ve poco afectado por el incumplimiento de la normalidad (Newbold, Carlson y Thorne, 2008). En todos los contrastes anteriores se asumió como umbral de significación  $p < 0,05$  que se corresponde con un nivel de confianza de 95%. Con este mismo nivel de confianza se calculó el Intervalo de Confianza de la diferencia de medias en cada variable.

Posteriormente se calculó el tamaño del efecto del cambio mediante de *d de Cohen* para cada variable dependiente. Se han asumido las indicaciones para situaciones con medidas intrasujetos pre y post intervención donde no exista grupo control. Para lo cual se utilizó como denominador de la diferencia entre las medias de los dos momentos la desviación típica de la medida pre tratamiento (Morris, 2008).

Para la metodología de  $ICF_{JT}$  (Jacobson y Truax 1991) se usaron los datos normativos ponderados, explicitados anteriormente (ver instrumento). En primer lugar, se calculó el umbral de significación clínica de la medida post tratamiento para cada escala, bajo el criterio “C”, ya que se asumió solapamiento entre la distribución funcional y disfuncional en las variables analizadas. Lo que implicó el uso de los estadísticos descriptivos de dos muestras de referencia (clínica y general). En un segundo lugar, se calculó el propio  $ICF_{JT}$  para cada paciente y en cada variable dependiente, teniendo en cuenta la fiabilidad informada en la validación del instrumento y la desviación estándar de la muestra clínica de referencia. Se consideró como cambio estadísticamente significativo aquellos valores de  $ICF_{JT}$  que en términos absolutos igualarán o superasen 1,96 (valor z de un nivel de confianza de 95% en la curva normal). Finalmente, atendiendo a los dos criterios generados se clasificó el cambio terapéutico acontecido. Para mayor detalle de este procedimiento el lector puede acudir la última sección del capítulo 4.

## 9.4. Resultados

### 9.4.1. Análisis mediante contrastes de significación clásicos de promedios

La tabla 30 recoge los principales datos descriptivos en lo referente a las variables dependientes en los dos momentos de medida, así como el análisis de la normalidad de la distribución de sus diferencias. En todas las variable se observa una reducción de las puntuaciones promedio en las medidas post-tratamiento con respecto a las pre tratamiento. Al contrario que en el estudio 1, las puntuaciones de cada escala no están ponderadas por el número de ítems, por lo que no tiene sentido comparar las medias entre dimensiones, ni su dispersión de puntuaciones ya que se ven directamente condicionadas por el rango posible de estas. Pero llama la atención la escasa diferencia entre los dos momentos en la variable **OQ-45 Interpersonal**, reduciendo en apenas una unidad la puntuación (sin ser la variable de menor rango de puntuación es la que menor diferencia presenta). Las pruebas de normalidad de las distribuciones de diferencias han resultado significativas en el caso de **OQ-45 Total** y **OQ-45 Síntomas**, por lo que en estas dos variables se complica la hipótesis nula de no diferencia con distribución normal, aunque por las razones explicitadas en el apartado anterior se decide seguir utilizando estadísticos de contraste paramétricos.

**Tabla 30***Estadísticos descriptivos de la muestra y pruebas de normalidad de distribución en estudio 2*

Variable resultado	Pre-test Min. / Max.	Pre-test M (DE)	Post- test Min. / Max.	Post- test M (DE)	Normalidad K.S. (p)
<i>OQ-45 Total</i>	22 / 125	77,52 (28,47)	39/89	65,41 (11,16)	0,095 (0,027)
<i>OQ-45 Síntomas</i>	11 / 81	43,78 (19,95)	20/52	36,37 (7,43)	0,109 (0,005)
<i>OQ-45 Inter Personal</i>	4 / 33	17,54 (6,95)	8/24	16,38 (3,23)	0,078 (0,141)
<i>OQ-45 Rol Social</i>	6 / 28	16,20 (5,17)	7/19	12,66 (2,93)	0,085 (0,073)

Notas: Min.= Puntuación Mínima; Max.= Puntuación Máxima; M = Media; DE = Desviación Estándar; K.S.= Kolmogorov-Smirnov (Prueba de normalidad de distribución diferencias); p = valor de significación del estadístico.

La Tabla 31 recoge los valores del estadístico de contraste realizado, junto a su significación estadística. En la misma tabla se recoge la diferencia de medias en cada variable junto su Intervalo de Confianza del 95%. Así mismo se presentan los valores de magnitud del efecto. En todas las variables resultado se obtuvieron diferencias significativas entre las medias pre-test y post-test, al realizar la prueba *t de Student*. Siendo el sentido de todas estas diferencias significativas hacia el disminución sintomática. Respecto a la magnitud del efecto de las diferencias, y siguiendo los criterios sugeridos por Cohen (1988) se observa que la variable **OQ-45 Rol Social** posee el mayor tamaño efecto entre las variables analizadas, pero no llega a superar el umbral de efecto importante ( $d \geq 0,80$ ) valorándose como efecto medio-alto. En las escalas **OQ-45 Total** y **OQ-45 Síntoma** la magnitud de la diferencia también se interpreta como moderada aunque su valor sea algo inferior al punto considerado como medio ( $d = 0,50$ ). Y en el caso de la variable **OQ-45 Interpersonal** su efecto es interpretado como muy pequeño o de poca relevancia ( $d \leq 0,20$ ).

**Tabla 31**

Resultados de contraste de significación de diferencias y magnitudes del efecto en estudio 2

Variable resultado	Pre-Post (IC 95%)	Estadístico: <i>t</i>	<i>p.</i>	Magnitud del efecto: <i>d</i>
<i>OQ-45 Total</i>	12,111 (8,31 ; 15,91)	6,323	0,000	0,43
<i>OQ-45 Síntomas</i>	7,413 (4,64 ; 10,18)	5,309	0,000	0,37
<i>OQ-45 Interpersonal</i>	1,160 (0,17 ; 2,15)	2,329	0,022	0,17
<i>OQ-45 Rol Social</i>	3,537 (2,71 ; 4,35)	8,652	0,000	0,68

Notas: Pre-Post= diferencia de medias entre los dos momentos; IC= Intervalo de confianza de la diferencia;  $t = t$  de Student (Prueba diferencias de medias para muestra relacionadas); *p* = valor de significación del estadístico; *d*= *d* de Cohen.

#### 9.4.2. Análisis de cambio individual mediante Índice de Cambio Fiable

A diferencia del estudio 1, se ha decidido presentar los resultados del análisis  $ICF_{JT}$  de forma grafica para mayor claridad expositiva. De todas formas, en el Anexo 3 se adjuntan las tablas de resultados con los valores numéricos individuales necesarios para la clasificación del cambio según el doble criterio  $ICF_{JT}$  y Punto corte funcional. En este aparatado se dedica un gráfico para cada variable de resultado evaluada (Figuras de 4 a 7). Para su interpretación se tiene que tener en cuenta que se representa la diferencia entre ambos momentos de medida ubicando el valor de la puntuación post tratamiento en el eje de ordenas y la del pre tratamiento en el eje de abscisas. La franja diagonal con la etiqueta “*No cambio*” representa el área delimitada por el producto Error Estándar de la Diferencia y 1,96 (valor *z* de un nivel de confianza de 95%), lo que identifica la magnitud de diferencia que debe superar el cambio de cada sujeto para ser considerado como estadísticamente fiable y no achacable a la variabilidad aleatoria o a la precisión del instrumento. De esta forma, se divide el grafico en dos sectores: uno por encima de la franja de *No Cambio*, donde se ubican los casos que han empeorado de

forma fiable, y otra por debajo o de la franja, donde se establecería el área de mejoría fiable. La línea discontinua horizontal identifica el punto de corte entre el nivel funcional y el nivel disfuncional del momento post tratamiento, de esta forma se añade el criterio para valorar si el cambio además de fiable es clínicamente significativo (la zona funcional quedaría por debajo de esta línea). También se ubica este mismo punto de corte en el eje de pre tratamiento, para así señalar qué casos se encontraban ya en el área funcional antes del tratamiento y realizar unas conclusiones más ajustadas a la realidad.

En la variable **OQ-45 Total** (Figura 4) se obtuvo un punto de corte de funcionalidad de 58,68 y el Error Estándar de la Diferencia en el valor 9,66. Ya que el tamaño muestral es  $n=100$  se va a obviar informar de porcentajes de clasificación ya que resultaría redundante. Se constata que 47 casos se sitúan fuera de la franja de “*No cambio*”, por lo que se establece que experimentan un cambio estadísticamente fiable. De los anteriores, 41 se ubican en el área de mejoría aunque ninguno alcanza el punto de corte de funcionalidad, clasificándose como “*Mejoría dentro de la disfuncionalidad*”. Los otros seis casos de cambio fiable, se encuentran en el área de empeoramiento (aunque en casi todos limítrofes al área de no cambio), situándose cinco de ellos por debajo del punto de corte de funcionalidad pese a su empeoramiento, clasificándose como de “*Empeoramiento dentro de la funcionalidad*”, y un único caso que sobrepasa el punto de corte de la funcionalidad en el momento post tratamiento clasificándose como “*Empeoramiento dentro de la disfuncionalidad*”. Por otro lado, 53 casos se sitúan dentro de la franja de “*No cambio*” por lo que no se puede considerar la diferencia entre los dos momentos de medida estadísticamente fiable. De estos, 23 se ubican por debajo del punto de corte de la funcionalidad en el post tratamiento, por lo que son de “*Sin cambio dentro de la funcionalidad*”, y 30 casos encima de ese punto de corte, clasificándose como “*Sin cambio dentro de la disfuncionalidad*”. Finalmente, es necesario matizar que en 30 casos se observa una puntuación pre tratamiento inferior al punto de corte calculado, resultando que 24 de ellos sigue permaneciendo en el área de funcionalidad sin evidenciar cambio fiable en el momento pre tratamiento.

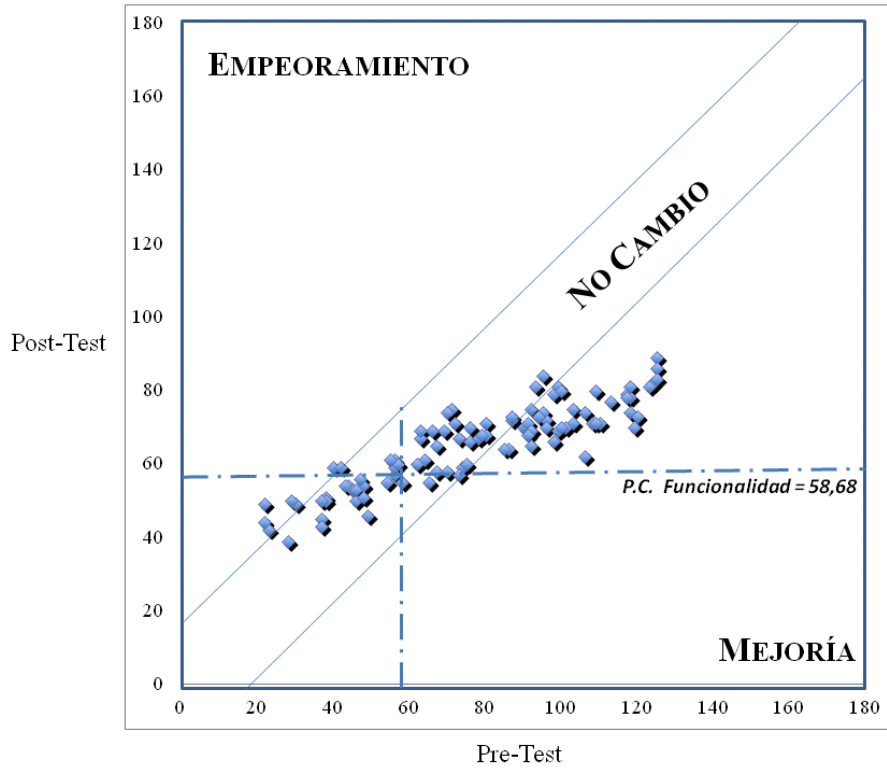
En la variable **OQ-45 Síntomas** (Figura 5) se obtuvo un punto de corte de funcionalidad de 34,59 y el Error Estándar de la Diferencia en el valor 6,61. En 34 casos se observa un cambio

estadísticamente fiable, distribuyendo su clasificación de la siguiente forma: 32 casos de “*Mejoría dentro de la disfuncionalidad*”, un caso de “*Recuperado*” y otro único caso de “*Empeoramiento dentro de la funcionalidad*”. Los otros 66 casos no evidencian cambio fiable, resultando: 35 casos como “*Sin cambio dentro de la funcionalidad*” y 31 casos como “*Sin cambio dentro de la disfuncionalidad*”. En 38 casos se observa una puntuación pre tratamiento en área de funcionalidad, de los cuales 34 permanecen en ella y sin cambio fiable en el momento post tratamiento.

En la variable **OQ-45 Interpersonal** (Figura 6) se obtuvo un punto de corte de funcionalidad en 13.07 y el Error Estándar de la Diferencia en el valor de 4,49. En nueve casos se observa un cambio estadísticamente fiable, distribuyendo su clasificación de la siguiente forma: ocho casos de “*Mejoría dentro de la disfuncionalidad*” y un único caso de “*Empeoramiento dentro de la disfuncionalidad*”. Los otros 91 casos no evidencian cambio fiable, resultando: 20 casos como “*Sin cambio dentro de la funcionalidad*” y 71 casos como “*Sin cambio dentro de la disfuncionalidad*”. En 34 casos se observa una puntuación pre tratamiento en área de funcionalidad, de los cuales 19 permanecen en ella y sin cambio fiable en el momento post tratamiento.

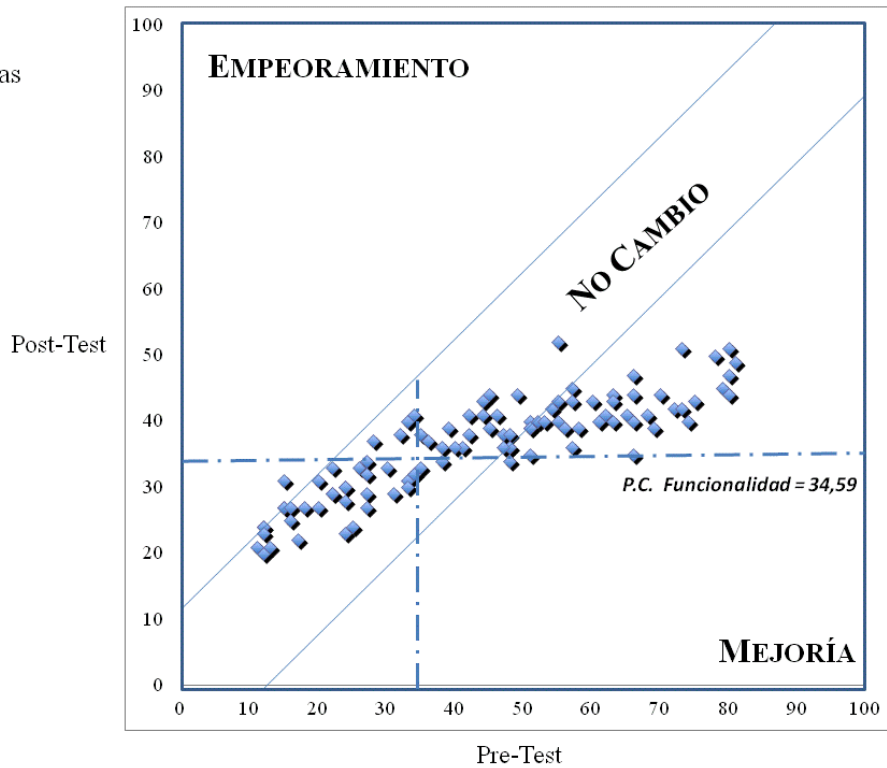
En la variable **OQ-45 Rol Social** (Figura 7) se obtuvo un punto de corte de funcionalidad de 10.85 y el Error Estándar de la Diferencia en el valor 4,76, generando el ancho de área de “*No cambio*” (relativo a su rango de puntuaciones) más amplio de las 4 variables. En ocho casos se observa un cambio estadísticamente fiable, distribuyendo su clasificación de la siguiente forma: siete casos de “*Mejoría dentro de la disfuncionalidad*” y un único caso de “*Recuperado*”. Los otros 92 casos no evidencian cambio fiable, resultando: 26 casos como “*Sin cambio dentro de la funcionalidad*” y 66 casos como “*Sin cambio dentro de la disfuncionalidad*”. En 12 casos se observa una puntuación pre tratamiento en área de funcionalidad, de los cuales siete permanecen en ella y sin cambio fiable en el momento post tratamiento.

OQ-45  
Total

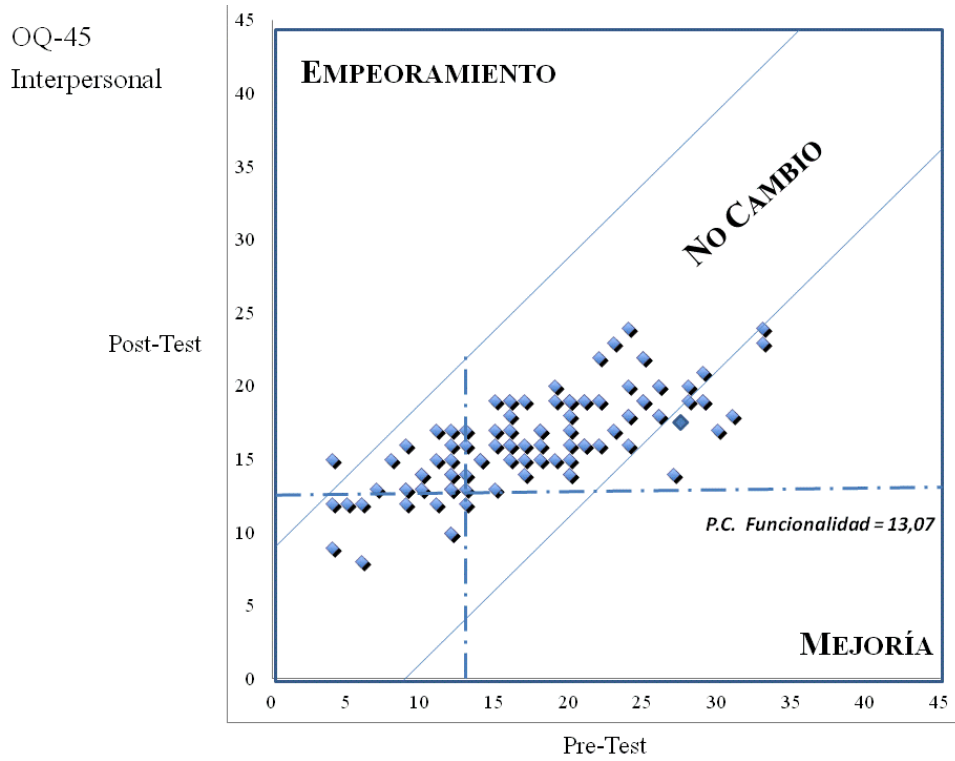


**Figura 4.** Resultados  $ICF_{IT}$  en variable OQ-45 Total en el estudio 2

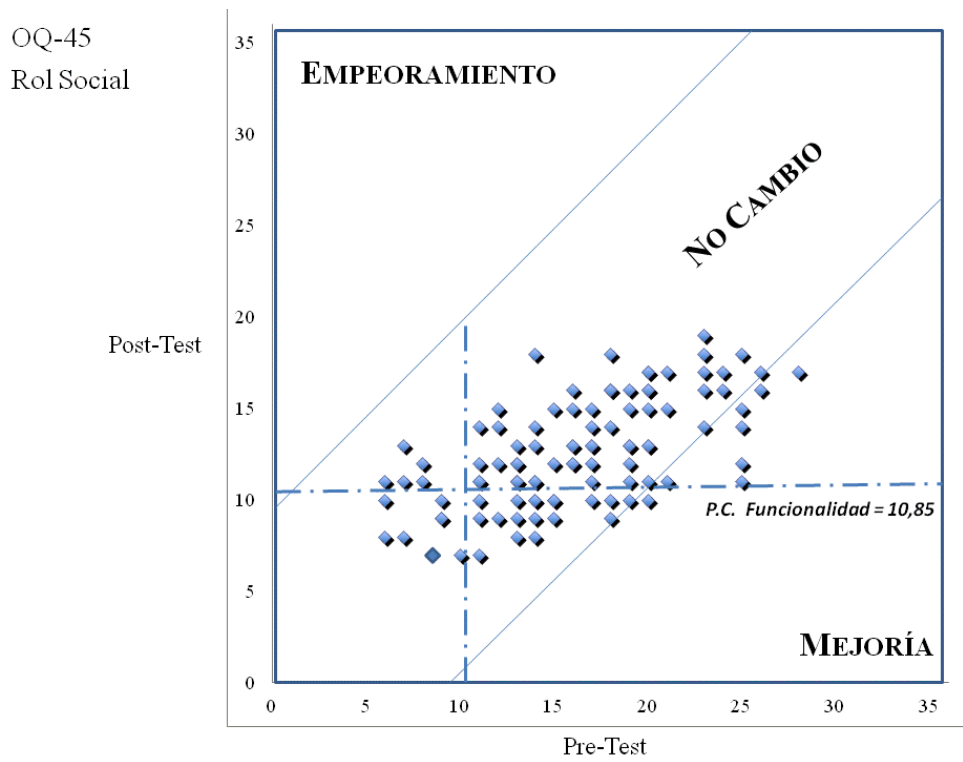
OQ-45  
Síntomas



**Figura 5.** Resultados  $ICF_{IT}$  en variable OQ-45 Síntomas en el estudio 2



**Figura 6.** Resultados  $ICF_{JT}$  en variable OQ-45 Interpersonal en el estudio 2



**Figura 7.** Resultados  $ICF_{JT}$  en variable OQ-45 Rol Social en el estudio 2

En relación a la evolución global de cada individuo y teniendo en cuenta todos índices de resultado, se observan varios datos relevantes. En 49 pacientes se registra una situación de no cambio fiable en todas las variables analizadas. En un único caso se observa un cambio fiable en todos los indicadores, y lo realiza en dirección a la funcionalidad (aunque no llegando a reducir su puntuación por debajo del punto de corte de funcionalidad en ninguno de ellos). En 34 pacientes se evidencia un cambio fiable del tipo de “*Mejoría dentro de la disfuncionalidad*” en dos o más indicadores. Respecto a lo anterior es destacable que de estos 29 casos hacen esta clasificación de forma simultánea en OQ-45 Total y OQ-45 Síntomas, y que en un único caso además de presentar dos indicadores mejorados figura otro más como “*Recuperado*”. Otros 8 casos presentan en un sola variable la clasificación de “*Mejoría dentro de la disfuncionalidad*”, resultando ser en seis pacientes la variable OQ-45 Total. Un único caso presenta en un solo indicador (OQ-45 Síntomas) la clasificación de *Recuperado* aunque en el resto de sus variables no se registren cambios fiables. Finalmente, en seis pacientes se observan cambios fiables en sentido de Empeoramiento en una o dos variables permaneciendo las demás sin cambio fiable.

## 9.5. Discusión y conclusiones

El método de análisis de promedios grupales señala como significativas las diferencias de medias en todas las variables de resultado. Sin embargo, el indicador de la magnitud del efecto refleja que unos tamaños del efecto moderados-bajos siendo inferior a 0,50 (*d* de Cohen) en todas las variables menos en **OQ-45 Rol Social**. Según este análisis, se concluye que ha existido una reducción estadísticamente significativa en todos los indicadores (en el sentido de la mejoría), resultando la variable **OQ-45 Rol Social** la que ha evidenciado un cambio más favorable, en contraste con la variable **OQ-45 Interpersonal** que ha sido en la que menor magnitud de cambio ha mostrado. Por otro lado, el método de análisis de cambio individual, desde unos criterios más restrictivos, plantean unas conclusiones diferentes para los resultados colectivos, discrepando especialmente en la consideración del cambio en **OQ-45 Rol Social**.

En las variables **OQ-45 Total** y **OQ-45 Síntomas** se observa cierta congruencia inicial entre el análisis de promedios y la metodología de casos ICF<sub>JT</sub>. La proporción de casos que realizan un cambio fiable hacia la funcionalidad en **OQ-45 Total** (41%) coincide con un tamaño

del efecto grupal medio-bajo ( $d = 0,43$ ). Lo mismo sucede con **OQ-45 Síntomas** aunque con unos valores algo más reducidos (33% de casos y  $d = 0,37$ ). Atendiendo a estos criterios se podría establecer que la variable **OQ-45 Total** ha mostrado mejores resultados que **OQ-45 Síntomas**, aunque siendo en ambos indicadores de una relevancia moderada.

Sin embargo, si se analizan los demás datos reportados por el método ICF<sub>JT</sub>, estas conclusiones no son del todo correctas. En **OQ-45 Total** se observan 6 casos de emparamiento fiable (aumento de puntuación en dirección disfuncionalidad) y ningún caso de cambio hacia la funcionalidad consigue llegar al punto de corte (“*Recuperado*”). Huelga decir que la mayoría de estos casos no llega a ubicarse en la zona de disfuncionalidad aunque siguen señalando una tendencia significativa de evolución desfavorable. Mientras que en **OQ-45 Síntomas** se registra un único caso de empeoramiento (aunque permanece en funcionalidad en el post tratamiento), y uno de los pacientes que cambia de forma fiable hacia la funcionalidad se clasifica como “*Recuperado*”. De esta forma se relativiza la mayor proporción de casos de cambio fiable de **OQ-45 Total**. De hecho, a nivel de relevancia clínica, esos casos de empeoramiento (especialmente el caso que pasa de puntuación funcional a disfuncional) tienen un peso específico que no se puede ignorar. De la misma forma que el caso que se clasifica como “*Recuperado*” en **OQ-45 Síntomas**. Por otro lado, si se observa las puntuaciones pre y post tratamiento en cada caso, se observa que en ambas variables, un tercio de los pacientes presenta una puntuación mayor en el momento post tratamiento. Este hecho se diluye en la consideración de promedios ya que se compensan los casos que empeoran con los que mejoran. El método ICF<sub>JT</sub>, si tiene en cuenta esta circunstancia clasificando estos casos como de no cambio fiable o empeorados según si el cambio se puede achacar a la variabilidad generada por la medida o no. Por todo ello, respecto a la metodología de contraste de promedios, las conclusiones de ICF<sub>JT</sub> para la evolución de estas variables resulta más negativa en **OQ-45 Total** y más favorable en **OQ-45 Síntomas**.

A pesar de lo anterior se evidencia una gran congruencia en la clasificación del cambio de **OQ-45 Total** y **OQ-45 Síntomas**, según el ICF<sub>JT</sub>. De hecho, el 48 % de los pacientes que realiza algún cambio fiable hacia la funcionalidad, lo realiza simultáneamente en estas variables. Circunstancia que no es de extrañar, al ser la sub-escala **OQ-45 Síntomas** la que más ítems aporta (25 de 45) al computo total de **OQ-45 Total**.

En la variable **OQ-45 Interpersonal** sí se puede apreciar mayor grado de acuerdo entre ambos métodos con respecto a las conclusiones colectivas, pero siendo necesario matizar una serie de observaciones. Aunque con el análisis de promedios grupales se haya una diferencia significativa entre los dos momentos de medida, el resto de datos que nos proporciona esta metodología señalan que esta diferencia es demasiado pequeña. En el intervalo de confianza de la diferencia de medias se constata que el extremo inferior (0,17) está muy próximo a 0, y que su valor de significación (0,022) es el más cercano al umbral de 0,05 de todas las variables. Con unos valores tan ajustados, en una muestra un poco más pequeña esta diferencia no hubiese resultado significativa. Además la magnitud del efecto señala un valor (0,17) inferior al considerado bajo ( $d = 0,20$ ). Por lo que se concluye que se trata de una diferencia muy poco relevante, y que a duras penas se puede rechazar la hipótesis nula de no diferencia. Con el método ICF<sub>JT</sub> se pueden sacar unas conclusiones similares al señalar que un 91% de la muestra no evidencia cambio fiable, y la mayoría de los pacientes se encuentra por encima del punto de corte de la funcionalidad en el momento post tratamiento. Además de los ocho casos que mejoran hacia la funcionalidad ninguno superan el punto de corte. Por ello, el cambio colectivo observado desde el método ICF<sub>JT</sub> también se definiría como poco relevante a nivel clínico y estadístico. El matiz importante que añade el método ICF<sub>JT</sub> es detectar el único caso que realiza un cambio fiable clínicamente significativo, que pasa de una puntuación previa por debajo del punto de corte a una post tratamiento que lo supera, empeorando de forma importante (aunque al ser un único caso no modifica en exceso las conclusiones colectivas).

Esta coincidencia entre ambos métodos en **OQ-45 Interpersonal** se puede explicar por la escasa magnitud de las diferencias (entre los dos momentos) en la mayoría de los casos. Sin embargo, hay un factor que ha incidido en la clasificación del método ICF<sub>JT</sub> que el método de análisis de promedios no tiene en cuenta: la fiabilidad de la escala. Si se observa la Figura 6, la franja de “*No cambio*” resulta más ancha que en las dos **anteriores** variables. Esto se debe a que el Error Estándar de las Diferencias, es relativamente mayor que las anteriores (relativo al rango posible de puntuaciones de la escala), siendo el valor del coeficiente de fiabilidad utilizado el culpable de que suceda esto. Mientras que el de las dos primeras escalas es mayor de 0,90, el coeficiente resultante de la generalización meta analítica de la escala **OQ-45 Interpersonal** es menor de 0,80. Aunque para esta variable, esta circunstancia no ha creado discrepancia entre lo

dos métodos, en la siguiente sí que ha generado una interpretación de resultados radicalmente diferente.

Según el análisis de promedios, **OQ-45 Rol Social** resulta ser la dimensión de resultados que mayor mejoría relativa registra, ya que su magnitud de efecto ( $d=0,68$ ) es la más alta. Aunque no alcanza el umbral de efecto importante ( $d=0,80$ ) sí supera el de efecto medio ( $d=0,50$ ), algo que no consiguen los demás indicadores. Por otro lado, el método ICF<sub>JT</sub> muestra una situación similar a la variable **OQ-45 Interpersonal**, clasificando la evolución de la mayoría de los pacientes como de cambio no fiable (92% de la muestra), discrepando totalmente con las conclusiones del análisis de promedios. A pesar de ello, intuitivamente observando las Figuras 6 y 7 se constata que el cambio acontecido en ambas variables es diferente. Si trazásemos una diagonal entre ambos ejes partiendo del 0, observaríamos la línea sobre la cual se sitúan los casos que puntúan exactamente lo mismo entre ambos momentos. La disposición de los casos de **OQ-45 Rol Social** se encuentra mucho menos agrupada alrededor de esta línea que **OQ-45 Interpersonal**. Y además se constata una clara inclinación de la nube de puntos hacia el cuadrante de mejora, tendencia que no tan marcada en **OQ-45 Interpersonal**. Observando con mayor detalle, se evidencia también que en **OQ-45 Rol Social** hay un número mucho mayor de casos que parten de un puntuación disfuncional y reducen su puntuación a un nivel funcional (área rectangular inferior derecha, generada al cruzar el punto de corte de la funcionalidad en los dos momentos de medida), lo que podría sugerir una relevancia clínica más coherente con el tamaño del efecto promedio. Pero debido a la gran amplitud relativa de la franja de “*No cambio*”, pocos casos (ocho pacientes) de los que realizan un cambio hacia la funcionalidad pueden no achacarlo a la variabilidad generada por el instrumento, y solo uno de estos consigue reducir su puntuación por debajo del punto de corte de la funcionalidad.

En la variable **OQ-45 Rol Social**, el bajo coeficiente de fiabilidad de la escala ( $\alpha = 0,658$ ) repercute de forma dramática en lo exigente que resulta el criterio ICF<sub>JT</sub>, al aumentar Error Estándar de las Diferencias. Se podría argumentar que en esta circunstancia la metodología resulta en exceso restrictiva, ya que no detecta los casos que podrían tener una relevancia clínica por la evidente magnitud del cambio. Sin embargo, el valor del coeficiente generalizado de nuestro meta análisis es inferior al umbral de 0,70, nivel a partir del cual la consistencia interna

de una escala se puede considerar aceptable, pudiendo cuestionar de hecho si con esta escala se esté evaluando un solo constructo (Campo-Arias y Oviedo, 2008; George y Mallery, 2013). Por lo que, no tener en cuenta este hecho sería añadir una fuente de error importante a las conclusiones, como de hecho sucede en el análisis de promedios.

Lo que sí se ha evidenciado como una limitación real del método ICF<sub>JT</sub> es la necesidad de que el instrumento de medida utilizado tenga una fiabilidad adecuada para poder ser sensible al cambio. De esta circunstancia se hacen eco Lambert, Hasen y Bauer (2008) señalando además que mayoritariamente los instrumentos de resultados que han mostrado alta fiabilidad y validez han sido los que registran sintomatología o funcionamiento global, (lo que se observa dentro del mismo OQ-45) existiendo una importante falta de medidas y valores normativos para otras áreas de resultados específicas. Por lo que, para cumplir con la reivindicación de Kazdin (2001) de ir más allá de lo meramente psicopatológico en la significación clínica, es necesario mayor investigación y desarrollo de medidas fiables.

Atendiendo al nivel de análisis del cambio individual de cada paciente, se puede resaltar las mismas ventajas del ICF<sub>JT</sub> subrayadas en el estudio 1: utilidad inmediata para el clínico (al observar la evolución de cada sujeto al margen de la tendencia grupal), y uso potencial de la independencia de análisis de cada caso (siendo posible la acumulación de casos para análisis futuros desde distintos parámetros). Un ejemplo dentro de la muestra de estudio de la primera ventaja es un paciente diagnosticado dentro del grupo de Trastornos Adaptativos (número 82 en tablas de resultados de Anexos). Se trata del único caso que experimenta un cambio fiable hacia la disfuncionalidad en la variable **OQ-45 Interpersonal** (ver Figura 6). Este empeoramiento es clínicamente trascendente, no solo por tener una gran magnitud (11 puntos en una rango de 45) sino porque implica la transición del estado del paciente de un nivel de gravedad no clínica a una puntuación considerada disfuncional por la comparación normativa. Resultando ser el único caso en esta variable que efectúa un cambio fiable de estas características en cualquiera de las dos direcciones. Su discrepancia con la tendencia general no termina aquí, ya que tanto **OQ-45 Síntomas** como en **OQ-45 Rol Social**, no efectúa un cambio fiable (permaneciendo la primera en área de funcionalidad en ambos momentos). La particular evolución de este caso no se tiene en cuenta en la consideración grupal. Con respecto a la posibilidad de acumular casos, se pueden

destacar algunos usos potenciales para la evolución de resultados futura del centro, tales como: realizar comparaciones estratificadas según grupo diagnósticos, terapeuta concreto o tipo de intervención.

Por último, de nuevo se comprueba que la metodología de análisis de promedios se ve mucho más afectada por la no existencia de grupo control que la de ICF<sub>JT</sub>. Al no controlar otras fuentes de variación de la puntuación que no sean el error de muestreo, resultan significativas las diferencias de medias de todas las variables, incluso la escasa diferencia en la variable **OQ-45 Interpersonal**. Y también se hace evidente el sesgo que se comete en la variable **OQ-45 Rol Social**, al destacarse en su tamaño del efecto respecto a las otras dimensiones aunque se trate de un medida de resultados menos confiable que las demás. El ICF<sub>JT</sub>, en esta situación, sigue sin poder asegurar relaciones de causalidad ( porque eso solo se consigue con otro tipo de diseño experimental) pero se muestra más precisa que la otra metodología a la hora de evaluar el cambio, no solo teniendo en cuenta la variabilidad del instrumento de medida sino también realizando una comparación normativa.

Se puede concluir que en el presente estudio (en este contexto de evaluación de resultados natural y sin grupo control) se ha vuelto ha hacer evidente que la metodología de ICF<sub>JT</sub> se muestra de mayor utilidad y permite conclusiones más rigurosas que la metodología clásica de contrastes de significación de promedios. Además la presentación grafica de los resultados del ICF<sub>JT</sub> ha permitido de forma sencilla e intuitiva sacar conclusiones colectivas sin verse diluidas las discrepancias individuales. Sin embargo, también se ha observado lo vulnerable que es la capacidad discriminante ICF<sub>JT</sub> a coeficientes de fiabilidad bajos como se ha podido ver en **OQ-45 Rol Social**.

## Capítulo 10

### **Estudio 3. Evaluación de resultados en una intervención sobre trastornos depresivos en el ámbito público. Comparación de diferentes estrategias de análisis de cambio individual.**

---

#### **10.1. Objetivos**

El propósito de este estudio es contrastar la propuesta metodológica de Jacobson y Truax (1991) para la evaluación de resultados psicoterapéuticos (Índice de Cambio Fiable) con otras metodologías de análisis de resultados en el marco de una intervención en trastornos depresivos en condiciones naturales.

Derivado de este objetivo general se pueden destacar dos objetivos específicos: Primero, comparar la metodología Índice de Cambio Fiable (ICF<sub>JT</sub>, a partir de ahora) con la metodología de contrastes de significación de promedios grupales y cálculo de tamaño del efecto. Segundo, contrastar el método ICF<sub>JT</sub> con otras propuestas metodológicas de análisis de cambio individual (GLN, EN, HA, TEI, DIT, ESG, DTW, IPP, IPI y CPM 50%).

#### **10.2. Hipótesis**

La metodología de ICF<sub>JT</sub> se mostrará superior a *otras alternativas de análisis de resultados*, tanto desde el punto de la utilidad clínica como de ajuste a la realidad, en un contexto de atención psicoterapéutica en personas diagnosticadas de un trastorno depresivo. Dos son las hipótesis a con:

- La metodología de ICF<sub>JT</sub> reportará más información útil, tanto para la determinación de la *efectividad de la intervención* del tratamiento de la depresión como para la toma de

decisiones del profesional, que la metodología de evaluación basada *en estadísticos de significación e indicadores del tamaño del efecto*.

- El método de ICF<sub>IT</sub> se mostrará más adecuado para la valoración del cambio de intensidad sintomática individual que las *propuestas alternativas de estimación de significación clínica* (GLN, EN, HA, TEI, DIT, ESG, DTW, IPP, IPI y CPM 50%).

### **10.3. Metodología**

#### **10.3.1 Participantes**

Se han considerado dos grupos diferenciados de participantes. El primero lo configura una muestra clínica de personas que acudieron con demanda de asistencia terapéutica a un Centro de Salud Mental de la Red de Salud Pública de la Comunidad Autónoma Vasca. Este grupo quedó formado por 100 personas (69% mujeres), de una edad comprendidas entre los 20 y 62 años (Media= 38,84; DE= 10,96). Su diagnóstico se ajustó a una de a una de las siguientes secciones del Eje 1 del DSM-IV TR (APA, 2002): Trastornos del estado del Ánimo (68% de la muestra) y Trastornos Adaptativos (32% de la muestra), siendo la sintomatología depresiva el signo psicopatológico predominante. Se consideraron como criterios de exclusión: diagnósticos de enfermedad mental grave (psicosis, deterioro cognitivo..), dificultades significativas en la lecto-escritura, el requerimiento de tratamiento farmacológico con neurolépticos o eutimizantes (litio) y la ausencia del consentimiento informado.

Para la configuración del segundo grupo, una vez conocidas las características de la muestra clínica, se buscaron controles sanos voluntarios que fueron apareados por sexo y edad con los pacientes clínicos. De esta manera se trataba de controlar la influencia de estas variables sociodemográficas, y generar un grupo de comparación donde no fuese esperable un cambio debido a su condición clínica y tratamiento entre los dos momentos de medida (necesario para alguna de las alternativas metodológicas de análisis de cambio individual). Por tanto, este grupo control quedó configurado por 100 participantes con la misma edad y género que la muestra de pacientes, y de cercana ubicación geográfica. Se consideraron como criterios de exclusión: necesitar o haber necesitado tratamiento psicológico o psiquiátrico en los últimos 10 años, dificultades significativas en la lecto-escritura y la falta de consentimiento informado.

### 10.3.2. Diseño

Se trata de un diseño cuasi-experimental de grupos no equivalentes, con medidas pre post facto, en el cual se han utilizado estrategias de equiparación de grupos mediante el apareado de casos por sexo y edad. En la Tabla 32 se ofrece una representación esquemática del diseño (Cammbell y Stanley, 1996; Alaminos y Castejon, 2006).

**Tabla 32**

*Representación gráfica del diseño del estudio 3*

Grupos	Medidas pre-programa	Programa experimental	Medidas post-programa
<i>Intervención</i>	O <sub>1</sub>	X	O <sub>2</sub>
<i>Control</i>	O <sub>1</sub>	-	O <sub>2</sub>

La equiparación de los grupos por sexo y edad se ha realizado por ser dos variables epidemiológicas que han mostrado su influencia sobre la expresión de la afectividad. De este modo, se bloquea el efecto de ambas variables mediante un procedimiento basado en el diseño y no en el control estadístico a posteriori. Por otro lado, se ha considerado un grupo control no equivalente (no clínico) para poder realizar todas las alternativas metodológicas de análisis de resultados individuales planificadas. En el análisis clásico (comparación de grupos), su utilidad principal es tener en cuenta el cambio espontaneo por el simple paso del tiempo en la medida resultado.

### 10.3.3. Variable dependiente

Como variable dependiente se ha tenido en cuenta la dimensión sintomática depresión. En este estudio se entiende la depresión como un conjunto de signos y síntomas relacionados que se

presentan de forma convergente (síndrome). También se ha asumido la hipótesis de continuidad que entiende la depresión como un fenómeno dimensional en el cual un individuo puede ser situado en un punto determinado del continuo de frecuencia/intensidad del síndrome, incluso sin llegar a cumplir criterios de diagnóstico clínico categorial. De esta forma su consideración clínica es una cuestión cuantitativa más que cualitativa (Beck 1976; Sakashita, Slade y Andrews 2007; Vallejo, 2011).

La depresión se caracteriza por presentar un estado de gran abatimiento emocional con predominio de sentimientos de tristeza, decaimiento, irritabilidad y malestar generalizado. Coherentemente con este estado se presentan diferentes manifestaciones sintomáticas, en forma de alteraciones de tipo: conductual, cognitivo, anímico y físico (Vázquez y Sanz, 2009). Un desarrollo conceptual más completo acerca de la depresión se puede encontrar en el segundo apartado del capítulo 6 de este trabajo.

#### **10.3.4. Variable Independiente**

Se establece como variable independiente el tratamiento psicoterapéutico realizado por un psicólogo clínico en un Centro de Salud Mental de la Red Pública en la provincia de Vizcaya. El terapeuta es un varón, de mediana edad, nativo de dicho territorio, con más de 20 años de experiencia clínica. La intervención se ajustó al procedimiento cognitivo-conductual estandarizado por Beck en su manual de tratamiento de la depresión (Beck, Rush, Shaw y Emery, 1983), aunque adaptándolo a las condiciones de asistencia del ámbito público. Se organiza el tratamiento en sesiones de 40-45 minutos, con una periodicidad de sesión cada dos semanas, exceptuando las dos primeras (dedicadas a entrevista inicial, evaluación y diagnóstico) separadas por una única semana. El estudio tuvo una duración de seis meses si bien se dio continuidad a la intervención psicoterapéutica en todos los casos que o precisaron.

#### **10.3.5. Instrumento**

Para medir la intensidad de la dimensión sintomática depresión se ha recurrido al Beck's Depression Inventory BDI (Beck, Rush, Shae y Emery, 1979), en su validación Española de Vázquez y Sanz (1997). Se trata de un instrumento de auto informe clínico configurado por 21 ítems. Cada uno de estos está formado por varias opciones de respuesta, en forma de

afirmaciones en relación a un mismo signo o síntoma depresivo. El sujeto debe elegir la afirmación de cada ítem que mejor se acerque a su estado durante la última semana. Los 21 síntomas representados en los ítems son los siguientes: tristeza, pesimismo, sentimientos de fracaso, insatisfacción, culpabilidad, sentimiento de castigo, auto decepción, auto culpabilización, ideas de suicidio, llanto, irritación, desinterés social, indecisión, fatiga, pérdida de apetito, disminución del peso, hipocondría y desinterés por sexo. En la corrección del cuestionario cada ítem se valora de 0 a 3 según la opción elegida (puntuándose 3 la afirmación que representa mayor intensidad sintomática), de esta forma se genera una puntuación total en el rango de 0 a 63 que cuantifica la intensidad depresiva general del sujeto. Aunque los criterios de episodio depresivo descritos en el DSM-IV y V (APA, 2001; 2013) están casi totalmente representados en el cuestionario (a excepción de los síntomas psicomotores), no se considera un instrumento diagnóstico y su uso debe limitarse a medir intensidad sintomática (Sanz y Vázquez, 1998).

Según la revisión de Sanz, Perdígón y Vázquez (2003) el BDI ha sido el instrumento más utilizado como medida sintomatológica depresiva tanto en el ámbito clínico como en el de la investigación. Aunque existe una segunda versión del cuestionario BDI-II (Sanz y Vázquez, 2011), su versión comercial con validación española estuvo disponible con posterioridad al inicio de la recogida de datos. Respecto a la versión del BDI utilizada, existen dos estudios de datos normativos en población española de los mismos autores (Vázquez y Sanz 1997; Sanz y Vázquez, 1998). Se ha decidido tomar los valores normativos del primero de ellos, por ser su muestra más heterogénea (especialmente con respecto a la variable edad). Resultando una puntuación media normativa de 7,5 con una desviación estándar de 6,7. Con respecto a la fiabilidad, en ambos estudios se obtuvo un coeficiente alfa de 0,83, siendo este valor el asumido, en el presente estudio, para los procedimientos de análisis de cambio individual que la requieran.

#### **10.3.6. Procedimiento**

En una primera fase se recogieron los datos de la muestra clínica entre los pacientes que habían acudido al Centro de Salud. A las personas cuya problemática psicológica principal fuese la sintomatología depresiva se les solicitó la participación en el estudio y el consentimiento informado para el uso de sus datos. Descartando los casos que implicasen alguno de los criterios

de exclusión explicitados anteriormente. Se siguió el procedimiento habitual del centro con respecto a la organización de las citas y atención del terapeuta. Se solicitó a cada paciente la cumplimentación del BDI, en su domicilio, antes del inicio de la segunda sesión, y de nuevo al de 6 meses en el mismo centro antes de comenzar otra sesión. Se generaron de esta forma dos momentos de medición fuera de las sesiones dedicadas a la intervención. El periodo de tiempo comprendido entre ambas medidas, en todos los casos, fue de 6 meses (24 semanas), lo que implicó una media de 14 sesiones de atención en el centro. Se continuo reclutando pacientes hasta contar con una muestra de 100 casos con los dos momentos de medida registrados. Con el fin de mantener la confidencialidad de los pacientes, se codificaron en la base de datos de forma anónima, generando únicamente una clave solo descodificadle por el terapeuta.

En la segunda fase se buscó aparear cada paciente de la muestra clínica con una persona del mismo ámbito geográfico, edad y sexo pero sin afección psicopatológica diagnosticada. Esto se realizó solicitando voluntarios en el rango de edad de la muestra clínica. Un miembro del equipo establecía contacto con el voluntario/a participante bien presencialmente o de forma telefónica. Durante este primer contacto daba explicaciones de la naturaleza del estudio y del objetivo de su colaboración solicitando consentimiento verbal de participación. En caso de acceder a la participación se comprobaban el cumplimiento de criterios de inclusión y exclusión y de ser satisfechos se les ofrecía un cuestionario en papel si la entrevista se realizaba de forma presencial o la posibilidad de responder al instrumento de una forma telemática, en el caso de que el contacto hubiera sido de forma telefónica. A este respecto, para facilitar la captación de los datos se desarrolló una aplicación web temporal, donde se reprodujo el BDI con las mismas instrucciones y redacción de ítems que el original. La aplicación ofrecía una explicación de los objetivos del estudio, se solicitaba el consentimiento de participación y se recogían datos sociodemográficos y filtros para garantizar los criterios de inclusión y exclusión. Un vez satisfechos éstos, el colaborador podía pasar a responder las 21 situaciones que propone el BDI.

Las características del estudio implicaba volver a contactar con los voluntarios/as y solicitarles una segunda respuesta al BDI seis meses después para emular el periodo entre el pre tratamiento y el post tratamiento en la muestra clínica. Para tal fin se les solicitó una forma de contacto (teléfono o e-mail) y se les designó una clave de identificación con la cual se

reconocerían los registros de datos de la línea base y seguimiento a los seis meses de cada participante. Únicamente el investigador responsable de la captación tenía acceso al listado de identificación nominal de participantes con sus direcciones y las claves asignadas, y mantenía custodia de este registro cumpliendo con los requisitos de guarda y confidencialidad que marcan los criterios de aplicación ética y legal de una investigación.

### **10.3.7. Análisis de datos**

Para el análisis de datos se ha utilizado la aplicación SPSS de IBM, en su versión 21.0. Para los cálculos y estadísticos no implementados en las opciones estándar del SPSS, se ha procedido a su programación mediante la sintaxis del propio programa (Ver Anexo 1).

En ambos grupos se ha realizado un análisis descriptivo y exploratorio de la variable dependiente en los dos grupos del estudio. Así mismo debido a la planificación del uso de determinadas pruebas paramétricas se ha comprobado el ajuste de la variable resultado al supuesto de distribución normal mediante la prueba *Kolmogorov-Smirnov*, y también al supuesto de homocedasticidad entre grupos mediante la prueba de *Levene*.

Para la metodología basada en significación estadística de promedios grupales se ha realizado un *modelo mixto de ANOVA* que implicó tener en cuenta las *medidas repetidas* de los dos grupos del estudio de forma simultánea. Debido a que de partida se sabe que existe diferencia entre los dos grupos en el nivel de depresión (uno es clínico y el otro es no clínico), el efecto de interés será intra-sujetos, concretamente la interacción de pertenecer a un grupo u a otro con el cambio producido entre los dos momentos de medida. En todos los contrastes anteriores se asumió como umbral de significación  $p < 0,05$  que se corresponde con un nivel de confianza de 95%. También se calcularon los estimadores del tamaño de los efectos reportados por el ANOVA realizado en forma de eta cuadrado ( $\eta^2$ ).

Para la metodología de ICF<sub>JT</sub> (Jacobson y Truax, 1991) se usaron los datos normativos de la validación del instrumento de medida explicitados anteriormente. Sin embargo, respecto a la muestra de referencia clínica se ha tomado la decisión de utilizar los valores de los estadísticos descriptivos del grupo clínico del presente estudio. Aunque existe una validación nacional del

instrumento en pacientes con trastornos psicológicos (Vázquez y Sanz, 1999), se trata de un muestra muy heterogénea, con presencia de problemáticas que no implican sintomatología depresiva. De hecho, el subgrupo de pacientes con diagnóstico de trastorno depresivo es más pequeño (n=48) que el grupo de nuestro estudio (n=100).

Con las dos muestras de referencia se calculó el umbral de significación clínica de la medida post tratamiento para cada escala, bajo el criterio “C”, ya que se asumió solapamiento entre la distribución funcional y disfuncional en la variable analizada. Ello implicó el uso de los estadísticos descriptivos de dos muestras de referencia (clínica y no clínica). En un segundo lugar, se calculó para el total de participantes el ICF<sub>JT</sub> del BDI teniendo en cuenta la fiabilidad informada en la validación del instrumento y la desviación estándar de la muestra clínica de referencia. Se consideró como cambio estadísticamente significativo aquellos valores de ICF<sub>JT</sub> que en términos absolutos igualarán o superasen 1,96 (valor z de un nivel de confianza de 95% en la curva normal). Finalmente, atendiendo a los dos criterios generados se clasificó el cambio terapéutico acontecido. Para mayor detalle de este procedimiento el lector puede acudir la última sección del capítulo 5.

Para la selección de métodos alternativos a ICF<sub>JT</sub> (análisis de cambio individual) se ha atendido a las revisiones de Lambert, Hansen y Bauer (2008), y de Pardo y Ferrer (2013) acerca de las principales propuestas de criterio interno para el ámbito clínico. Además se ha optado por utilizar únicamente las últimas versiones de los índices de cada autor, tal y como se plantea en otros estudios de comparación de este tipo de métodos (McGlinchey, Atkins y Jacobson, 2002; Bauer, Lambert y Nielsen, 2004). Debido a que sólo se cuenta con dos momentos de medida en la variable resultado, se han descartado métodos que requieren de múltiples medidas, tales como el uso de curvas de crecimiento en Modelo Lineal Jerárquico HLM (Speer y Grenbaum, 1995). Teniendo en cuenta todo lo anterior, se han utilizado los siguientes métodos: Gullisken-Lord-Novick -GLN- (Hsu, 1995), Edward-Nunnally EN (Speer, 1992), Hageman-Arrindel HA (Hageman y Arrindel, 1999a; 1999b), Tamaño del Efecto Individual TEI (Kazis, Anderson y Meenan, 1989), Diferencia Individual Tipificada DIT (Stucki, Liang Fossel y Katz, 1995), Estadístico de Sensibilidad de Guyatt ESG (Guyatt, Bombardier y Tugwell, 1986), Intervalo para Pronóstico Promedio IPI e Intervalo para Pronóstico Individual IPI (Crawfor y Howell, 1998;

Crawfor y Garthwaite, 2006). Adicionalmente se ha incluido un noveno método, no reflejado en las revisiones citadas, aunque popular en los estudios psiquiátricos de efectividad de tratamientos farmacológicos: el Criterio Porcentaje de Mejoría 50% CPM 50% (Hiller y Schindler, 2011; Hiller, Schindler y Lambert, 2012). Para facilitar la exposición, en adelante se hará referencia a cada método por su acrónimo.

En el capítulo 5 se explicita el planteamiento y formulación de cada método de análisis individual presente en este estudio, sin embargo es necesario matizar una serie de elementos que se han establecido para sus cálculos y criterios de clasificación. Para todos los índices que contemplan en su cálculo la desviación estándar y/o la media de una muestra de referencia o normativa, se ha utilizado los valores del estudio de BDI en población general de Vázquez y Sanz (1997). Lo mismo se aplica para las situaciones donde se requiera un coeficiente de fiabilidad (normalmente para el cálculo del Error de Estándar de Medida), donde se ha utilizado el coeficiente alfa del estudio referido (ver apartado Instrumento), a menos que el método designase específicamente otro cálculo o referente. Si el método contaba con diferentes umbrales de significación estadística, se fija en  $\pm 1,96$ , a menos que se explicita un único umbral específico o implique un intervalo de contraste con la puntuación post tratamiento. Para los índices que establecen sus umbrales según los criterios de tamaño del efecto de Cohen (1988), se ha clasificado tanto con el umbral mínimo (0,20) como con el criterio de  $\pm 1,96$ , tal y como lo plantea Pardo y Ferrer (2013). Para el establecimiento del punto de corte de funcionalidad requerido en el CPM 50% se ha utilizado el mismo que el calculado para el ICF<sub>JT</sub>. En aquellos estadísticos que necesiten de un grupo control donde no se espere cambio sistemático (ESG, IPP e IPI), se ha utilizado como tal la muestra no clínica captada para el estudio. La asunción de todos estos criterios, tiene como objeto la homogenización de los elementos comunes de los diferentes métodos para poder realizar una comparación más rigurosa de su funcionamiento.

Para generar la clasificación de los diferentes métodos se han designado tres categorías: Mejora, No cambio y Empeora. Esto se ha hecho así porque los métodos del estudio son alternativas al método interno de ICF<sub>JT</sub>, siendo compatibles con el criterio de significación clínica por punto de corte de funcionalidad. Estudios precedentes de comparación (McGlinchey, Atkins y Jacobson, 2002; Bauer, Lambert y Nielsen, 2004) han establecido la clasificación de

cambio teniendo en cuenta el criterio interno de cada método concreto pero combinado con la de significación clínica de punto de corte de Jacobson y Truax (1991), generando un mínimo de cuatro categorías (Recuperado, Mejorado en disfuncionalidad, Sin Cambio y Empeorado). En el presente estudio se entiende que la discrepancia entre métodos la crea el umbral de cambio estadísticamente significativo, pudiendo obviarse para la comparativa el criterio externo de punto de corte. La excepción a esto es el método HA, ya que también especifica un cálculo alternativo para el punto de corte con la funcionalidad, por lo que en ese caso en concreto se planifica un análisis adicional con cuatro categorías para el estudio de su concordancia con ICF<sub>JT</sub>.

Para la comparación estadística entre métodos, inicialmente, se ha calculado el porcentaje de acuerdo de clasificación de cada método con respecto a los demás utilizando *Tablas de Contingencia* para registrar la concordancia observada en la clasificación del cambio de todos los sujetos en conjunto (de ambos grupos). Posteriormente, se han calculado los *Coefficientes Kappa* entre pares de métodos para controlar la concordancia esperada por azar, aunque de forma separada para el grupo clínico y el grupo no clínico (ya que plantean dos situaciones de cambio muy diferentes). También se ha estudiado si las diferencias entre pares de clasificaciones puedan atribuirse al azar o no, utilizando para ello la prueba de los *Rangos con Signo de Wilcoxon*, y de nuevo de forma separada para el grupo clínico y el grupo no clínico. Para los dos estadísticos anteriores se asumió como umbral de significación  $p < 0,05$ .

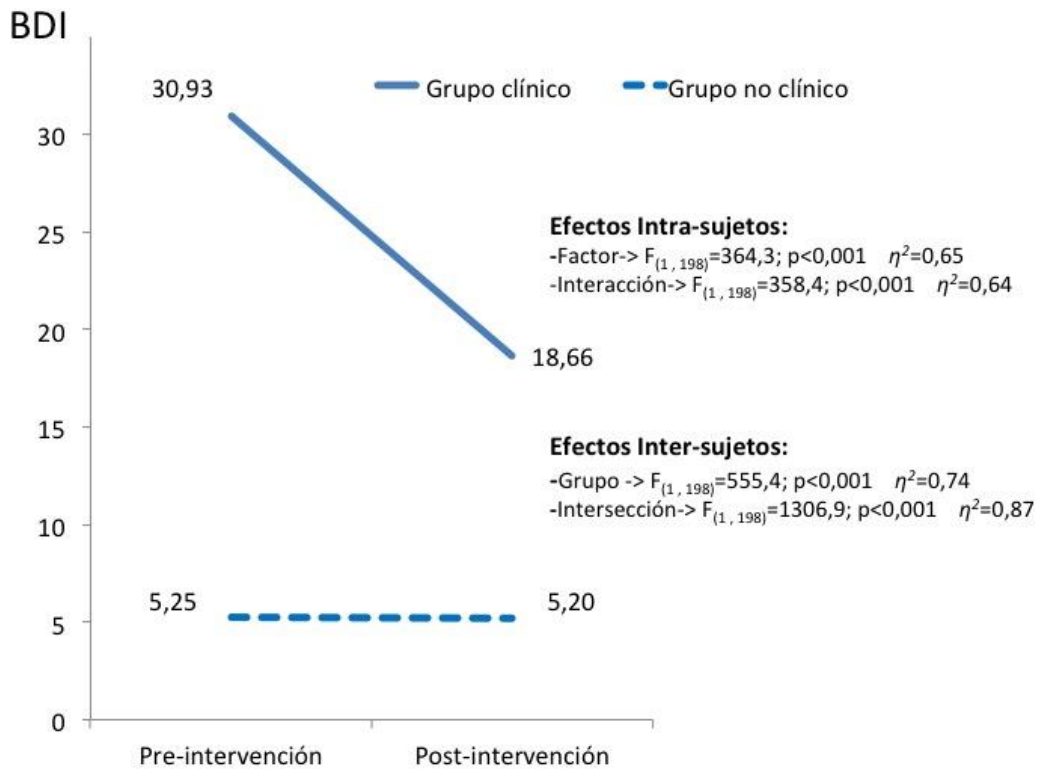
## **10.4. Resultados**

### **10.4.1. Análisis mediante contrastes de significación clásicos grupales**

El análisis descriptivo de las muestras ha reportado para el grupo clínico una puntuaciones medias en el BDI de 30,93 (DE= 8,11) en el pre tratamiento, y de 18,66 (DE= 7,42) en el post tratamiento. Mientras que el grupo no clínico presenta unas puntuaciones medias de 5,25 (DE= 4,15) en el inicio y de 5,20 (DE= 4,54) en el segundo momento de medida. Mostrando una situación de reducción sintomática en el grupo clínico, y prácticamente de no cambio en el grupo no clínico.

El estudio de la normalidad de la variable dependiente y el contraste de homocedasticidad entre grupos han resultado estadísticamente significativos ambos casos, complicando la asunción de los supuestos para el uso de la prueba paramétrica planificada. A pesar de ello, se ha decidido utilizar el contraste ANOVA por tener más potencia y ofrecer más información que su alternativa no paramétrica (donde, por ejemplo, sería imposible estudiar el efecto interacción). Se justifica su uso en el contexto de este estudio al poseer ambos grupos un tamaño grande ( $n=100$ ) y tener el mismo número de participantes, situación en la cual el análisis de la varianza se ve poco sesgado por el incumplimiento de los anteriores supuestos. Adicionalmente, se matiza que a pesar de tratarse de una situación de medidas repetidas, no es necesario generar una prueba de esfericidad ya que sólo existen dos niveles en el factor (Berkovits, Hancock y Nevitt, 2000; Newbold, Carlson y Thorne, 2008; Pardo, San Martín, 2010).

En la Figura 8 se presenta los resultados del ANOVA realizado, así como la representación gráfica del cambio acontecido en ambos grupos. Los contraste inter-sujetos han resultado significativos tanto para el grupo como para la intersección, señalando que los grupos parten de una diferencia alta en las puntuaciones basales y que ésta se mantiene en el post tratamiento ( $\eta^2= 0,87$  y  $\eta^2= 0,74$ ). Los contrastes intra-sujeto también han resultado significativos tanto para el factor (tiempo) como para la interacción de grupo con el factor. La diferencia entre el momento pre y post tratamiento se muestra muy grande ( $\eta^2= 0,65$ ), pero también el efecto de interacción de pertenecer a un grupo u otro en el cambio acontecido ( $\eta^2= 0,64$ ), lo que justifica estadísticamente lo que se puede observar gráficamente: el grupo clínico reduce su puntuación de manera muy relevante mientras que el grupo no clínico se mantiene estable. Todos los tamaños del efecto obtenidos se consideran grandes según el criterio de  $\eta^2 > 0,14$  (Cohen, 1988).

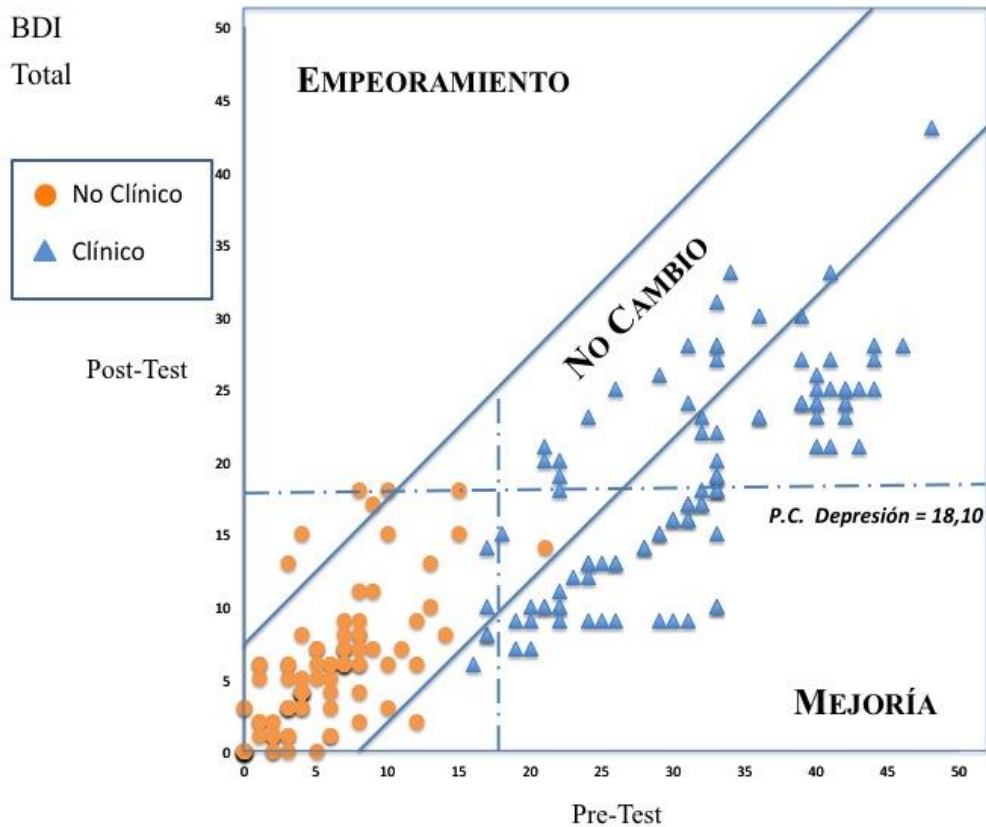


**Figura 8.** ANOVA mixto con medidas repetidas en grupos del estudio 3

#### 10.4.2 Análisis de cambio individual mediante metodología ICF<sub>JT</sub>.

Para mayor claridad expositiva se van a presentar en este apartado los resultados de ICF<sub>JT</sub>, de forma resumida en un gráfico. A pesar de ello, en el Anexo 3 se adjuntan todos los resultados individuales que se han generado para clasificar el cambio de cada participante. La Figura 9 representa las diferencias individuales entre los dos momentos de medida en la puntuación del BDI en ambos grupos de estudio, ubicando el valor de la puntuación post-tratamiento en el eje de ordenadas y el del pre-tratamiento en el eje de abscisas. La franja diagonal del gráfico genera tres áreas de clasificación según el criterio interno de cambio fiable. La línea discontinua horizontal separa el área disfuncional de la funcional respecto a la puntuación post-tratamiento según el criterio de punto de corte de significación clínica (aunque también se señala este mismo punto de

corte para la puntuación pre tratamiento para visualizar en que área de funcionalidad se encontraba el participante en el momento inicial).



**Figura 9.** Clasificación de cambio individual  $ICF_{IT}$  en puntuación BDI del estudio 3

El Error Estándar de la Diferencia ha resultado en un valor de 3,91, que al multiplicarlo por 1,96, genera un intervalo de  $\pm 7,66$  puntos. Si la diferencia entre la puntuación post tratamiento y la puntuación pre tratamiento de un participante ha resultado estar dentro de este intervalo, el cambio se ha considerado ‘no estadísticamente fiable’, por lo que se ha ubicado al participante en la franja de “No cambio” de la Figura 10.X. Si la diferencia supera los límites del intervalo, se ubicará en el área de “Empeoramiento” (puntuación post mayor que pre, lo que implica que aumenta la intensidad sintomática) o en área de “Mejoría” (puntuación pre mayor que post, es decir, disminuye la intensidad sintomática). Para el criterio de significación clínica, el punto de

corte ha resultado fijado en una puntuación de 18,1 (debajo del cual se ubica el área de funcionalidad). Este valor ha coincidido casualmente con la puntuación donde el cuestionario ha mostrado mayor exactitud diagnóstica (sensibilidad, especificidad y poder de predicción) en distintos estudios psicométricos, y en el cual existe gran consenso para designarlo como el umbral de la prueba más adecuado para diferenciar personas con una afección clínica de depresión de las que no (Kendall, Hollon, Beck, Hammen e Ingram, 1987; Rudd y Rajab, 1995; Sanz y Vázquez, 1998).

En el grupo clínico se observa que en 80 pacientes acontece un cambio estadísticamente fiable (al tratarse de grupos de 100 participantes se va obviar informar de porcentajes). De los cuales 46 también se encuentran por debajo del punto de corte de funcionalidad en el momento post tratamiento, clasificándose su cambio como *“Recuperado”*; y los otros 34 se quedan por encima del punto de corte designándose su cambio como *“Mejoría dentro de la disfuncionalidad”*. Por otro lado, de los 20 pacientes que no efectúan un cambio fiable, 15 se sitúan por encima del punto de corte, por lo que se considera su situación *“Sin cambio fiable dentro de la disfuncionalidad”*, mientras que los otros cinco participantes sí muestran una puntuación post tratamiento por debajo del punto de corte clasificándose como *“Sin cambio fiable dentro de la funcionalidad”*. Finalmente, es necesario matizar que de todos los participantes, cinco ya se encontraban por debajo del punto de corte de funcionalidad antes del tratamiento, de los cuales dos han efectuado un cambio de suficiente magnitud como para ser considerado fiable (en dirección a la mejoría), mientras que el cambio de los otros tres no ha podido ser considerado fiable.

En el grupo no clínico, se constata que solo seis participantes han mostrado un cambio de estadísticamente fiable. Y de éstos, cinco han aumentado su puntuación en el segundo momento de medida, aunque se mantienen por debajo del punto de corte de funcionalidad, por lo que su cambio se ha clasificado como *“Empeoramiento dentro de la funcionalidad”* (tres de ellos se sitúan justo en el límite con la disfuncionalidad). El otro participante que efectúa un cambio fiable lo hace en sentido de mejoría, disminuyendo su puntuación dentro del área de la funcionalidad. Los 94 participantes que no realizan un cambio fiable se sitúan por debajo del punto de corte de la funcionalidad clasificándose todos ellos como *“Sin cambio dentro de la*

*funcionalidad*”. Finalmente, es necesario señalar que un único participante de todo el grupo puntuaba por encima del punto de corte de funcionalidad en el momento inicial, y a pesar de situarse en el área de funcionalidad en la segunda medida, este cambio no ha resultado estadísticamente fiable.

### **10.4.3. Análisis y contraste de cambio individual entre diferentes metodologías**

Aunque en el presente apartado se presentan únicamente las clasificaciones globales de cada método de análisis, en el Anexo 3 se adjuntan los valores de cada participante que se han utilizado para generar dichas clasificaciones. Por otro lado, debido a que el método Hageman-Arrindel (HA) ha tenido un funcionamiento totalmente anómalo e indiscriminado para las muestras del estudio (señalando cambio en todos los casos incluso cuando la diferencia entre las dos puntuaciones fuese cero), no se ha incluido en las tablas de resultados del presente apartado. Matizar también acerca de HA que la fiabilidad de las diferencias en el grupo clínico ha resultado en un valor aceptable (0,49) mientras que en el grupo no clínico no ha alcanzado el nivel mínimo explicitado por los autores para su uso ( $r_{dd} > 0,40$ ). Los valores individuales de HA también están incluidos en el Anexo 3.

En la Tabla 33 se pueden observar los recuentos de clasificación para los dos grupos del estudio según los diferentes métodos empleados en su análisis. Se evidencia que los métodos ESG y IPP, resultan los menos restrictivos a la hora de detectar cambio (menor tasa de clasificados como no cambio que los demás métodos) en ambos grupos. Por el contrario, el método TEI con el umbral 1,96 se muestra como el más restrictivo a la hora clasificar cambio, tanto a mejor como a peor.

**Tabla 33**

Recuentos de clasificados de cambio individual según los diferentes métodos en el estudio 3

	Grupo Clínico (n=100)			Grupo No Clínico (n=100)		
	Mejora	No cambia	Empeora	Mejora	No cambia	Empeora
Índice Cambio Fiable Jacobson y Truax	80	20	0	1	94	5
Método Gullisken Lord-Novik	74	26	0	1	94	5
Método Edward-Nunnally	81	19	0	10	79	11
Tamaño del efecto individual $\geq 1,96$	25	75	0	1	96	3
Tamaño del efecto individual $\geq 0,20$	95	5	0	36	37	27
Diferencia individual Tipificada $\geq 1,96$	71	29	0	3	92	5
Diferencia individual Tipificada $\geq 0,20$	95	5	0	36	37	27
Estadístico de sensibilidad Guyatt $\geq 1,96$	82	18	0	3	92	5
Estadístico de sensibilidad Guyatt $\geq 0,20$	99	1	0	36	37	27
Diferencia Tipificada de Wyrwich	84	16	0	6	89	5
Intervalo para Pronostico Promedio	80	7	13	44	30	26
Intervalo para Pronostico Individual	67	33	0	1	94	5
Criterio Porcentaje de Mejoría 50%	79	21	0	27	73	0

La Tabla 34 recoge el acuerdo porcentual de concordancia observada entre métodos, teniendo las clasificaciones en ambos grupos al mismo tiempo. Las Tablas 35 y 36, informan del índice *Kappa* entre pares de métodos, y si la diferencia de clasificación resulta significativa (prueba *Wilcoxon*), aunque de forma separada para el grupo clínico y para el grupo no clínico. Para interpretar los *Índices Kappa* se ha recurrido a la guía de niveles de concordancia ofrecida por Altman (1991): < 0,20 pobre / 0,21-0,40 baja / 0,41-0,60 moderada / 0,61-0,80 buena / 0,81-1 muy buena.

**Tabla 34***Porcentajes de acuerdo entre métodos de clasificación cambio individual en conjunto del estudio 3*

	ICF <sub>JT</sub>	GLN	EN	TEI 1,96	TEI 0,20	DIT 1,96	DIT 0,20	ESG 1,96	ESG 0,20	DTW	IPP	IPI	CPM 50%
ICF <sub>JT</sub>	—	97%	91%	71%	64%	95%	64%	98%	62%	96%	59%	94%	83%
GLN		—	89%	74%	61%	96%	61%	95%	59%	93%	58%	97%	82%
EN			—	64%	72%	89%	72%	93%	70%	94%	68%	86%	85%
TEI_1,96				—	36%	75%	36%	70%	34%	67%	33%	78%	59%
TEI_0,20					—	61%	100%	66%	98%	69%	77%	58%	74%
DIT_1,96						—	61%	95%	59%	92%	58%	95%	82%
DIT_0,20							—	66%	98%	69%	77%	58%	74%
ESG_1,96								—	64%	98%	62%	92%	84%
ESG_0,20									—	67%	77%	56%	72%
DTW										—	63%	89%	85%
IPP											—	55%	72%
IPI												—	79%
CPM_50%													—

*Nota:* ICF<sub>JT</sub>= Índice Cambio Fiable Jacobson y Truax; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; TEI 1,96/0,20=Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para Pronostico Promedio; IPI=Intervalo para Pronostico Individual; CPM50%= Criterio Porcentaje de Mejoría 50%.

Respecto al método ICF<sub>JT</sub>, se evidencia un destacado porcentaje de acuerdo con GLN, EN, DTW, DIT con umbral 1,96, ESG con umbral 1,96, IPI y con CPM 50%. Al observar los estadísticos de concordancia por grupos, se constata un acuerdo total con GLN e IPI en la situación de no cambio mayoritario que representa el grupo no clínico, que disminuye en el contexto de cambio hacia mejoría que representa el grupo clínico (aunque sigue siendo muy buena), mostrándose entonces estos métodos sutilmente más exigentes que ICF<sub>JT</sub> para señalar cambio. Con el método EN no muestra diferencias significativas de clasificación y además se constata un alto grado de acuerdo para el grupo clínico, pero esta concordancia se reduce drásticamente en el grupo no clínico, mostrándose menos restrictivo que ICF<sub>JT</sub>. Un patrón de concordancia similar al anterior es el que se muestra con el método CPM 50 %, aunque en este caso la concordancia es casi inexistente para el grupo no clínico. Por otro lado, con los métodos

TEI, DIT y ESG, se observa una concordancia baja o pobre en el umbral de 0,20, en ambos grupos; resultando además estadísticamente significativas la diferencias de sus clasificaciones respecto a ICF<sub>JT</sub> en el grupo clínico. Sin embargo, estos mismos estadísticos resultan con una buena concordancia si se establece el umbral en 1,96; exceptuando el TEI para la muestra clínica una concordancia muy baja, resultando este último mucho más restrictivo para calificar cambio. Con DTW resultan diferencias significativas en clasificación para ambos grupos, sin embargo, su nivel de concordancia no baja de moderado y es muy alta en el grupo clínico. El método IPP también difiere significativamente en ambos grupos pero su nivel de concordancia es mucho menor que con DTW; moderado en grupo clínico y prácticamente inexistente en grupo no clínico.

En el método **GLN** se observa un perfil similar a ICF<sub>JT</sub> con respecto porcentajes de acuerdo con los otros métodos, aunque resultando con un nivel sutilmente más alto con los métodos que resultan más restrictivos que ICF<sub>JT</sub> (como TEI 1,96 e IPI), y algo menor con los que resultan un poco menos restrictivos que ICF<sub>JT</sub> (como EN, DTW). Los estadísticos de concordancia resaltan el acuerdo con el método IPI, con un nivel muy alto en ambos grupos; de hecho el GLN es el método que de todos mayor concordancia muestra con IPI. En relación a los demás métodos, sus niveles de concordancia siguen una tendencia parecida al ICF<sub>JT</sub> aunque con el mismo matiz enunciado para los porcentajes de acuerdo. Al contrario que el ICF<sub>JT</sub>, el GLN no muestra diferencias estadísticamente significativas (*Wilcoxon*) en clasificación con DIT con umbral 1,96, y sí las evidencia con respecto al método EN.

El método **EN** también se observa un patrón de porcentajes de acuerdo parecido a ICF<sub>JT</sub>, aunque al contrario a **GLN**, este se inclina más hacia los métodos menos restrictivos. Su nivel de acuerdo e índice de concordancia más alto se da con ESG con el umbral 1,96 y en el grupo clínico, no existiendo diferencias significativas en clasificación entre los dos en ambos grupos. Con DTW también mantiene un nivel de concordancia destacado: muy alto para el grupo clínico y moderado para el grupo no clínico, y tampoco muestra diferencias significativas en ninguno de los dos grupos. Por otro lado, TEI con umbral 1,96 y ESG con umbral 0,20 son los métodos con los que menor concordancia se puede observar.

**Tabla 35**

*Comparación por pares entre clasificaciones de cambio individual en grupo clínico del estudio 3*

	ICF <sub>JT</sub>	GLN	EN	TEI 1,96	TEI 0,20	DIT 1,96	DIT 0,20	ESG 1,96	ESG 0,20	DTW	IPP	IPI	CPM 50%
ICF <sub>JT</sub>	___	0,83	0,90	0,15	0,35	0,76	0,35	0,94	0,08	0,87	0,51	0,67	0,91
GLN	p<0,05	___	0,80	0,21	0,26	0,82	0,26	0,77	0,06 <sub>NS</sub>	0,70	0,51	0,83	0,86
EN	NS	p<0,05	___	0,15	0,37	0,73	0,37	0,97	0,08	0,90	0,59	0,65	0,94
TEI_1,96	p<0,001	p<0,001	p<0,001	___	0,03 <sub>NS</sub>	0,24	0,03 <sub>NS</sub>	0,14	0,01 <sub>NS</sub>	0,12	0,09	0,28	0,16
TEI_0,20	p<0,001	p<0,001	p<0,001	p<0,001	___	0,23	1	0,39	0,32	0,43	0,15	0,19	0,33
DIT_1,96	p<0,05	NS	p<0,05	p<0,001	p<0,001	___	0,23	0,70	0,05 <sub>NS</sub>	0,64	0,47	0,81	0,78
DIT_0,20	p<0,001	p<0,001	p<0,001	p<0,001	NS	p<0,001	___	0,39	0,32	0,43	0,15	0,19	0,33
ESG_1,96	NS	p<0,05	NS	p<0,001	p<0,001	p<0,05	p<0,001	___	0,09	0,93	0,55	0,62	0,91
ESG_0,20	p<0,001	p<0,001	p<0,001	p<0,001	p<0,05	p<0,001	p<0,05	p<0,001	___	0,10	0,04 <sub>NS</sub>	0,04 <sub>NS</sub>	0,07
DTW	p<0,05	p<0,05	NS	p<0,001	p<0,05	p<0,001	p<0,05	NS	p<0,001	___	0,50	0,56	0,84
IPP	p<0,05	NS	p<0,001	p<0,001	p<0,001	NS	p<0,001	p<0,001	p<0,001	p<0,001	___	0,41	0,60
IPI	p<0,001	p<0,05	p<0,001	p<0,001	p<0,001	NS	p<0,001	p<0,001	p<0,001	p<0,001	NS	___	0,70
CPM_50%	NS	p<0,05	NS	p<0,001	p<0,001	p<0,05	p<0,001	NS	p<0,001	p<0,05	p<0,05	p<0,05	___

*Nota:* Cuadrante inferior izquierdo presenta niveles de significación para comparación de pares de prueba **Wilcoxon**. NS= No significativo (p>0,05). Cuadrante superior izquierdo presenta **Coefficientes Kappa** para estudio de nivel concordancia entre métodos. Los coeficientes son estadísticamente significativos como mínimo al nivel de p<0,05, exceptuando los indicados con el sub índice <sub>NS</sub>. ICF<sub>JT</sub>= Índice Cambio Fiable Jacobson y Truax; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; TEI 1,96/0,20=Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para Pronostico Promedio; IPI=Intervalo para Pronostico Individual; CPM 50%= Criterio Porcentaje de Mejoría 50%.

**Tabla 36**

*Comparación por pares entre clasificaciones de cambio individual en grupo no clínico del estudio 3*

	ICF <sub>JT</sub>	GLN	EN	TEI 1,96	TEI 0,20	DIT 1,96	DIT 0,20	ESG 1,96	ESG 0,20	DTW	IPP	IPI	CPM 50%
ICF <sub>JT</sub>	___	1	0,40	0,79	0,10	0,85	0,10	0,85	0,10	0,69	0,09	1	0,01 <sub>NS</sub>
GLN	NS	___	0,40	0,79	0,10	0,85	0,10	0,85	0,10	0,69	0,09	1	0,01 <sub>NS</sub>
EN	NS	NS	___	0,28	0,35	0,51	0,35	0,51	0,35	0,65	0,29	0,40	0,29
TEI_1,96	NS	NS	NS	___	0,07	0,65	0,07	0,65	0,07	0,51	0,06	0,79	0,02 <sub>NS</sub>
TEI_0,20	NS	NS	NS	NS	___	0,13	1	0,13	1	0,18	0,59	0,10	0,43
DIT_1,96	NS	NS	NS	NS	NS	___	0,13	1	0,13	0,83	0,11	0,85	0,10 <sub>NS</sub>
DIT_0,20	NS	NS	NS	NS	NS	NS	___	0,13	1	0,18	0,59	0,10	0,43
ESG_1,96	NS	NS	NS	NS	NS	NS	NS	___	0,13	0,83	0,11	0,85	0,10 <sub>NS</sub>
ESG_0,20	NS	NS	NS	NS	NS	NS	NS	NS	___	0,18	0,59	0,10	0,43
DTW	p<0,05	p<0,05	NS	NS	NS	NS	NS	NS	NS	___	0,15	0,69	0,22
IPP	p<0,05	p<0,05	p<0,05	p<0,05	NS	p<0,05	NS	NS	NS	p<0,05	___	0,09	0,35
IPI	NS	NS	NS	NS	NS	NS	NS	NS	NS	p<0,05	p<0,05	___	0,01 <sub>NS</sub>
CPM_50%	p<0,001	p<0,001	p<0,001	p<0,001	p<0,05	p<0,001	p<0,05	p<0,001	p<0,05	p<0,001	NS	p<0,001	___

*Nota:* Cuadrante inferior izquierdo presenta niveles de significación para comparación de pares de prueba **Wilcoxon**. NS= No significativo (p>0,05). Cuadrante superior izquierdo presenta **Coefficientes Kappa** para estudio de nivel concordancia entre métodos. Los coeficientes son estadísticamente significativos como mínimo al nivel de p<0,05, exceptuando los indicados con el sub índice <sub>NS</sub>. ICF<sub>JT</sub>= Índice Cambio Fiable Jacobson y Truax; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; TEI 1,96/0,20=Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para Pronostico Promedio; IPI=Intervalo para Pronostico Individual; CPM 50%= Criterio Porcentaje de Mejoría 50%.

El método **TEI** muestra unos porcentajes de acuerdo con los demás métodos en su mayoría menores que los de los tres anteriores. Se destaca el acuerdo con el método IPI cuando se establece el umbral de TEI en 1,96, y con el DIT y ESG cuando los tres tienen el umbral de 0,20. En el umbral de 1,96 todos los niveles de concordancia (*Kappa*) en el grupo clínico son, en el mejor de los casos, bajos. En el grupo no clínico este hecho se atenúa bastante acercándose a buenos niveles de concordancia con ICF<sub>JT</sub>, GLN e IPI. Dado que el grupo no clínico se da una situación de poco cambio, se observa que la discrepancia con los otros métodos en el umbral 1,96 se debe fundamentalmente a ser más restrictivo que todos los demás a la hora de clasificar como mejorado o empeorado. En el umbral 0,20 sucede lo contrario que lo anterior: sus niveles concordancia son menores con la mayoría de los métodos en el grupo no clínico respecto al grupo clínico. Exceptuando DIT y el ESG en el umbral 0,20 con los que concuerda totalmente en el grupo no clínico (circunstancia que mantiene con el DIT 0,20 en el grupo clínico). De forma general, se observa un nivel de concordancia mayor en el umbral 0,20 con los métodos menos restrictivos: DIT 0,20; ESG 0,20; DTW y CPM 50%.

En el método **DIT** con el umbral 1,96 se observa un perfil muy similar al de porcentajes de acuerdo de GLN, con el cual, de hecho, se señala un 96% de acuerdo. En el umbral 0,20 sus concordancias resultan ser las mismas que las del TEI 0,20, existiendo un 100% de acuerdo entre ambos en las clasificaciones para los dos grupos. Se destaca un nivel de concordancia muy bueno con los métodos IPI y GLN cuando se usa el umbral 1,96, tanto para el grupo clínico como no clínico, lo que identifica un mayor grado de acuerdo (para ese umbral) con los métodos un poco más restrictivos que el ICF<sub>JT</sub>. En el umbral 0,20 resulta algo más restrictivo que el método ESG 0,20 para el grupo clínico y, aunque tiene acuerdo total con este mismo método en el grupo no clínico, no es destacable ya que matemáticamente son idénticos para esa situación concreta (ver explicación en el apartado de análisis de datos).

Con el método **ESG** se observa un gran parecido con perfil de porcentajes de acuerdo de ICF<sub>JT</sub> (en el umbral 1,96) y con el de DIT 0,20 (en el umbral 0,20). Sus niveles de concordancia señalan un comportamiento de clasificación en el umbral 1,96 muy similar a EN en el grupo clínico, y más cercano a GLN e ICF<sub>JT</sub> en el grupo no clínico, resultando un poco más restrictivo que EN teniendo en cuenta ambos grupos. Por otro lado, en el umbral 0,20 se muestra como el

método menos restrictivo de todos y niveles de concordancia muy bajos o nulos para las clasificaciones de ambos grupos, exceptuando con DIT 0,20 en el grupo no clínico por las razones explicitadas anteriormente.

En el método **DTW** se observa unos porcentajes de acuerdo destacados con ICF<sub>JT</sub>, GLN, EN, DIT 1,96 y ESG 1,96. Los niveles de concordancia revelan que el mayor nivel de acuerdo lo muestra con el método ESG 1,96 en la clasificación de ambos grupos. Se observa que con el método EN ha evidenciado también nivel de concordancia muy bueno en el grupo clínico que disminuye en el grupo no clínico. Siguiendo la comparativa con el anterior método, DTW se muestra un poco más permisivo para calificar cambio que EN para el grupo clínico, y más restrictivo en el grupo no clínico. Con respecto ICF<sub>JT</sub>, GLN, EN, DIT 1,96 y ESG 1,96 se muestra menos restrictivo en ambos grupos, lo que genera que para el grupo de poco cambio (no clínico) tenga unos niveles de acuerdo menores con estos métodos que en el grupo clínico.

El método **IPP** se muestra como el método con menores porcentajes de acuerdo evidenciados, en ningún caso superior al 80%. Muestra su mayor nivel de acuerdo general con los métodos menos restrictivos: TEI, ESG, DIT (los tres en el umbral 0,20) y con CPM 50%. Observando los niveles de concordancia *Kappa* se observa que no supera en ningún caso el nivel de moderado, y en el grupo no clínico únicamente muestra concordancia algo mayor con los citados métodos menos restrictivos. Analizando la clasificación y los resultados individuales se ha hecho evidente un funcionamiento anómalo del método. Al formar el intervalo de confianza pronóstico, en algunos casos, el valor real de la puntuación pre tratamiento no se encuentra incluido en el intervalo. Esto ha generado clasificaciones distorsionadas, como señalar cambios a peor cuando la persona o no había cambiado en absoluto o había cambiado a mejor por pocos puntos.

El método **IPI** presenta unos altos porcentajes de acuerdo con ICF<sub>JT</sub>, GLN, EN, DIT y ESG (los dos últimos únicamente cuando toman como umbral 1,96). Con respecto a los niveles de concordancia se destaca un nivel muy bueno con los métodos GLN y DIT 1,96 para la clasificación en ambos grupos, además con DIT 1,96 no muestra diferencias significativas en ninguno de los grupos. Con ICF<sub>JT</sub> y ESG 1,96 su concordancia es mucho mayor en el grupo de poco cambio que en el grupo clínico. Se muestra más restrictivo a la hora de clasificar cambio

que los métodos con los que existe una concordancia alta. Sin embargo, solo alcanza a un nivel de concordancia alta con TEI 1,96 (el método más restrictivo del estudio) en el grupo no clínico.

Y, finalmente, en el método **CPM 50%** se destaca un mayor porcentaje de acuerdo con EN, DTW, ESG 1,96 y ICF<sub>JT</sub>. Los niveles de concordancia con los demás métodos se ven drásticamente afectados por el grupo que está clasificando (clínico o no clínico). Esto se debe a que en el método CPM 50% no se establece un criterio de empeoramiento como en los demás métodos, y sólo se tiene en cuenta el cambio a mejor. Dado que en el grupo clínico el cambio solo tiene ese sentido, el acuerdo con los demás métodos es mucho mayor en el grupo clínico donde sí hay personas que aumentan notablemente su puntuación en segundo momento de medida. De esta forma en el grupo clínico se observa una clasificación muy cercana a ICF<sub>JT</sub> y EN ( $Kappa < 0,90$ ), presentando niveles muy semejantes de concordancia que los presentados por éstos. Se alinea con los métodos que resultan menos restrictivos que IPP y DIT 1,96 pero que no son tan permisivos como el umbral 0,20 en los métodos ESG, DIT y TEI. Aunque es con estos últimos con los únicos que llega a mostrar una concordancia moderada en la clasificación del grupo no clínico siendo, en los demás casos, o bajas o nulas.

## **10.5. Discusión y conclusiones**

El análisis de promedios grupales hace concluir que pertenecer a un grupo u otro es muy relevante para el cambio observado ( $\eta^2 = 0,64$ ), siendo este de una gran magnitud entre ambos momentos ( $\eta^2 = 0,65$ ). Dado que de partida los grupos no son equivalentes en gravedad sintomatológica, se deben relativizar las conclusiones de causalidad. Sin embargo, se ejerce más control que en los estudios anteriores al tener cuenta el cambio espontáneo por el paso del tiempo utilizando grupo con equivalencia en ciertas variables de interés (como sexo y edad). En cualquier caso, y observando también los estadísticos descriptivos, al igual que en los dos estudios precedentes, este método de análisis arroja unas estimaciones muy favorables sobre el efecto de la intervención en la medida resultado.

El ICF<sub>JT</sub> se muestra en coherencia con la apreciación general del análisis grupal, aunque matiza la relevancia del cambio y señala los casos discordantes con la tendencia grupal. El hecho de que un 80% de la muestra clínica haya realizado un cambio fiable en dirección a la mejoría, y que no exista empeoramiento en ningún paciente de este grupo, se muestran en congruencia con las positivas conclusiones del análisis de promedios. Sin embargo, se aprecia que en sólo 46 pacientes se puede calificar su cambio como clínicamente relevante (57% de los que realizan un cambio fiable). El método permite identificar a 34 pacientes (43% de los que cambian fiablemente) que, a pesar de realizar un cambio estadísticamente significativo, aún se encuentran en una situación de gravedad clínica en el momento post-tratamiento. Se vuelve a confirmar la observación de Jacobson y Truax (1991) de que el análisis individual suele ser menos favorable que el análisis grupal cuando se evalúa eficacia psicoterapéutica. Por otro lado, en el análisis individual del grupo no clínico, que en términos promedios no muestra prácticamente cambio alguno, sí se señalan cambios relevantes para algunos participantes. Sólo 36 participantes mantienen la misma puntuación entre ambos momentos de medida, aunque las variaciones de los que cambian son en su mayoría escasas y por debajo de lo que se ha considerado cambio fiable. Sin embargo, cinco personas presentan un empeoramiento significativo y cercano a traspasar el umbral hacia la disfuncionalidad, un evento relevante que pasa desapercibido en la consideración grupal y plantea un argumento adicional a favor de la diferencia entre grupos y el efecto beneficioso de la intervención: no solo los participantes no tratados no cambian sino que la mayoría de los pocos que cambian de forma fiable lo hacen a peor.

En la comparativa de métodos de análisis de cambio individual, se observa que los métodos EN y GLN, basados en el ajuste del planteamiento ICF<sub>JT</sub> para suplir el efecto de regresión a la media, se comportan de una manera muy similar al ICF<sub>JT</sub> para la clasificación en el grupo clínico (nivel de concordancia superior a 0,81 con ambos métodos). Es llamativo que el método EN, que pretende corregir el sesgo aludido, resulte notablemente más permisivo en el grupo no clínico que ICF<sub>JT</sub>. Estos resultados coinciden con estudios precedentes (Speer y Greenbaum, 1995; Bauer, Lambert y Nielsen 2004; Atkins, Bedics, Mc Glinchey y Beuchaine, 2005) donde los métodos computacionalmente más complejos que ICF<sub>JT</sub> no se muestran más sensibles en la calificación, evidenciándose un gran tasa acuerdo con estos (se repite la tendencia de que GLN sea ligeramente más restrictivo y EN algo menos restrictivo). Esto complica la asunción de que

el efecto de regresión a la media tenga el impacto que se le ha conferido para clasificar cambio fiable (Hsu, 1995; Lambert, Hansen y Bauer, 2008).

El método HA, planteado también como una alternativa más precisa que  $ICF_{JT}$ , se ha comportado en el presente estudio como un sistema de clasificación totalmente indiscriminado. Este hecho discrepa con estudios anteriores (Bauer, Lambert y Nielsen 2004; Atkins, Bedics, Mc Glinchey y Beuchaine, 2005; Ronk, Hooke y Page 2012) donde el método HA había resultado ser el más exigente, de todos los contrastados, a la hora de clasificar cambio. La situación del presente estudio plantea una baja dispersión de puntuaciones tanto en el grupo de estudio como en el de referencia, y una diferencia de medias en el grupo clínico muy grande, generando una calificación de cambio incluso en los casos de cambio nulo. Aunque sería necesario generar un estudio más pormenorizado para detectar la causa matemática de este funcionamiento anómalo, está no parece estar contemplada en el planteamiento original, ya que en el grupo clínico sí se cumple el requisito del mínimo nivel de fiabilidad de las diferencias ( $r_{dd} > 0,40$ ). Este evento se añade a otros inconvenientes como la complejidad de su cálculo que hacen del método HA una alternativa con escasa utilidad práctica real (Bauer, Lambert y Nielsen 2004).

Los métodos basados en la tipificación de las diferencias entre ambos momentos de medida (TEI, ESG y DIT) se han mostrado demasiado poco exigentes en el umbral de 0,20. Se constata que un cambio de dos puntos entre las dos medidas ha sido suficiente para considerarlo como significativo (incluso con un solo punto con ESG en el grupo clínico). Este comportamiento en el estudio, unido al porcentaje inaceptable de falsos positivos mostrado (para este umbral) en el estudio de Pardo y Ferrer (2012), hacen desaconsejable el uso de estos métodos como alternativa al  $ICF_{JT}$ . Por otro lado, en el umbral de 1,96 ESG y DIT han tenido un funcionamiento más congruente con las demás alternativas. ESG se ha acercado más al método EN en su clasificación (menos restrictivo que  $ICF_{JT}$ ) y DIT al método GLN (más restrictivo que  $ICF_{JT}$ ). Además las conclusiones de Pardo y Ferrer (2012) para este umbral señalan una tasa muy baja de falsos positivos. Por lo que DIT y ESG con umbral 1,96, se muestran como una alternativa factible para clasificar cambio. Sin embargo, estos no tienen en cuenta la fiabilidad del instrumento de medida (tal y como lo hacen de forma directa  $ICF_{JT}$ , EN, GLN y HA), por lo que es necesario señalar que se obvia una posible fuente de error. Por su parte, el umbral de 1,96 para TEI resulta en

exceso restrictivo haciendo falta un cambio de 16 puntos en el grupo clínico para ser considerado como tal, y mostrándose con escasa congruencia con todos los demás métodos para este mismo grupo. Dado el funcionamiento extremo del método TEI no se configura como alternativa válida a  $ICF_{JT}$  en ninguno de los dos umbrales estudiados.

El método DTW resulta muy concordante en su clasificación con ESG en el umbral 1,96 y, por tanto, cercano también al método  $ICF_{JT}$ . Sin embargo, resulta más permisivo que ambos a la hora de señalar cambio. El planteamiento de DTW se basa, al igual que  $ICF_{JT}$ , en la generación de un error de medida a través de la fiabilidad del instrumento y la desviación estándar de referencia; sin embargo, no tienen en cuenta que se están utilizando dos muestras (dos momentos de medida) generando un error estándar menor que en el caso de  $ICF_{JT}$ . De hecho, es este matiz el que corrigen Christensen y Mendoza (1986) de la propuesta primigenia del  $ICF_{JT}$  (Jacobson, Follette y Revenstorf, 1984; 1986). En nuestro estudio, el DTW no ha resultado tan indiscriminado como los umbrales 0,20 de DIT, ESG y TEI, pero sí ha sido uno de los métodos que más fácilmente han clasificado cambio. La propia autora del método (Wyrwich, 2004) plantea un posible aumento del umbral de significación para compensar que efectivamente se han utilizado dos momentos de medida.

Los dos métodos que generan intervalos de pronóstico de la puntuación post tratamiento (IPP e IPI) se han mostrado bastante discrepantes entre sí en la clasificación de ambos grupos (concordancia moderada para el grupo clínico y baja para el grupo no clínico). El IPP, como ya ha explicitado anteriormente, ha clasificado de una forma poco fiable y errática a los dos grupos (señalando cambios en la dirección opuesta al sentido de la diferencia entre los dos momentos). Estos resultados son coincidentes con los hallados en el estudio de Pardo y Ferrer (2012) que señala un porcentaje medio de falsos positivos del 79% en las simulaciones de muestras de tamaño  $n=100$ . Por lo que se evidencia como un método no aceptable para clasificar cambio individual. Por otro lado, el IPI se comporta de una forma semejante al GLN, aunque resulta un poco más restrictivo. Este mayor nivel de exigencia mostrado en el estudio no le aleja en exceso de  $ICF_{JT}$ , pero como detectan Pardo y Ferrer (2012) sí supone aumentar la sensibilidad (evitar falsos positivos). Por ello, el IPP se presenta como una alternativa viable a  $ICF_{JT}$ , aunque se pueden señalar dos impedimentos relevantes para su uso en la práctica habitual: aumenta la

sensibilidad a costa de la especificidad (evitar falsos negativos), y necesita de un grupo control sin cambio sintomático para generar el intervalo (algo complicado de conseguir en el contexto clínico real).

Del método CPM 50% se conocían a priori sus limitaciones: se plantea para calificar cambios en grupos clínicos y sólo tiene en cuenta el cambio en el sentido de la mejoría (Hiller y Schindler, 2011). Por ello, la gran discrepancia mostrada con los demás métodos en el grupo no clínico ha sido la esperada. Esta circunstancia limita las conclusiones que se pueden sacar del comportamiento de concordancia del método en el grupo no clínico, siendo necesario obviar los casos que empeoran. Por otro lado, los resultados del grupo clínico han permitido hacer totalmente comparable el CPM 50% con las demás alternativas. De esta forma, y a pesar de establecer un criterio de cambio de lógica muy diferente a  $ICF_{JT}$ , se ha mostrado un alto nivel de acuerdo con éste en la clasificación de la muestra clínica. Esta situación es totalmente inversa a la clasificación del grupo no clínico, donde resulta mucho menos restrictiva a la hora de clasificar como mejorados que  $ICF_{JT}$ . Se hace evidente lo poco exigente que resulta el método para señalar cambio cuando el participante tiene su puntuación inicial cerca del punto de corte con la funcionalidad o por debajo de él. Es decir, sólo mostrará un comportamiento similar al  $ICF_{JT}$  con grupos de paciente que puntúen alta intensidad en el momento post tratamiento, resultando más indiscriminado con los pacientes menos graves. Teniendo en cuenta ambos grupos resulta menos restrictivo que  $ICF_{JT}$ , tal y como sucede en el estudio de Hiller, Schindler y Lambert (2012). Se puede, concluir como lo hacen los autores del estudio referido, que el CPM 50% no se puede plantear como una alternativa a  $ICF_{JT}$ , sino como un complemento (al tener en cuenta elementos que el  $ICF_{JT}$  no contempla tales como la gravedad inicial de los pacientes).

En conclusión, el  $ICF_{JT}$  se ha vuelto a mostrar como un método más útil que las estrategias de análisis de promedios, tanto para la determinación de la efectividad del tratamiento de forma colectiva, como para la toma de decisiones clínica individual. A pesar de que la situación evaluada ha generado conclusiones grupales más compatibles entre ambos métodos que en los estudios anteriores, el  $ICF_{JT}$  ha seguido ofreciendo más información relevante al evaluador: matizando el éxito terapéutico en el grupo clínico, y detectando casos de cambio relevante (casi de consideración clínica en algunos de ellos) en el grupo no clínico.

Con respecto a la comparativa de métodos de análisis individual, el  $ICF_{JT}$  se ha evidenciado claramente más adecuado y fiable que los métodos HA, IPP y TEI para las circunstancias del estudio. Por otro lado, y obviando los umbrales 0,20 (demasiado indiscriminados), se ha constatado un nivel de concordancia alto de  $ICF_{JT}$  con las demás alternativas. Situándose como un método moderado en relación al nivel de exigencia para clasificar cambio: algo más permisivo que IPI, DIT 1,96 y GLN, y más restrictivo que EN, ESG 1,96 y DTW. Este comportamiento medido coincidente con estudios anteriores, junto a su facilidad computacional (frente a EN o GLN) y sus menores requerimientos previos (frente a IPI o ESG) llevan a confirmarlo como el método más idóneo para el análisis de cambio individual (Lambert, Hansen y Bauer, 2008; Gómez y Roussos, 2012). Sin embargo, teniendo en cuenta los resultados del presente estudio, y las reflexiones de estudios previos sobre la detección de falsos positivos y de aporte de información complementaria (Pardo y Ferrer 2012; Hiller, Schindler y Lambert, 2012), concluimos deseable para estudios de efectividad de tratamientos añadir un segundo criterio de análisis individual adicional al  $ICF_{JT}$ . Eligiendo este complemento entre los que se hayan desarrollado bajo lógica diferente (no basada en el error estándar de la medida) y que no haya presentado un comportamiento anómalo, tales como IPI, CPM 50% o DIT 1,96.



## Capítulo 11

### Estudio 4: Uso del índice de cambio fiable en la evaluación de la efectividad de intervenciones clínicas. Aplicación en un programa formativo de adhesión al tratamiento en asma (\*<sup>1</sup>)

---

#### 11.1. Objetivo

Este estudio tiene el objetivo de contrastar la metodología de ICF<sub>JT</sub> con la metodología habitual de significación estadística de diferencias grupales, en el marco de un ensayo clínico (con *grupo control*) realizado para determinar la *efectividad* (en términos de mejoría de un *indicador bio-métrico objetivo* en condiciones de práctica real) de una *intervención educativa* en profesionales sanitarios. Cuyo contexto clínico es el seguimiento y tratamiento de pacientes con una patología respiratoria *crónica* en los centros ambulatorios de la *Red Pública de Salud*.

#### 11.2. Hipótesis

La metodología de ICF<sub>JT</sub> se mostrará más sensible y reportará más información útil para la valoración de la efectividad de *una intervención educativa en Asma* (en el sentido de mejoría *espirométrica de capacidad pulmonar*), que la metodología habitual de análisis de promedios y significación estadística de las pruebas de contraste de hipótesis nula.

---

<sup>1</sup> El contenido del presente capítulo ha sido aceptado a publicación en una revista científica. Las diferencias entre el contenido de este estudio y el artículo publicado se deben fundamentalmente a ajustes de formato y a una mayor extensión de la discusión. La referencia de la publicación es la siguiente: **Montero M. et al. (2015) Uso del índice de cambio fiable en la evaluación de la efectividad de intervenciones clínicas: aplicación en un programa formativo en asma. Atención Primaria: <http://dx.doi.org/10.1016/j.aprim.2014.12.007>**

## 11.3. Metodología

### 11.3.1. Procedimiento y participantes

A través de la información de los historiales médicos se contactó con 100 pacientes que estaban siendo tratados por asma en el servicio de neumología de cinco centros de salud de asistencia comunitaria de la provincia de Vizcaya. Fueron seleccionados por cumplir los siguientes criterios de inclusión: presentar diagnóstico de asma con afección moderada o severa, presentar sintomatología clínica en el último año, estar tratados con grupos de fármacos equivalentes, no padecer otras enfermedades asociadas que pudiesen interferir en el curso de la enfermedad y que no hubiesen recibido intervención psicológica o psiquiátrica en el último año. Fueron convocados a una entrevista en la que se les solicitaba participación en el estudio y en caso de consentimiento se realizaba la recogida de datos clínicos conformando los datos de línea base. Se registró la información diagnóstica de cada uno de los participantes, sus resultados en las pruebas funcionales de seguimiento (espirometría forzada) y las prescripciones farmacológicas realizadas por el médico. Cada participante de forma individual fue entrevistado por un clínico durante 60 minutos recogiendo o contrastando las variables descriptivas relevantes. Se les informó que se estaban reuniendo datos sobre los enfermos asmáticos a fin de conocer mejor su situación. Se solicita consentimiento informado para el uso de sus datos y se les comunica que volverán a ser entrevistados en una segunda entrevista transcurridos seis meses.

Después de finalizar tomar las medidas pre-tratamiento se iniciaron las sesiones de intervención en los profesionales sanitarios en el “*Programa de Mejora de Adhesión al Tratamiento*” que configura la variable independiente del estudio. Se generaron dos grupos de pacientes según fuesen a recibir atención de los profesionales intervenidos, condición “*Experimental*”, o no, condición “*Control en lista de espera*”. A los seis meses de la primera evaluación se establece una segunda recogida de datos en cada participante para constatar el efecto diferencial de la intervención donde se registran las mismas variables actualizadas a ese momento.

En la Tabla 37 se registran las principales características de la muestra, así como su distribución en los grupos control (n = 38) y experimental (n = 38). Se observa que se encuentran homogéneamente distribuidas en ambos grupos, no existiendo diferencias significativas, exceptuando el sexo que después de la asignación no quedó balanceado, circunstancia tenida en cuenta para el posterior análisis de resultados.

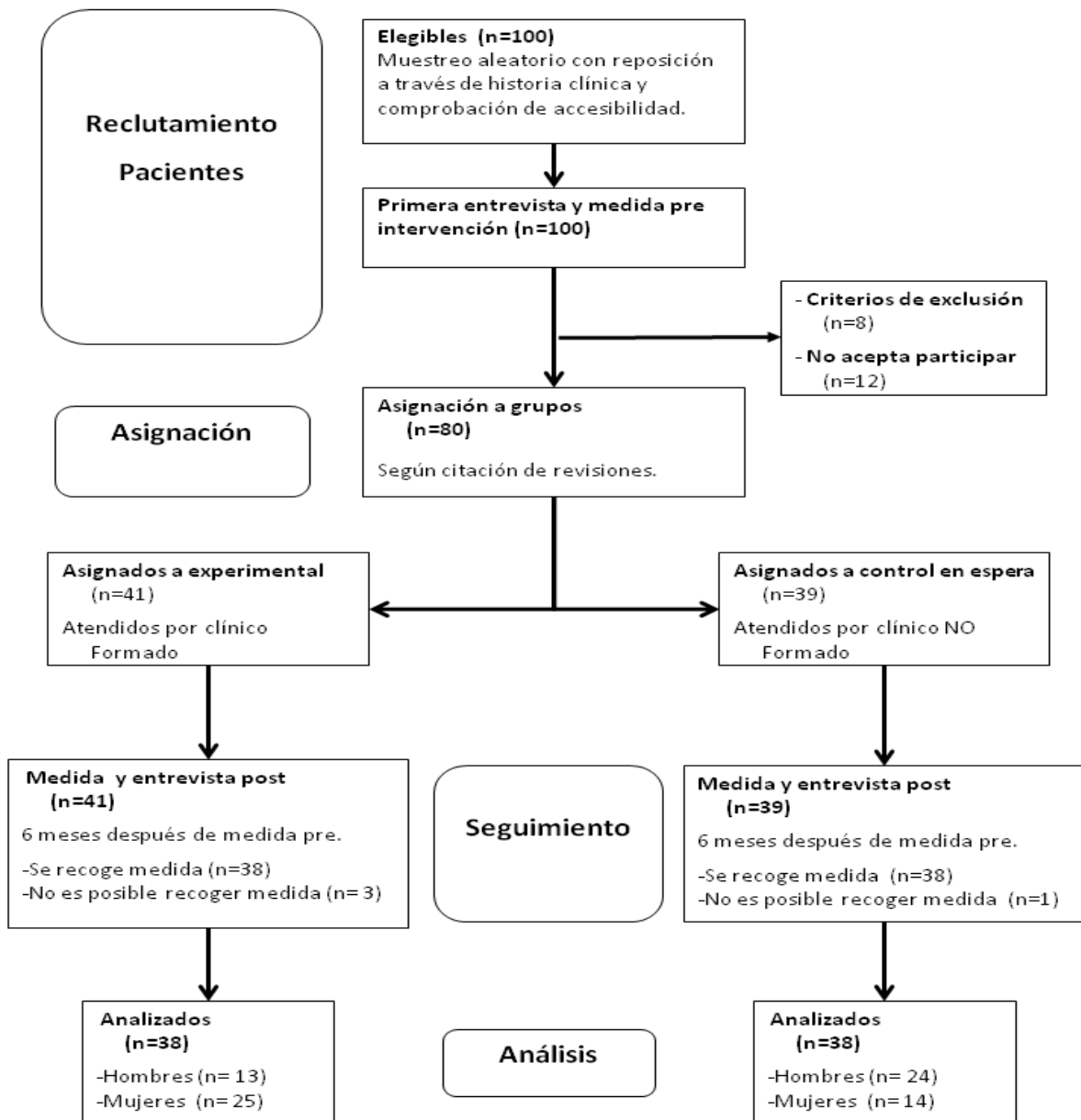
**Tabla 37**

*Datos de caracterización muestral del estudio 4*

		Total (n=76)		Grupo Experimental (n= 38)		Grupo Control (n= 38)		Prueba	g.l.	p.
		n	%	n	%	n	%			
<b>Sexo</b>	Hombre	37	48,7	13	34,2	24	63,2	6,37	1	0,01
	Mujer	39	51,3	25	65,8	14	36,8			
<b>Fuma</b>	No	51	67,1	23	60,5	28	73,7	1,49	1	0,22
	Si	25	32,9	15	39,5	10	26,3			
<b>Edad inicio</b>	Infancia	21	27,6	9	23,7	12	31,6	2,12	2	0,34
<b>Síntomas</b>	Adolescencia	26	34,2	16	42,1	10	26,3			
	Adulthood	29	38,2	13	34,2	16	42,1			
<b>Incidencia</b>	No hubo	38	50	20	52,6	18	47,4	0,84	2	0,65
<b>Síntomas</b>	Leves	25	32,9	13	34,2	12	31,6			
<b>24 horas</b>	Moderados	13	17,1	5	13,2	8	21,1			
<b>Cumplir</b>	No tomo	6	7,9	2	5,3	4	10,5	1,92	3	0,59
<b>Tratamiento</b>	Tomo menos	28	36,8	15	39,5	13	34,2			
<b>24 horas</b>	Tomo más	4	5,3	3	7,9	1	2,6			
	Correctamente	38	50	18	47,4	20	52,6			
		<b>M</b>	<b>DE</b>	<b>M</b>	<b>DE</b>	<b>M</b>	<b>DE</b>	<b>t</b>	<b>g.l.</b>	<b>p.</b>
<b>Edad</b>	Años	41,2	15,2	40,9	13,2	41,5	17,2	-0,16	69,4	0,87
<b>Dosis</b>	Antiinflamatorios	1414,5	589,6	1410,5	556,0	1419	629,0	-0,06	74	0,95
<b>Dosis</b>	Broncodilatador	256,7	340,1	258,8	320,2	254	370,7	0,52	55	0,95

Notas.- La prueba para el caso de variables nominales el Ji cuadrado y para variables continuas la t de Student con aplicación de las pruebas de corrección en caso de incumplimiento de requisitos. M = Media; DE =Desviación Estándar; g.l. = grados de libertad; p= valor de probabilidad de significación.

Del total de pacientes seleccionados, un 8% fueron descartados tras la entrevista inicial por la constatación del incumplimiento de alguno de los criterios de inclusión y un 12% no aceptó participar. Durante el desarrollo del estudio se perdieron 4 casos, de modo que la muestra de análisis quedó constituida finalmente por un total de 76 participantes. En la Figura 10 se presenta el esquema del estudio. La edad de los participantes oscila entre los 15 y 65 años (Media = 41,2; DE = 15,2), siendo 37 varones y 39 mujeres.



**Figura 10.** Esquema general del estudio 4 en diagrama de flujo.

### 11.3.2. Diseño

El presente estudio es un ensayo clínico pragmático (naturalístico) de 2 cohortes sincrónicas (Hotopf, Churchill y Lewis, 1999). La figura 10 recoge el planteamiento y desarrollo del ensayo en forma de diagrama de flujo según los estándares internacionales de información de ensayos clínicos (Schulz, Altman y Moher, 2010).

### 11.3.3. Análisis de datos

Las variables de caracterización fueron expresadas en medias y desviaciones típicas, y en frecuencias y porcentaje según fueran de tipo escalar o nominal. Para realizar el análisis de resultados según el paradigma clásico se realizó un ANOVA mixto con medidas repetidas según las dos condiciones experimentales. Para el contraste bivariado de diferencias de medias o proporciones entre grupos se realizaron pruebas *F de Snedecor* y *Ji Cuadrado*, respectivamente, aplicándose las correcciones oportunas en caso de heterocedasticidad (análisis de *Brown-Forsythe*) o incumplimiento de normas de aplicación (Prueba exacta de *Fisher*). En todos los contrastes anteriores se asumió como umbral de significación  $p < 0,05$  que se corresponde con un nivel de confianza de 95%.

Para evaluar la significación estadística del cambio individual de cada sujeto se calculó el Índice de Cambio Fiable (ICF) según lo establecido por Jacobson y Truax (1991):  $ICF = [(PEF_{post} - PEF_{pre}) / S_{dif}]$ , donde,  $S_{dif}$  es el error estándar de la diferencia de las medidas [ $S_{dif} = DE \cdot \sqrt{(1 - R_{xx}) \cdot \sqrt{2}}$ ]; siendo  $DE$  la desviación estándar en la muestra clínica en el momento post tratamiento y  $R_{xx}$  la fiabilidad del instrumento de medida. Se consideró estadísticamente significativo si  $ICF = \geq 1,96$  o  $ICF = \leq -1,96$  (valor  $z$  correspondiente al punto en una curva normal de niveles de confianza del 95%). De este algoritmo se deriva que un cambio significativo se produce cuando  $PEF_{post} - PEF_{pre} \geq 1,96 * S_{dif}$ , o  $(PEF_{post} - PEF_{pre}) \leq -1,96 * S_{dif}$ . El cambio clínico ha sido expresado de forma gráfica para facilitar una interpretación más intuitiva.

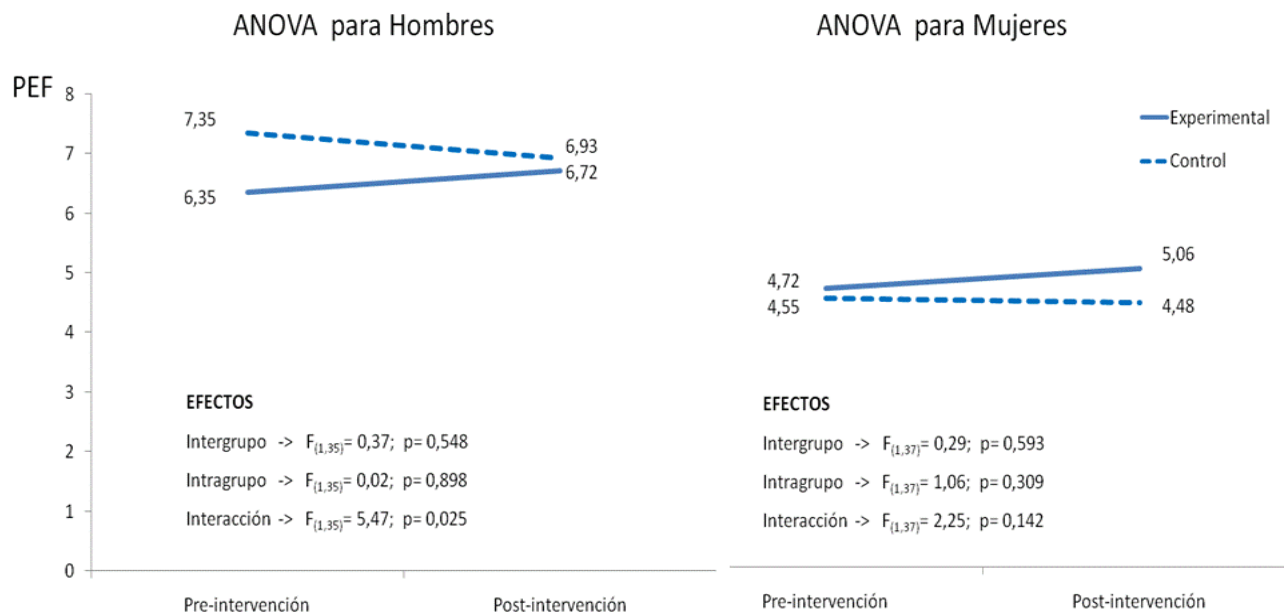
Para poder establecer un criterio de significación clínica del cambio individual se calculó el Punto de Corte (*PC*) propuesto por Jacobson y Truax (1991) a partir del cual considerar al sujeto más cerca de la media de la población funcional que de la disfuncional. Siendo  $PC = [(DE_f \cdot M_c) + (DE_c \cdot M_f)] / (DE_f + DE_c)$ , donde  $M_f$  y  $DE_f$  corresponden a la Media y Desviación Estándar de la muestra funcional-normativa (Roca et al. 1986); y  $M_c$  y  $DE_c$  corresponden con la Media y Desviación Estándar de la muestra clínica. Este punto de corte supone una estimación donde se pondera las medias de ambas distribuciones en función de la variabilidad inherente a cada una de estas distribuciones (funcional y disfuncional). Como segundo criterio para la valoración del cambio clínico, se tomó como punto de corte normativo los valores promedios de referencia ( $PC = 9,51$  y  $6,13$ , para hombres y mujeres, respectivamente) del estudio de población general de Roca y colaboradores (1986).

## **11.4. Resultados**

### **11.4.1. Análisis mediante contrastes de significación clásicos de promedios**

En la Figura 11 se presentan los resultados del ANOVA mixto con medidas repetidas. Dado que la variable 'sexo' presenta diferencias fisiológicas en capacidad pulmonar, las pruebas de hipótesis han sido realizadas de forma independiente para cada categoría de esta variable.

Se observa un patrón similar de cambio en hombres y mujeres; un ligero crecimiento de la capacidad pulmonar (PEF) en el grupo experimental y una ligera reducción o estabilidad en el grupo control. No obstante, tan sólo se ha observado significación estadística en el efecto de interacción para el caso de los hombres ( $F_{(1,35)} = 5,47$ ;  $p = 0,025$ ), no alcanzando significación estadística ninguno de los demás contrastes intergrupo o intrasujeto.

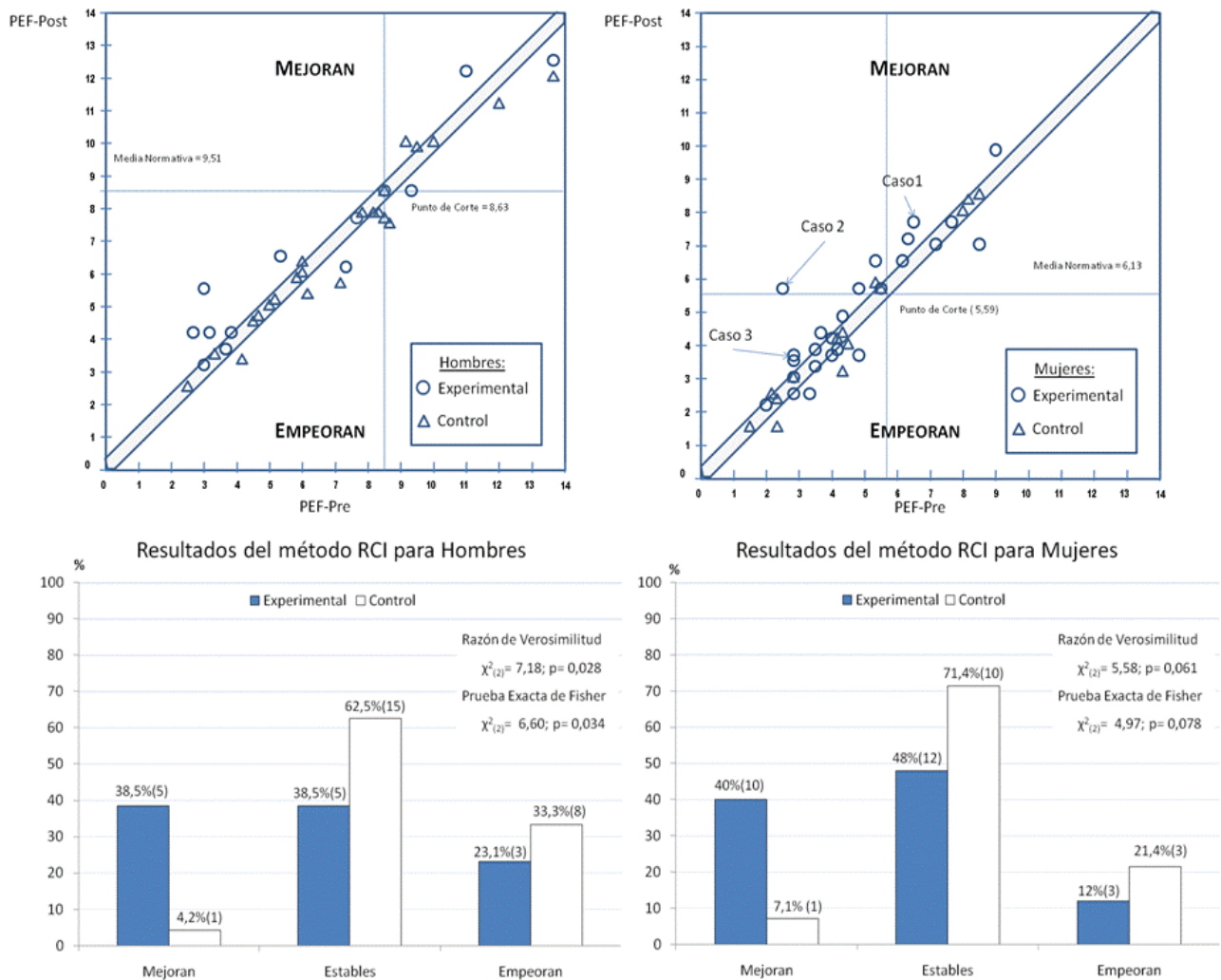


**Figura 11.** ANOVA medidas repetidas para las dos condiciones experimentales del estudio 4.

#### 11.4.2. Análisis de cambio individual mediante Índice de Cambio Fiable

La mitad superior de la Figura 12 muestra la distribución de valores en las gráficas de cambio terapéutico (una para cada sexo) y la estimación del cambio fiable mediante el cálculo del error estándar de la diferencia ( $S_{dif} * 1,96$ ). Ésta estimación viene representada por la franja diagonal que cruza los dos gráficos. Todos los casos que se encuentran incluidos en ella representan participantes cuyo cambio en la puntuación PEF antes y después de la intervención no han variado de forma estadísticamente significativa. Los casos que se encuentran en el espacio inferior son personas que han evolucionado hacia un empeoramiento de su capacidad respiratoria y presentan  $ICF \leq -1,96$ , por lo que el cambio ha de considerarse estadísticamente significativo. Y los casos situados en el espacio superior han evolucionado hacia una mejoría, con valores  $ICF \geq +1,96$ , indicando un cambio estadísticamente significativo. Por otro lado, el criterio de significación clínica viene representado en ambos gráficos por la línea horizontal que

parte del eje de ordenadas (puntuación post-tratamiento), representando el PC entre funcionalidad (por encima de la línea) y disfuncionalidad ( por debajo de la línea). En el Anexo se adjunta una tabla de resultados con los valores numéricos individuales que se han calculado para esta clasificación.



**Figura 12.** Resultados según Índice Cambio Fiable, separados por sexo del estudio 4

De esta forma se puede identificar la siguiente clasificación del cambio acontecido en los **hombres**. En el grupo experimental (n=13) cinco pacientes se consideran en situación de “*No Cambio*”, uno en “*Empeoramiento dentro de la funcionalidad*”, dos “*Empeoramiento en disfuncionalidad*”, cuatro en “*Mejoría dentro de la disfuncionalidad*” y uno como “*Recuperado*”(Mejoría dentro de funcionalidad). Mientras que en el grupo control (n=24) 15 pacientes se clasifican en “*No Cambio*”, dos en “*Empeoramiento dentro de la funcionalidad*”, seis en “*Empeoramiento en disfuncionalidad*”, ninguno en “*Mejoría dentro de la disfuncionalidad*” y uno como “*Recuperado*”.

También se puede señalar la clasificación de cambio en las **mujeres**. En el grupo experimental (n=25) 12 pacientes se consideran en situación de “*No Cambio*”, uno en “*Empeoramiento dentro de la funcionalidad*”, dos en “*Empeoramiento en disfuncionalidad*”, cuatro en “*Mejoría dentro de la disfuncionalidad*” y seis “*Recuperado*”. Mientras que en el grupo control (n=14) 10 pacientes se clasifican en “*No Cambio*”, ninguno en “*Empeoramiento dentro de la funcionalidad*”, tres en “*Empeoramiento en disfuncionalidad*”, ninguno en “*Mejoría dentro de la disfuncionalidad*” y uno como “*Recuperado*”.

Por otro lado, en la mitad inferior de la Figura 12 se ofrece la proporción de casos que presentan mejoría ( $ICF \geq 1,96$ ), estabilidad ( $1,959 < ICF < -1,959$ ) o empeoramiento ( $ICF \leq -1,96$ ) en cada uno de los grupos de intervención y para cada uno de los sexos (únicamente criterio de cambio estadísticamente fiable). Se observa que, en el caso de los **hombres**, existen diferencias estadísticamente significativas ( $p= 0,034$ ) que reflejan una mayor proporción de casos que mejoran en el grupo experimental (38,5% vs 4,2%), frente a una mayor proporción de casos estables o de empeoramiento en el grupo control. Patrón similar se halla entre el grupo de **mujeres**, si bien en este caso la prueba exacta de comparación de proporciones no ha alcanzado significación estadística ( $p= 0,078$ ). En el Anexo 3, se adjunta los análisis de resultados individuales que se han calculado para generar los graficos de la Figura 12.

## 11.5. Discusión y conclusiones

Se ha mostrado los resultados de la intervención según dos metodologías de análisis de resultados, uno basado en los métodos clásicos y otro alternativo consistente en la estimación del Índice de Cambio Fiable. Los resultados muestran que el segundo método permite captar con mayor precisión los cambios clínicos habidos.

Dado el carácter didáctico de este trabajo quisiéramos hacer algunas matizaciones respecto a la variable de resultado contemplada. El índice de cambio fiable es un tipo de metodología ampliamente utilizado cuando la variable de resultado es de tipo 'score' (utilización de cuestionarios que valoran la percepción subjetiva respecto a un fenómeno y que están afectos de un error de medida valorado por su grado de fiabilidad) y mucho menos utilizado en valoraciones que se efectúan mediante métodos más objetivables. La aplicación de esta metodología a una variable de resultado como puede ser la función respiratoria es novedosa, pero no por ello inviable. Quizá pueda ser más discutible la utilización del PEF como único indicador de la función respiratoria, pero como se ha indicado se ha optado por éste por ofrecer unos resultados que permiten una mayor claridad expositiva en la demostración de los hallazgos y conclusiones alcanzadas con el método clásico de análisis frente al alternativo del índice de cambio fiable. Sin duda alguna, un estudio encaminado a valorar los resultados e impacto de un programa de formación en la adherencia al tratamiento de asma debiera contemplar una combinación de criterios de resultados como se aconseja en la guía GINA (2010), pero excedería los objetivos del presente estudio. Hemos considerado conveniente realizar estas matizaciones con el fin de abordar con mayor especificidad la discusión de los resultados, a la cual se procede.

El análisis clásico, basado en el contraste de medias por pruebas ANOVA, no permite aceptar una hipótesis de diferencias en la mejora de la capacidad pulmonar (PEF) del grupo intervenido respecto al control, por lo que se llegaría a la conclusión de que el programa de formación y mejora de la adhesión al tratamiento que se ha ofrecido a los profesionales sanitarios no ha tenido efecto sobre esta medida de resultado, lo que impide la constatación de su efectividad.

No obstante, el análisis de los resultados a través de la estimación del índice fiable de cambio sí evidencia diferencias entre la evolución de los pacientes de los dos grupos. Únicamente atendiendo al primer criterio de la metodología de ICF<sub>JT</sub> (si el cambio es estadísticamente fiable o no) se observa que tanto en hombres como en mujeres el grupo control presenta una proporción mayor de casos de “*No Cambio*” y de “*Empeoramiento*” (ver columnas de “*Estables*” y “*Empeoran*” en gráficas de la mitad inferior de Figura 3) que el grupo experimental. Y además la diferencia más grande entre proporciones de clasificación se da en la condición de “*Mejoría*”, resultando claramente mayor para ambos sexos en el grupo experimental que en el grupo control. Al aplicar un estadístico de significación para estas diferencias ha resultado significativa la diferencia de proporciones entre el grupo control y el grupo experimental para los hombres pero no para las mujeres. A pesar de lo anterior, si se amplía el análisis con la consideración de la significación clínica (criterio de PC funcionalidad), se observa que en el grupo experimental el 24% de las mujeres han clasificado su cambio en la categoría de “*Recuperado*”, es decir que además de cambiar de forma estadísticamente fiable en dirección a la mejoría, se han situado por encima del punto de corte de funcionalidad, frente al único caso del grupo control que se ubica en la misma categoría. Por otro lado, se observa una mayor proporción de casos clasificados “*Empeoramiento en disfuncionalidad*” (cambio fiable de reducción de PEF por debajo de PC) en el grupo control que en el grupo experimental. Por todo ello, los datos reportados por el método ICF<sub>JT</sub>, hacen difícil mantener que no han existido diferencias de cambio clínico relevantes entre ambos grupos de estudio. Concluyendo que de forma colectiva se observa una mejor evolución del grupo experimental.

Aunque el ensayo clínico se planificó con grupos equivalentes en tamaño y características, el procedimiento de asignación aleatoria a las condiciones experimentales (generado a partir de orden de citación preestablecido en los centros ambulatorios) provocó que no hubiese el mismo número de hombres y mujeres en los grupos control y experimental. Dado que el sexo tiene un relevancia crucial para cualquier indicador de capacidad pulmonar (Sociedad Española de Neumología y Cirugía Torácica –SEPAR– et al., 2009), se ha tenido que desglosar el análisis de resultados separando hombres y mujeres para evitar un importante sesgo. Esto ha generado subgrupos de análisis mucho más reducidos en tamaño ( $n < 30$ ), de lo esperado inicialmente, reproduciendo la principal dificultad metodológica detectada por la revisión Cochrane (Haynes,

Ackloo, Sahota, McDonald y Yao, 2008) sobre los estudios de efectividad de programas de adhesión al tratamiento. Situación que propicia que si únicamente se atendiera a los en los estadísticos de significación de promedios y las limitadas conclusiones que se pueden sacar de ellos en este contexto experimental, el presente estudio no aportaría evidencias sobre los beneficios del programa realizado. Sin embargo, el método  $ICF_{JT}$ , se ha mostrado mucho más flexible y preciso a la hora de acometer el mismo análisis, sin verse afectado por la reducción de los grupos de análisis. Señalando que no se debe desestimar los efectos del programa en el grupo a estudio, por lo que aporta argumentos a favor de la realización de este tipo de intervenciones en el ámbito clínico. Por supuesto, el reducido tamaño del estudio no permite generalizar unas conclusiones definitivas sobre su efectividad, sin embargo, debido a la fácil modularidad del método  $ICF_{JT}$ , se podría apilar los datos recogidos con los de futuros estudios que apliquen de la misma forma el programa (por ejemplo en otras comarcas de la comunidad) generando un evidencia mucho más consistente.

Además de las ventajas señaladas del método  $ICF_{JT}$  para la determinación de la efectividad en términos generales, se puede destacar su utilidad para las decisiones clínicas inmediatas ante cada caso individual. En la gráfica de resultado de mujeres (en la mitad superior de la Figura 12) se han señalado tres casos paradigmáticos diferentes de evolución clínica favorable según esta metodología. El caso 3 es el de una mujer que ha progresado hacia la mejoría de su capacidad pulmonar ( $PEF_{post} > PEF_{pre}$ ) y dicho cambio ha sido estadísticamente significativo. Sin embargo, no alcanza el punto de corte que define la funcionalidad por lo que no se puede considerar su situación como normalizada, resultando su puntuación PEF post intervención aun de consideración clínica. El caso 2 es una paciente que además de presentar una mejoría estadísticamente significativa, su puntuación tras la intervención se sitúa por encima del punto de corte. Se interpreta esta mejoría como clínicamente relevante, situando su puntuación más cerca a la media de PEF de la población general que la del grupo de referencia clínico, evidenciándose de esta forma una recuperación clínica desde un estado inicial de disfuncionalidad. Por su parte, el caso 1 se halla en la misma situación que el caso 2, ya que muestra un cambio estadísticamente significativo y clínicamente relevante, pero en este caso la interpretación es distinta a la del caso 2 ya que esta persona antes de la intervención ya se encontraba en una

situación de funcionalidad siendo una mejoría significativa aunque de menor impacto clínico que el caso 2 que parte de una situación más grave.

Todos estos matices individuales son de inmediato acceso al clínico y en su presentación grafica fácilmente interpretables, sin ser necesario esperar a generar estadísticos grupales, ni tener una formación metodológica específica para su comprensión. Esta ventaja pragmática hace que la inclusión de este método para evaluar resultados en entornos sanitarios reales sea más idónea y fluida que los métodos clásicos, ya que se plantea de utilidad para el clínico en la toma de decisiones de su día a día.

En conclusión, en este estudio el método clásico ha mostrado ciertas limitaciones para lo toma de decisiones respecto a si ha existido o no un cambio en la variable resultado. La metodología basada en puntuaciones fiables de cambio ha permitido una estimación más adecuada de la proporción de casos que han experimentado cambio antes y después de la intervención, permitiendo a sí mismo, para cada caso individual, la constatación de si dicho cambio (además de ser estadísticamente significativo) resulta ser clínicamente relevante. Señalando en cómputo una evolución más favorable para el grupo experimental que para el grupo control. Por otra parte, la metodología ICF<sub>JT</sub> y su presentación gráfica de los resultados de cambio se plantea una adecuada herramienta para que el clínico pueda valorar sus decisiones respecto a los sujetos individuales.



## Capítulo 12

### Discusión y conclusiones generales

---

Aunque cada uno de los cuatro estudios que integran la parte empírica del trabajo cuentan con su propia discusión de resultados, en este capítulo final se pretende dar respuestas a las hipótesis generales inicialmente planteadas para la tesis en su conjunto. Para ello, en esta discusión se tendrán en cuenta de forma global los resultados de todos los estudios. Después de esto, se establecerán las conclusiones generales del trabajo. También en este apartado, se explicitarán las principales limitaciones que se han encontrado en el desarrollo de la presente investigación. Y finalmente, se hará referenciase a las aportaciones prácticas del trabajo y se sugerirán posibles líneas futuras de investigación e implementación de la metodología ICF<sub>JT</sub>.

#### 12.1. Contraste de hipótesis y discusión

*1º Hipótesis:* La metodología ICF<sub>JT</sub> se mostrará *más idónea y útil* que la metodología de significación de promedios en el análisis de resultados en condiciones de práctica clínica natural.

En los cuatro estudios se ha planteado la comparación entre el análisis de resultados mediante pruebas de significación de diferencias grupales (contraste de hipótesis nula habitual) y el análisis de resultados individuales con el método ICF<sub>JT</sub>.

En el *estudio 1* las conclusiones derivadas de las pruebas significación y las magnitudes del efecto de las diferencias de medias ( $p < 0,001$ ;  $d > 0,50$ ) son muy favorables, en sentido de reducción en todas las dimensiones sintomáticas seleccionadas. Mientras que el ICF<sub>JT</sub> señala dos

indicadores con muy bajo porcentaje de casos que han experimentado un cambio realmente fiable (>25%), y una sola dimensión que ha superado el 40% de casos que se han podido considerar recuperados de su gravedad clínica (clínicamente significativo). Por otro lado, el ICF<sub>JT</sub> ha permitido observar las discrepancias individuales en cada indicador respecto a la tendencia grupal, y observar la heterogeneidad tanto de perfiles sintomáticos iniciales como áreas mejoradas después del tratamiento.

En el *estudio 2* las pruebas de significación de las diferencias de medias han vuelto a resultar en todos los indicadores significativas en sentido de reducción de gravedad ( $p < 0,05$ ), aunque las magnitudes del efecto han mostrado como muy poco relevante el efecto de uno de la indicadores ( $d < 0,20$ ), y moderadamente relevantes los demás. Por su parte con ICF<sub>JT</sub> ningún indicador ha registrado una proporción superior al 50% de pacientes con cambios estadísticamente fiables hacia mejoría o empeoramiento, siendo estos en su mayoría pacientes que mejoran pero siguen sin alcanzar niveles funcionales. Tres de las cuatro dimensiones evaluadas muestran casos que empeoran de forma estadísticamente fiable y solo dos muestran un único caso como recuperado. A nivel de análisis de curso clínico individual el ICF<sub>JT</sub> ha detectado casos que, independientemente de dirección de cambio mayoritario de la muestra, han empeorado o mejorado con tal magnitud que han cambiado de forma fiable de un área a otra de funcionalidad.

El *estudio 3* ha mostrado el mayor grado de acuerdo entre los dos métodos de los estudios realizados. La prueba de significación a través análisis de la varianza, ha señalado un cambio estadísticamente significativo entre los dos momentos de medida en la variable resultado ( $p < 0,001$ ) siendo además significativo y muy relevante el efecto de pertenecer a uno u otro grupo para este cambio (interacción  $p < 0,001$  y  $\eta^2 = 0,64$ ), observándose en las medias grupales que el grupo experimental reduce la intensidad sintomática media en gran medida y el grupo control permanece sin apenas cambio. El ICF<sub>JT</sub> muestra que un 80% de los casos del grupo experimental realizan un cambio estadísticamente fiable en dirección a la mejoría, mientras que en el grupo control el 94 % de los participantes no realiza cambio fiable hacia ninguna dirección. A nivel colectivo se concluye lo mismo con los dos métodos (resultados muy favorables de la intervención) pero el ICF<sub>JT</sub> matiza la significación clínica (46% de recuperación funcional en

grupo experimental). Es al nivel del curso clínico de cada participante donde el ICF<sub>JT</sub> muestra su valor añadido, detectando en el grupo control cambio fiable de empeoramiento (seis casos) que se quedan al límite de la disfuncionalidad).

El *estudio 4* concluye a través del análisis de la varianza grupal que no ha existido una diferencia estadísticamente significativa (ni en hombres ni en mujeres) entre los dos momentos de medida (intragrupos  $p > 0,05$ ), ni entre los dos grupos de estudio (intergrupos  $p < 0,05$ ), únicamente mostrándose efecto de interacción para los hombres ( $p < 0,05$ ). No pudiendo evidenciarse que el programa educativo haya influido sobre la variable resultado de forma significativa. Sin embargo, el ICF<sub>JT</sub> sí ha detectado un patrón diferencial la clasificación del cambio en los pacientes de los dos grupos de estudio, y también ha señalado pacientes que realizan un cambio de tal magnitud que se puede considerar clínicamente relevante. El grupo control ha mostrado una tendencia mayor a no realizar cambios fiables que el grupo experimental (en hombres 62,5% vs 38,5% y en mujeres 71,4% vs 48%), mientras que el grupo experimental ha clasificado una proporción claramente mayor de casos que han mejorado de forma fiable que el grupo control (en hombres 38,5% vs 4,2% y en mujeres 40% vs 7,1%). Repitiéndose esta tendencia también para los casos de mejoría clínicamente significativa (en funcionalidad). Y a nivel individual el ICF<sub>JT</sub> ha permitido hacer visible gráficamente para el profesional el curso clínico diferencial de cada paciente incluso entre los que se han clasificado en la misma categoría de cambio.

En los cuatro estudios se ha observado una mayor riqueza descriptiva del ICF<sub>JT</sub> para la valoración de los efectos de las intervenciones realizadas. A nivel grupal se ha estudiado la efectividad de forma más ajustada a la realidad, y de hecho ha resultado, en tres de los cuatro estudios, un método más exigente y conservador en sus conclusiones que la consideración de promedios habitual (mucho más indiscriminada). El *estudio 4* ha sido el único donde el ICF<sub>JT</sub> ha resultado ser más favorable en sus conclusiones de efectividad que el análisis clásico, aunque como se argumentará más adelante esto se haya debido más al efecto del tamaño grupal del análisis sobre las pruebas de significación clásicas que a una característica del ICF<sub>JT</sub>. Además de esta utilidad colectiva, con los mismos cálculos el ICF<sub>JT</sub> ha permitido generar una clasificación individual de aplicabilidad directa sobre la toma de decisiones clínica, además ha sido posible

presentar de forma gráfica, haciendo fácilmente accesible y entendible para el profesional no metodólogo, la evolución particular de cada paciente en los indicadores elegidos.

Por todo ello, **se confirma la 1º hipótesis**: en las circunstancias de los estudios realizados el método ICF<sub>JT</sub> ha **resultado más útil y adecuado** para el análisis de resultados que el uso de pruebas de significación de diferencias grupales. Esta afirmación se especifica con el contraste de las siguientes sub-hipótesis.

*Sub-Hipótesis 1.1:* El ICF<sub>JT</sub> se evidenciará más adecuado para la evaluación el cambio en diferentes *medidas de resultado* (referidas a intensidad sintomática, funcionalidad en vida cotidiana y a normalidad biométrica) que la significación de promedios.

En los *estudios 1 y 3* se han seleccionado medidas de resultado que reflejan intensidad sintomática de tipo psicopatológico (auto percibida por el paciente). En las pruebas de significación de promedios cualquier aumento o disminución de la puntuación entre los dos momentos de medida, ha afectado a las medias grupales y por tanto las conclusiones finales. En cambio, con el ICF<sub>JT</sub> se ha generado un error estándar de la medida que ha establecido un criterio de magnitud mínima para considerar el cambio de puntuación como real. Debido a que los instrumentos de medida utilizados habían mostrado un buen nivel de fiabilidad ( $\alpha > 0,80$ ), la exigencia de cambio del ICF<sub>JT</sub> ha sido moderada aunque más restrictiva que las pruebas de significación habituales. Se ha constatado que la magnitud del cambio de algunos pacientes ha sesgado en gran medida las conclusiones grupales del análisis clásico de promedios (fácilmente apreciable en tablas de resultados del *estudio 1*, dimensiones Ansiedad Fóbica y Somatización). Siendo también afectado en su análisis, por la circunstancia inversa donde la tendencia mayoritaria grupal ha ocultado varios casos que han empeorado de forma relevante (como en el cambio en sintomatología depresiva en el grupo control del *estudio 2*). Mientras que las conclusiones de ICF<sub>JT</sub> no se ha visto sesgadas por estas dos circunstancias.

En el *estudio 2* también se ha utilizado un instrumento de medida de resultados terapéuticos auto informado por el paciente, pero que incluye indicadores funcionalidad directa en su vida cotidiana (relaciones interpersonales y ajuste a rol social). Mientras que los indicadores que contemplan intensidad sintomática se han comportado de forma similar a los estudios 1 y 3, (presentando también las mismas ventajas del método de ICF<sub>JT</sub> sobre el análisis clásico), las dos dimensiones de resultados no directamente clínicas han mostrado unos comportamientos particulares. En el indicador de relaciones interpersonales se ha dado una tendencia de cambio de muy escasa magnitud en prácticamente el 90% de la muestra, por lo que las conclusiones de ambos métodos han sido muy similares, aunque algo más detalladas por parte del ICF<sub>JT</sub> por señalar las pocas excepciones han existido. En cambio ambos métodos han discrepado en gran medida en la dimensión de Rol Social, resultando la variable de evolución más favorable según el análisis clásico de diferencia de medias, y un indicador con muy baja proporción de casos de cambio fiable (8%) según ICF<sub>JT</sub>. En esta circunstancia la baja fiabilidad de la escala ( $\alpha > 0,70$ ) ha generado que el ICF<sub>JT</sub> sea muy exigente con la magnitud del cambio para considerar el cambio como tal. Aunque se presenta como una limitación a la hora de discriminar el cambio en el grupo estudiado, el no considerar este elemento implica un sesgo de sobre estimación de cambio que el análisis de significación de promedios no tiene en cuenta.

El *estudio 4* ha planteado como indicador de resultado una medida biometrica de capacidad pulmonar, obtenida mediante prueba de espirometría forzada. Al contrario que las anteriores, la mejoría terapéutica se interpreta como aumento de la magnitud del indicador. Pese a sus particularidad respecto a las variables de resultado de los demás estudios (prueba objetiva, indicador biológico y alta precisión de la medida), su comportamiento con el método ICF<sub>JT</sub> ha sido similar, aunque resultando de particular interés el criterio de significación clínica frente al método de análisis clásico. En esta ocasión la mayor parte de la discrepancia entre ambos métodos, se ha debido más a factores derivados del diseño del estudio que a la variable resultado elegida (ver sub-hipótesis 1.3).

Por todo lo anterior, **se confirma la sub-hipótesis 1.1:** la metodología de ICF<sub>JT</sub> ha resultado de más idóneo para valorar el cambio en las diferentes medidas de resultado contempladas en los estudios que el uso de pruebas de significación de diferencias grupales.

*Sub-Hipótesis 1.2:* El ICF<sub>JT</sub> será más útil en la evaluación de resultados de intervenciones en *distintas problemáticas de salud* (ansiedad, depresión y asma) que el procedimiento habitual basado en análisis promedios grupales.

Los tres primeros estudios han implicado pacientes con problemáticas psicopatológicas con predominio de la sintomatología ansiosa o depresiva de curso *agudo*. Diferenciándose el *estudio 2* por su planteamiento de evaluación más genérico (mayor heterogeneidad de diagnósticos, medida de resultado menos específica y foco sobre el desempeño de un centro en vez de en una intervención de un solo profesional). En la valoración de la evolución de las problemáticas de cada estudio, el método clásico se ha mostrado muy optimista respecto a los efectos de la intervención, señalando mejorías clínicas de magnitud relevante en los tres estudios. En cambio, con ICF<sub>JT</sub>, se han matizado estos resultados especialmente por el segundo criterio de significación clínica (punto de corte con funcionalidad), lo que ha generado un valor añadido a la hora de valorar la importancia real del efecto conseguido, oscilando la proporción de casos que se han considerado *recuperados* entre un 30% y un 60% de la muestra en los indicadores de los *estudios 1 y 3*. Estableciendo, además, una información que es de utilidad inmediata para el terapeuta que con el análisis de promedios no se facilita. Por otro lado, en el *estudio 2* ha existido una discrepancia mucho mayor entre ambos métodos, con un contexto particular donde la magnitud global de la mejoría ha sido menor que en los otros dos estudios. Aunque el análisis de promedios no ha sido tan favorable como en los otros dos estudios (en cuanto a conclusiones de evolución clínica), ha seguido manteniendo unas conclusiones positivas en sentido de mejoría relevante. Mientras que el ICF<sub>JT</sub>, ha descrito una evolución donde solo un tercio de los casos han presentado una *mejoría* fiable (siendo la situación de no cambio la más frecuente), y donde la clasificación de cambio de *empeoramiento* significativo ha sido más frecuente que la consideración de *recuperado* en la mayoría de los indicadores analizados. Por lo que no solo se plantea una situación global mucho más negativa, sino que además subraya agravamientos clínicos relevantes que el método clásico obvia totalmente.

El estudio 4 se ha realizado con pacientes asmáticos. Esto implica una problemática de salud de muy distinta índole a las anteriores, debido a dos razones: la naturaleza nosológica del diagnóstico es biomédica y el curso del trastorno crónico. Observando el comportamiento más conservador del ICF<sub>JT</sub>, en los estudios precedentes, (con respecto a los análisis de significación clásicos) se podría esperar que su exigencia fuese excesiva para los cambios de menor magnitud probables en el contexto de una enfermedad crónica. Sin embargo, la clasificación del cambio resultante ha señalado evoluciones clínicas de todas las categorías del ICF<sub>JT</sub>, incluyendo las que implican el cumplimiento del segundo criterio de cambio clínica significativo (punto de corte con funcionalidad): *Recuperado* (19% del grupo experimental) y *Empeorado en disfuncionalidad* (13% del grupo experimental). Además el ICF<sub>JT</sub>, ha permitido diferenciar la evolución de los dos grupos de estudios (no cambio fiable en grupo experimental 44% versus 68% en grupo control) , caracterizándolos con mucha mayor precisión y riqueza descriptiva que el análisis de significación de diferencias grupales. Resultando este último muy afectado por la coyuntura en la que ha tenido que realizar el análisis de datos, no pudiendo ofrecer más conclusiones que el posible efecto de pertenecer a un grupo o al otro sobre la evolución de los hombres. Y tampoco ha reflejado la disparidad de evoluciones dentro de los grupos, reflejando un cambio grupal mínimo para ambas condiciones.

Teniendo en cuenta lo reflejado en los cuatro estudios, **se confirma la sub-hipótesis 1.2:** la metodología de ICF<sub>JT</sub> ha resultado más útil y precisa para evaluar los resultados de las intervenciones en las diferentes problemáticas de salud estudiadas que el procedimiento habitual de análisis de significación de diferencias grupales.

*Sub-Hipótesis 1.3:* El ICF<sub>JT</sub> resultará más conveniente para la determinación de la *efectividad* de las intervenciones ante *diferentes condiciones de diseño experimental* (pre experimental sin grupo control, cuasi experimental con grupo control no equivalente, y ensayo clínico con grupo control) que el análisis de promedios habitual.

Los *estudios 1 y 2* se han planteado bajo un diseño experimental pragmático sin grupo control, tratando de modificar lo mínimo indispensable el proceder habitual de los profesionales en su centro de trabajo (público o privado). Priorizando así la validez externa de los resultados analizados frente al foco explicativo de un contexto más controlado. En esta situación, el análisis de significación de las diferencias grupales ha maximizado la importancia de los efectos de las intervenciones, señalando en todos los indicadores cambios estadísticamente significativos y de una magnitud relevante. La excepción ha sido el indicador OQ- 45 Interpersonal del *estudio 2*, el cual ha mostrado un tamaño del efecto demasiado pequeño, aunque estadísticamente significativo (debido al tamaño muestral n=100). Al no contar con grupo control el método de análisis habitual, se ha encontrado con la imposibilidad de ejercer ningún tipo de filtro frente a fuentes de variabilidad de la medida que no se deben a la propia intervención. Por otro lado, el ICF<sub>JT</sub> si ha incluido en su análisis la consideración de posibles fuentes de variación ajenas al efecto de la intervención, resultando más conservador a la hora de evaluar los resultados (algo explicitado en el contraste de las hipótesis anteriores, y de bastante relevancia por ejemplo en la variable Rol Social del *estudio 2*). En estos dos estudios para la determinación colectiva de la efectividad de los tratamientos realizados, el ICF<sub>JT</sub> ha sido mucho menos indiscriminado y ajustado a la realidad de las evoluciones clínicas que el método clásico (ver las comparaciones del contraste de la 1<sup>o</sup> *hipótesis general* ).

El *estudio 3* ha vuelto a plantear la evaluación de resultados de una intervención en un contexto de practica real, pero añadiendo un grupo control no equivalente. Aunque se ha planteado equivalente en sexo, edad y contexto geográfico, no lo ha sido en su condición clínica al tratarse de población sin diagnostico psicopatológico. Por ello, su utilidad se ha limitado a

controlar el cambio espontaneo del indicador de medida. En estas circunstancias, el análisis mediante pruebas de significación clásicas ha señalado un cambio muy relevante en la medida de resultado en sentido de mejoría ( $p < 0,001$  y  $\eta^2 = 0,65$ ), con una gran importancia de la pertenencia a uno los grupos de estudio ( $p < 0,001$  y  $\eta^2 = 0,64$ ), y una enorme diferencia en la magnitud de la medida entre ambos grupos ( $p < 0,001$  y  $\eta^2 = 0,74$ ). Caracterizando la evolución grupo control como de no cambio, frente al grupo experimental que plantea un claro cambio a mejor entre ambos momentos de medida. Aunque el ICF<sub>JT</sub>, establece unas conclusiones cercanas al análisis clásico (debido lo extremo de los resultados analizados) ha mostrado dos ventajas sobre este a la hora de valorar de la efectividad de la intervención: ha matizado la relevancia del efecto observado a través de la significación clínica (46% del grupo clínico clasificado como *recuperado*) y ha realizado una caracterización más precisa de la evolución de grupos de estudio (5% del grupo control realiza un cambio fiable a peor).

En el *estudio 4* se ha desarrollado un ensayo clínico con unas condiciones experimentales más controladas que en los anteriores estudios, con generación de grupo control clínico y asignación aleatoria a una de las dos condiciones experimentales. Sin embargo, a pesar de que el grupo control ha sido equivalente en las demás variables de interés al grupo tratado, la distribución de mujeres y hombres ha sido significativamente desigual entre ambos ( $p < 0,05$ ). Esto ha provocado que haya sido necesario generar grupos de análisis de resultados separados por sexo, con el fin de evitar este sesgo relevante en la medida de resultado. Esta situación ha afectado en gran medida al análisis grupal clásico, ya que los estadísticos utilizados se han visto muy afectados por el reducido tamaño de los grupos. No detectando diferencias significativas entre los grupos, ni entre los dos momentos de medida ( $p > 0,05$ ), y únicamente señalando para los hombres un efecto de interacción de pertenecer a uno de los grupos sobre le evolución del indicador ( $F_{(1\ 35)} = 5,47$ ;  $p = 0,025$ ). Por otro lado, ICF<sub>JT</sub>, no solo ha verificado una situación de cambio real entre los dos momentos de medida (más de un 50 % del grupo experimental efectúa un cambio fiable) sino que además a diferenciado un patrón diferencial de evolución clínica entre grupo control y grupo experimental en ambos sexos (ver contraste de *1º hipótesis general*). Al no ser relevante el tamaño muestral para ICF<sub>JT</sub> se ha podido generar un análisis útil para valorar los efectos del programa con grupos tan reducidos.

Por todo lo anterior, **se confirma la sub-hipótesis 1.3:** la metodología de ICF<sub>JT</sub> ha resultado más conveniente para la determinación de la efectividad de las intervenciones realizadas, ante diferentes condiciones de diseño experimental, que los análisis clásicos de significación de diferencias grupales.

En relación a esta primera hipótesis (y a sus sub hipótesis), se ha podido observar que en los tres primeros estudios se cumple la tendencia, recurrentemente señalada en la literatura, de que el análisis individual realiza una interpretación menos favorable de los resultados en psicoterapia que el análisis grupal de estos (Jacobson y Truax 1991; Beretta et al. 2005; Lambert y Ogles, 2009). Esta circunstancia ha confirmado la diferenciación teórica entre significación práctica (informada por los tamaños del efecto grupales) y la significación clínica (importancia del cambio acontecido en la evolución de cada paciente), y como la primera no llega a reflejar los efectos reales de las intervenciones (Kazdin, 2001; Iraurgi, 2010; Preacher y Kelly, 2012; Gómez y Roussos, 2012). Los tamaños del efecto de los indicadores resultantes en los estudios han sido similares a los referidos a los estudios meta-analíticos sobre eficacia de la psicoterapia, oscilando la interpretación de su magnitud entre moderada y grande (Chambless y Ollendick, 2001; Haby, Donnelly, Corry, y Vos, 2006 Shedler, 2010). Sin embargo, al analizar la relevancia clínica de la evolución de cada paciente, la magnitud del efecto grupal se ha visto relativizada, tal y como ha sucedido en otras investigaciones (Bauer, Lambert y Nielsen, 2004; Hiller, Schindler y Lambert, 2012) por dos factores fundamentalmente: la consideración de magnitud de los cambios individuales de mejoría (si son fiables y si se acercan a la funcionalidad, representando la discrepancia principal en los *estudios 1 y 3*) y la dirección de los cambios individuales (pacientes que de hecho empeoran significativamente dentro de una tendencia grupal que no lo hace como sucede en el *estudio 2*).

Sobre la idoneidad del método ICF<sub>JT</sub> ante diferentes circunstancias clínicas, se ha puesto en duda su adecuación fuera de las problemáticas psicológicas de curso agudo, señalando que las magnitudes de cambio necesarias para cumplir los dos criterios del método son excesivas en trastornos de curso crónico y que en ciertas circunstancias cambios menores deberían considerarse clínicamente relevantes (Kazdin 1999; Wise, 2004). Sin embargo, en el *estudio 4* las clasificaciones de cambio se han establecido de forma equivalente a los anteriores estudios (y

con pacientes clasificados en todas las categorías de cambio posibles), e incluso ha resultado menos restrictivo con los resultados globales que el propio análisis de promedios grupales. También se ha criticado al ICF<sub>JT</sub>, por establecer criterios dicotómicos de mejoría demasiado rígidos como para capturar el curso clínico de cada paciente (Tengey, 1996; Hiller y Schindler 2011). Pero el reporte de resultados ICF<sub>JT</sub> de los cuatro estudios, ha permitido mostrar claramente (especialmente en su representación gráfica) las diferentes evoluciones de los participantes, discriminando cambios relevantes incluso en pacientes de poca gravedad clínica inicial (como en el *estudio 1*) y diferenciando varios tipos de evolución dentro de un misma categoría de clasificación (como se ejemplariza en la discusión del *estudio 4*).

En cuanto a los distintos indicadores de resultado utilizados, se ha vuelto a constatar lo observado en anteriores estudios con ICF<sub>JT</sub>: las medidas de resultado de diferente naturaleza (sintomáticos vs no sintomáticas) han presentado mayor discrepancia entre sí que los propios métodos de análisis utilizados (Beckstead, 2003; Ronk, Hooke y Page, 2012). Algo que se puede constatar fácilmente comparando la congruencia dentro de cada paciente entre medidas del *estudio 1* (todas sintomáticas), con la de los indicadores del *estudio 2* (sintomáticas y de funcionalidad social). Se evidencia por tanto la necesidad de contar con diversos indicadores además de los directamente clínicos para poder valorar el impacto real de la intervención tal y como señala Kazdin (1999; 2001). Sin embargo, también se ha puesto de manifiesto en el *estudio 2* lo mucho que afecta al comportamiento del método una baja fiabilidad en la medida utilizada, y resultando las escalas con menor fiabilidad las que no eran de tipo sintomático. Coyuntura que ha coincidido con una de las principales dificultades para el uso del ICF<sub>JT</sub> detectadas por Lambert, Hasen y Bauer (2008): la mayoría de las medidas de resultado clínicas que han demostrado alta fiabilidad, validez y cuentan con datos normativos de referencia son las de tipo sintomático, mientras que los referentes a otro tipo de áreas de resultados son escasos cumpliendo estas características.

Por último, los resultados de los estudios han puesto de manifiesto una de las principales limitaciones señaladas respecto a los análisis estadísticos clásicos y al paradigma de contraste de hipótesis nula: su gran sensibilidad al tamaño muestral (Levine, Weber, Hullett, Park y Lindsey 2008). Lo que ha llevado resaltar como significativas diferencias mínimas en los promedios

(*estudio 2*) o a obviar diferencias grupales cuando el número de sujetos de cada grupo de análisis ha sido pequeño (*estudio 4*). Por otro lado, el ICF<sub>JT</sub> ha dado respuesta a una complicación metodológica recurrente detectada en el área particular del *estudio 4* (adhesión al tratamiento en enfermedades crónicas) precisamente relacionada con los tamaños muestrales de los estudios de efectividad de programas (Haynes, Ackloo, Sahota, McDonald y Yao, 2008).

*2º Hipótesis:* La metodología ICF<sub>JT</sub> resultará *más adecuada* para la evaluación del cambio y su significación clínica que *otras alternativas metodológicas de análisis de resultados individuales* (GLN, EN, HA, TEI, DIT, ESG, DTW, IPP, IPI y CPM 50%).

En el *estudio 3* es donde se ha realizado la comparativa del ICF<sub>JT</sub> con los métodos alternativos seleccionados. Aunque se ha tratado de un mismo estudio la existencia de dos grupos experimentales han permitido realizar el contraste entre métodos en dos situaciones muy diferentes: de cambio generalizado (grupo clínico) y de no cambio mayoritario (grupo control).

En los métodos IPP y HA se han generado clasificaciones de cambio totalmente anómalas en ambos grupos (señalando cambio cuando se mantenía la misma puntuación en los dos momentos o explicitándolo en sentido inverso del cambio real) por lo que según los datos de este estudio no se pueden considerar alternativas factibles al ICF<sub>JT</sub>.

Los métodos EN y GLN se han comportado de manera similar a ICF<sub>JT</sub> en la clasificación del grupo clínico ( $Kappa > 0,80$ ), en el cual se esperaría un mayor influjo del efecto de regresión a la media al mostrar unas puntuaciones pre tratamiento tan elevadas (justificación de las alternativas EN y GLN). Sin bien, la clasificación no ha sido idéntica resultando algo más restrictiva para GLN y más permisiva para EN. En cambio, para el grupo control el acuerdo ha sido total entre GLN y ICF<sub>JT</sub> mientras que EN se ha distanciado bastante de estos ( $Kappa = 0,40$ ) al generar una clasificación de cambio mucho más permisiva en un grupo donde la magnitud general del cambio ha sido reducida. Y el método DTW también basado en el error estándar de medida ha tenido un comportamiento cercano a EN, aunque resultando incluso más laxo en la clasificación del grupo clínico (84% de cambios fiables).

Los métodos basados en la tipificación de las diferencias individuales (TEI, ESG y DIT) se han mostrado demasiado poco exigentes para clasificar cambio para los dos grupos en el umbral 0,20, llegando a identificar relevante cualquier mínimo cambio en la medida de resultado. Sin embargo, en el umbral 1,96 ESG y DIT han generado clasificaciones tanto en grupo clínico y como grupo no clínico con buena concordancia con  $ICF_{JT}$  ( $Kappa > 0,75$ ), llegando para el grupo clínico a tener un nivel de acuerdo muy alto ( $ICF_{JT}$ -ESG 1,96  $Kappa=0,94$  y prueba *Wilcoxon* no significativa). Por otro lado, el método TEI en el umbral 1,96 se ha mostrado como el más exigente a la hora de clasificar cambio de todos los estudiados. Este discrepa con  $ICF_{JT}$  y con todas las demás alternativas, especialmente en la clasificación del grupo clínico ( $Kappa < 0,30$ ), exigiendo un cambio de más 15 puntos en la puntuación del BDI para ser considerado como tal, resultando en exceso restrictivo.

El Criterio de Porcentaje de Mejoría 50%, se ha mostrado muy de acuerdo con la clasificación  $ICF_{JT}$  del grupo clínico (prueba *Wilcoxon* no significativa y  $Kappa < 0,91$ ) aunque ha resultado totalmente discordante para el grupo control. Adicionalmente a la limitación del método CPM 50% ya conocida de antemano (solo clasifica cambio en sentido de mejoría), se ha observado un comportamiento mucho menos restrictivo que  $ICF_{JT}$  para considerar cambio significativo (en grupo control 27% vs 6 % proporción de cambio significativo), llevándole a acercarse a la permisividad del umbral 0,20 de los métodos TEI, ESG y DIT. Se limita la idoneidad del método únicamente en condiciones de alta gravedad clínica inicial, con mucho margen de mejoría posible.

El método IPI ha establecido un criterio de identificación de cambio para el grupo clínico no muy alejado de los métodos basados en el error estándar de medida ( $Kappa > 0,61$ ) aunque visiblemente más restrictivo que todos ellos. A pesar de que ha resultado el segundo método más exigente de los analizados (con la magnitud del cambio necesaria para ser considerado como tal), no ha resultado tan desmesurado como el TEI 1,96. Y aunque se ha mostrado como un método más conservador que  $ICF_{JT}$ , la clasificación ha resultado plausible con la evolución de los grupos del estudio, y por otro lado ambos métodos han coincidido totalmente a la hora de clasificar el grupo control ( $Kappa=1$ ). Se ha evidenciado como un método que con respecto a  $ICF_{JT}$  aumenta su sensibilidad a costa de una menor especificidad. Su idoneidad por tanto vendrá supeditada por

las prioridades y objetivos de análisis de resultados (evitar falsos positivos o evitar falsos negativos).

Todo lo anterior nos lleva a **confirmar parcialmente la 2º hipótesis**: el ICF<sub>JT</sub> ha resultado indudablemente más adecuado que HA, IPP, TEI, DIT 0,20 y ESG 0,20. Con respecto a los demás métodos no ha discrepado en exceso en sus clasificaciones alcanzando niveles de concordancia altos con mucho de estos, y resultado ser moderado en su exigencia para clasificar cambio (existiendo métodos más permisivos y también más restrictivos). Teniendo en cuenta esto y al mostrarse como uno de los métodos computacionalmente sencillos de calcular y con menos requisitos previos necesarios, se puede confirmar la idoneidad del método para la evaluación de resultados clínicos en contextos de practica real. Sin embargo, es arriesgado concluir superioridad del ICF<sub>JT</sub> con respecto a todos las demás métodos que también han tenido un comportamiento adecuado. Dada la variación nivel de exigencia y la diferente naturaleza de los criterios aplicados, sería más idónea la combinación del ICF<sub>JT</sub> con otro de estos métodos, que su uso en solitario. Pudiendo aportar estas ventajas o perspectivas adicionales cuando estas fuesen interesante para los objetivos del análisis de resultados (como la precisión del IPP o la consideración de la gravedad clínica inicial de CP50% )

En relación a esta comparativa, lo primero que se puede contrastar es el hecho de que el efecto de regresión a la media no ha tenido el gran impacto que teóricamente le conferían Speer (1992) y Hsu (1995) en su crítica del ICF<sub>JT</sub>. En sentido, las conclusiones del presente estudio se muestran en coherencia con el consenso actual sobre este punto (Lambert, Hansen y Bauer, 2008; Iraugui, 2010). Sin embargo, los resultados no han mostrado la misma equivalencia práctica entre métodos ICF<sub>JT</sub> y GLN que se había observado en estudios anteriores (Bauer, Lambert y Nielsen, 2004; Atkins, Bedics, McGlinchey y Beauchaine, 2005), resultando con un alto nivel de congruencia pero significativamente diferentes en el grupo clínico. Otra discrepancia con la evidencia previa ha sido el funcionamiento anómalo de la propuesta de Hagerman y Arrindal (1999a), que en las condiciones particulares de esta investigación no solo no se ha planteado tan restrictivo como en estudios anteriores (McGlichey, Atkins y Jacobson, 2002; Ronk, Hooke y Page, 2012), sino que ha resultado totalmente indiscriminado para señalar cambio.

Por otro lado, la excesiva laxitud encontrada en el umbral 0,20 para los métodos de tipificación de diferencias individuales, y la obtención de numerosos falsos positivos con el método de Intervalo para el Pronóstico Promedio (Crawford y Garthwaite, 2006), ha coincidido con el análisis acerca de su comportamiento de Pardo y Ferrer (2013). También en coherencia con ese mismo estudio, se ha observado un mayor nivel de exigencia y sensibilidad para el cambio del Intervalo para el Pronóstico Individual (Crawford y Garthwaite, 2006; Pardo y San Martín, 2010) y la Diferencia Individual Tipificada (Guyatt, Bombardier y Tugwell, 1986 ) en el umbral 1,96, en comparación al  $ICF_{JT}$ . Con respecto a la relación entre  $ICF_{JT}$  y el Criterio de Porcentaje de Mejoría 50% (Hiller y Schindler, 2011), se han alcanzado similares conclusiones que en el estudio de Hiller Schindler y Lambert (2012), señalando una menor exigencia general del CPM 50% para clasificar cambio pero mayor en los casos de alta gravedad clínica inicial, y resaltando la naturaleza complementaria de ambos criterios.

Sobre el comportamiento del  $ICF_{JT}$  respecto a las demás alternativas metodológicas, los resultados han apoyado la conclusión general de anteriores revisiones acerca de un nivel de exigencia moderado, y una relativa facilidad de computación (Lambert, Hasen y Bauer 2008; Lambert y Ogles, 2009). Sin embargo, al contrario que dichas revisiones, no se ha concluido la preferencia del método en solitario frente a los demás métodos de análisis individual. Dado la distinta naturaleza de los alternativas metodológicas, la complementariedad entre índices puede ser clave para adaptarse a las distintos contextos clínicos (Hiller, Schindler y Lambert, 2012). Siendo la características y prioridades de estos contextos determinantes para seleccionar estos métodos (Jacobson, Robers, Berns y McGlinchey, 1999; Gómez y Roussos, 2012). Y aunque se ha evidenciado la flexibilidad y adecuación del método  $ICF_{JT}$  en los cuatro estudios, se ha contrastado en la literatura situaciones donde claramente se limita su utilidad clínica como único indicador de relevancia clínica, tales como cursos de enfermedad degenerativa donde el cambio es de hecho significativo (Kazdin, 1999) o en indicadores de resultado donde influye el efecto de la practica al medir competencia o habilidad con el mismo instrumento varias veces (Lewis et al. 2007)

## 12.2. Conclusiones generales

Derivado de lo expuesto en el anterior apartado se pueden concretar una serie de conclusiones finales:

- Atendiendo los resultados de los cuatro estudios, se ha comprobado que el método de Jacobson y Truax (1991),  $ICF_{JT}$ , ha sido de mayor utilidad y más ajustado a la realidad (en el análisis de resultados en situaciones natural de práctica clínica) que el análisis clásico de significación de diferencias promedios (paradigma de inferencia estadística mediante contraste de hipótesis nula) y consideración de tamaños del efecto grupales. Esta superioridad se ha mostrado tanto desde el punto de vista de la determinación grupal de efectividad de las intervenciones, como de la utilidad para la toma de decisiones profesional respecto a cada individuo.
- Se ha confirmado la adecuación del  $ICF_{JT}$  al análisis de evolución clínica en diferentes circunstancias clínicas. No viéndose afectado en su comportamiento por el uso en patologías crónicas o medidas de resultado no directamente sintomáticas o clínicas. Aunque se ha hecho evidente lo sensible que es el método a instrumentos con poca fiabilidad, especialmente si esta es inferior a 0,70. Por otro lado, el método de análisis clásico si ha visto limitada y mermada su utilidad en las diferentes condiciones estudiadas, comprobándose su gran sensibilidad al tamaño muestral y su incapacidad para reflejar los cambios reales acontecidos en los grupos analizados, a menos que todos los participantes siguiesen la tendencia mayoritaria.
- El  $ICF_{JT}$  ha mostrado un comportamiento de clasificación de cambio claramente más adecuado que los siguientes métodos alternativos de análisis individual: Intervalo para Pronóstico Promedio (IPP), Tamaño del Efecto Individual (TEI), y método Hageman-Arrindell (HA). Esto ha sido debido al comportamiento anómalo o demasiado extremo de estos en la clasificación de cambio de los dos grupos analizados. Es destacable lo indiscriminado que ha resultado el método HA en el estudio, ya que no hay referente de investigaciones anteriores donde haya sucedido algo parecido (anteriormente había sido

señalado como de los métodos más restrictivos y exigentes con la magnitud del cambio). También se han observado métodos con un comportamiento inadecuado por ser excesivamente permisivos: Diferencia Individual Tipificada (DIT), Estadístico de Sensibilidad de Guyatt (ESG). Aunque estos últimos únicamente en el umbral 0,20 de significación.

- Al igual que en investigaciones anteriores, se ha comprobado que el posible efecto de regresión a la media no plantea una gran distorsión real en el análisis de  $ICF_{JT}$ . Esto se ha evidenciado por el alto nivel de concordancia de clasificación de cambio con los métodos centrados en corregir precisamente este sesgo: Edward-Nunnally (EN) y Gullisken Lord-Novick (GLN). Se muestra la paradoja de que el método EN resulta menos restrictivo para clasificar cambio que  $ICF_{JT}$  en las dos situaciones de cambio analizadas. A pesar de ello, es necesario matizar que la equivalencia del  $ICF_{JT}$  con estos métodos ha resultado ser tan alta como en estudios anteriores.
- El  $ICF_{JT}$  ha manifestado un comportamiento moderado con respecto a los métodos alternativos de análisis individual que no han resultado inadecuados en su clasificación. Situándose a medio camino entre los más exigentes con la magnitud de cambio (Gullisken Lord-Novick, Diferencia Individual Tipificada  $\geq 1,96$  e Intervalo para Pronostico Individual) y los menos exigentes (Edward-Nunnally, Estadístico de sensibilidad Guyatt  $\geq 1,96$ , Diferencia Tipificada de Wyrwich y Criterio Porcentaje de Mejoría 50%). Además ha planteado un destacado nivel de acuerdo con todos ellos lo que refuerza la designación de anteriores investigaciones como método de elección. Sin embargo, a la luz del aceptable funcionamiento de otras alternativas metodológicas con distintos criterios de cambio, se concluye lo idóneo de su complementariedad con índices de análisis individual de orden lógico diferente seleccionado según las características y prioridades del contexto clínico donde se utilice.

### 12.3. Limitaciones

En el curso de la investigación, se han hecho evidentes una serie de limitaciones o problemáticas que es necesario explicitar para relativizar las conclusiones obtenidas:

- En los *estudios 3 y 4* no se han dado las condiciones idóneas para el uso de la prueba de significación paramétrica de Análisis de la Varianza por no cumplirse los supuestos de normalidad e igualdad de varianzas. Mientras que en el *estudio 3* esto se ve atenuado por el hecho de contar con grupos de análisis de gran tamaño y el mismo número de sujetos por grupo, en el *estudio 4* no se ha contado con estas circunstancias. Aunque se había planificado un número de sujetos por grupo superior a 30 en los dos grupos, la desigualdad en la distribución de hombres y mujeres entre grupo experimental y control ha obligado a generar sub-grupos de análisis mucho más pequeños. Se tomó la decisión de seguir realizando la prueba paramétrica por su mayor potencia y por la imposibilidad de constatar la significación de los efectos de interés mediante alternativas no paramétricas.
- El instrumento psicométrico utilizado en el *estudio 2* (OQ-45) se eligió por contar con una orientación mixta en sus escalas (teniendo en cuenta la evolución clínica en sentido sintomático pero también el impacto en su funcionalidad social general) y haber sido destacado en la literatura científica internacional acerca de resultados de psicoterapia. Sin embargo, el instrumento aún no cuenta con una validación en la población española (aunque sí en otros países de habla hispana como Chile), por lo que se decidió aprovechar una versión del cuestionario que en estos momentos está en proceso de validación local (ver Anexo 2), y obtener mediante estrategias meta analíticas unos valores de referencia con los datos de otras validaciones en países occidentales.
- Para medir el cambio clínico se han tomado únicamente dos momentos de medida en todos los estudios (antes y después de la intervención), ya que no ha sido factible obtener más medidas debido al tiempo limitado y a tener que adecuarse al desempeño clínico habitual. Esto ha provocado que no se hayan podido contemplar ciertas alternativas

metodológicas de análisis individual como el Modelado Lineal Jerárquico (HLM), y se haya perdido la oportunidad de añadir mayor riqueza a las observaciones sobre los análisis de resultados, con parámetros tales como la estabilidad del cambio en cada paciente.

- El planteamiento del presente trabajo ha tratado de responder a la demanda de mayor respaldo empírico del funcionamiento del método ICF<sub>JT</sub>, ante diferentes circunstancias clínicas e indicadores de resultado. Sin embargo, el ámbito de los tres primeros estudios del trabajo se han centrado en la psicoterapia en trastornos de ansiedad y del ánimo, con indicadores de resultado en su mayoría de tipo sintomático. Aunque se han planteado contextos clínicos diferentes (*estudio 4*) y medidas de resultado de funcionalidad general (*estudio 2*), no se han incluido medidas de resultado importantes en el ámbito de la salud general tales como Calidad de Vida, ni indicadores de resultado relacionados con la recuperación de capacidad cognitiva, ni problemáticas de salud mental crónicas como los trastornos psicóticos. Todo ello, resta capacidad de generalización de las conclusiones a los diferentes ámbitos de intervención psicológica posibles.

#### **12.4. Implicaciones, aplicaciones y futuras líneas de investigación**

El presente trabajo se enmarca en la línea de investigación sobre metodologías de análisis de resultados vinculadas a la relevancia clínica de las intervenciones psicológicas. Focalizándose sobre la propuesta de análisis de cambio individual de Jacobson y Truax (1991) aunque también contrastando su funcionamiento con otras metodologías que tienen el mismo objetivo y similar naturaleza. Se ha recogido la demanda de la literatura previa, de generar evidencia en diferentes contextos reales sobre la idoneidad de este tipo de análisis (Jacobson, Roberts, Berns y McGlinchey, 1999; Lambert y Ogles 2009; Gómez y Roussos, 2012).

En este sentido, el presente trabajo ha aportado varias verificaciones empíricas del comportamiento del método ICF<sub>JT</sub> en condiciones de práctica real comprobando las posibilidades y ventajas de su aplicación. Se ha realizado en contextos psicoterapéuticos que reproducen las condiciones de anteriores estudios (Lambert, Hansen y Bauer 2008), pero también con aplicaciones innovadoras como su uso con una medida biométrica en el curso de un trastorno médico crónico (generándose además de este estudio una publicación en revista científica).

Por otro lado, se ha realizado el contraste del ICF<sub>JT</sub> con el método más ampliamente utilizado en la investigación de resultados terapéuticos: el basado en la significación estadística de diferencias grupales promedio. Las distintas circunstancias evaluadas han mostrado con datos reales, las limitaciones que presenta el paradigma de contraste de hipótesis nula para los fines particulares de esta área de investigación, incluso añadiéndole la consideración del tamaño del efecto. Dando más soporte empírico a las observaciones y críticas, sobre esta forma de afrontar el análisis de resultados, generadas en los últimos años (Kazdin, 2001, Iraurgi, 2010, Lewis et al. 2007; Nuzzo, 2014).

Además de los dos elementos anteriores, uno de los objetivos de esta tesis ha sido recopilar y mostrar las principales propuestas para el análisis de relevancia clínica de cambio individual en el contexto de la psicología clínica. Se entiende que el valor añadido de este tipo de métodos no solo es de interés para el investigador (en el estudio de la efectividad general de las intervenciones), sino que es de uso práctico para el profesional clínico en su día a día, facilitando la necesaria integración de ambos ámbitos (Kazdin, 2008; Beutler, 2009; Castoguy, 2011; Henton, 2012). Sin embargo, el clínico no vinculado con el mundo académico puede tener dificultades para comprender y poner en práctica estas alternativas metodológicas (dada su relativa dificultad computacional), por lo que además de resumir y exponer estos métodos en la revisión teórica, se han adjuntado los algoritmos desarrollados (sintaxis de SPSS Anexo 1), para automatizar estos procesos en el *estudio 2*, además de ejemplarizar la posibilidad de presentar los resultados de forma gráfica facilitando su interpretación. Consideramos esta aportación de particular interés para la divulgación y la implementación de este tipo de metodologías a contextos de práctica profesional real.

También se han contrastado el funcionamiento y la adecuación de estas distintas alternativas de análisis individual ante dos situaciones posibles de cambio (gran mejoría mayoritaria y estabilidad o poco cambio). Lo que ha servido para continuar la línea de estudios anteriores al respecto (Bauer, Lambert y Nielsen 2004; Ronk, Hooke y Page, 2012; Hiller, Schindler y Lambert, 2012; Pardo y Ferrer, 2013), aumentando la evidencia sobre las conclusiones alcanzadas en estos (tales como el reducido impacto del efecto de regresión a la media, la moderación de la exigencia del ICF<sub>JT</sub> o la inadecuación de ciertos métodos) e incluso mostrando nuevas coyunturas no detectadas anteriormente (como el funcionamiento anómalo del método Hageman-Arrindell).

Finalmente, teniendo en cuenta los desarrollos realizados, las conclusiones alcanzadas y también las limitaciones que se han evidenciado, se proponen las siguientes líneas futuras de investigación y desarrollo:

- A partir de los algoritmos presentados en el Anexo 1, desarrollar una herramienta informática o una MACRO de SPSS con un interfaz más amable para el usuario y fácil de utilizar. Incluyendo en su programación los valores de referencia de los principales instrumentos psicométricos de resultados terapéuticos que dispongan de adecuada validación nacional. De manera que el clínico solo tenga que seleccionar el instrumento, el tipo de método que quiera utilizar e introducir las puntuaciones de los diferentes momentos de medida, configurándose el resto de parámetros de forma automática. Y cuya salida de resultados pueda venir apoyada por la representación gráfica del cambio acontecido para facilitar su interpretación.
- Estudiar de forma específica la idoneidad del ICF<sub>JT</sub> y las demás alternativas de análisis de cambio individual en contextos de trastorno mental de curso crónico o de enfermedad mental degenerativa.

- Continuar investigando la adecuación del ICF<sub>JT</sub> para medir el cambio relevante en indicadores de resultado que van más allá de la reducción sintomática u otras variables directamente clínicas, pero que sí tienen importancia para el individuo (tales como funcionalidad en vida diaria o calidad de vida).
- Relacionado con lo anterior, desarrollar un mayor número de instrumentos de medida, que cubran estas otro tipo de dimensiones de resultado clínicos, generando datos de referencia normativos, y demostrando fiabilidad y validez de la herramienta.
- Contrastar el funcionamiento ICF<sub>JT</sub> (y de otros métodos similares) ante medidas de resultado que puedan ser influidas por el efecto de la práctica (tales como indicadores de capacidad cognitiva basada en resolución de tareas). Y si se muestra necesario generar nuevos desarrollos metodológicos para dar respuesta este tipo de situaciones.
- Replicar el contraste entre ICF<sub>JT</sub> y análisis de significación de diferencias grupales en condiciones de máximo control experimental e idoneidad total para el uso de pruebas paramétricas. Para comprobar la posible aportación del ICF<sub>JT</sub> a contextos experimentales de estudio controlados de eficacia.

## **BIBLIGRAFÍA Y ANEXOS**

---



## Bibliografía

---

- Abraira, V. y Pérez de Vargas, A. (1996). *Métodos Multivariantes en Bioestadística*. Madrid: Ediciones Centro de Estudios Ramón Areces.
- Abramson, L. Y., Seligman, M. E. y Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 87(1), 49-74.
- Abt, L. y Bellak, L. (1967). *Psicología Proyectiva*. Buenos Aires: Paidós.
- Ackerman, S. J., Benjamin, L. S., Beutler, L. E., Gelso, C. J., Goldfried, M.R., Hill, C., et al. (2001). Empirically supported therapy relationships: Conclusions and recommendations of the Division 29 Task Force. *Psychotherapy*, 38(4), 495–497.
- Adarraga, P., y Zaccagnini, J. L. (1992). DAI: A knowledge-based system for diagnosing autism. *European Journal of Psychological Assessment*, 8(1), 25-46.
- Addis, M. E. (1997) Evaluating the treatment manual as a means of disseminating empirically validated psychotherapies. *Clinical psychology: Science and Practice*, 4(1), 1-11.
- Addis, M. E., Wade, W. A. y Hatgis, C. (1999). Barriers to dissemination of Evidence-Based Practices: Addressing practitioners' concerns about manual-Based psychotherapies, *Clinical Psychology: Science and Practice*, 6(4), 6430–6441.
- Adefna, R., Leal, A., Izquierdo, F.T., Castellanos, J.A. y Arocha, R. (2003) De la medicina basada en la opinión a la medicina basada en la evidencia. *Revista Cubana de Cirugía*, 42(1), [en línea], (fecha de consulta: 13 Septiembre 2013). Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0034-74932003000100008](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0034-74932003000100008)
- Adkins, D. (1994) *Elaboración de tests: Desarrollo e interpretación de los tests de aprovachamiento*. México: Trillas.
- Agras W. E. y Berkowitz, R. (1980). Clinical research in behavior therapy: Halfway there?. *Behavior Therapy*, 11(4), 472-487.
- Agudelo, D., Buela-Casal, G. y Spielberger, G. (2007). Ansiedad y depresión; el problema de la diferenciación a través de los síntomas. *Salud Mental*, 30(2), 33-41.

- Ahmedani, B. K. Peterson, E. L., Wells K. E. y Williams, L. K. (2013). Examining the relationship between depression and asthma exacerbations in a prospective follow-up study. *Psychosomatic Medicine*, 75(3), 305-310.
- Aiken, L. R. (1996). *Tests Psicológicos y Evaluación. Octava edición*. Mexico: Prentice Hall Hispano Americana.
- Alamilla-López, E. y Jiménez, J. C. (2010). Contraste de Hipótesis: Clásico vs Bayesiano. *Revista digital Matemática, Educación e Internet*, 11(1), [en línea], (fecha de consulta: 20 Agosto 2014). Disponible en: [http://tecdigital.tec.ac.cr/revistamatematica/ARTICULOS\\_V11\\_N1\\_2010/NAlamilla\\_ConstrastedeHipotesis/1\\_NAlamilla\\_JJimenez\\_Constraste%20de%20hipotesis.pdf](http://tecdigital.tec.ac.cr/revistamatematica/ARTICULOS_V11_N1_2010/NAlamilla_ConstrastedeHipotesis/1_NAlamilla_JJimenez_Constraste%20de%20hipotesis.pdf)
- Alcántara, G. (2008) La definición de salud de la Organización Mundial de la Salud y la interdisciplinariedad. *Sapiens*, 9(1), 93-107.
- Alonso, J. (2004). *Evaluación psicológica. Coordinadas, procesos y garantías*. Madrid: Ediciones Universidad Autónoma de Madrid.
- Álvarez, L., Arce, J. y De Diuk, L. (1994). Las técnicas proyectivas. En M. Casullo, A. Cayssials, L. Liporace, M. Arce y L. Álvarez (Eds.) *Proyecto de vida y Decisión Vocacional* (pp. 284-309). Buenos Aires : Paidós.
- Álvarez, R. (2007). *Estadística aplicada a las ciencias de la salud*. Madrid: Ediciones Díaz de Santos.
- Álvarez-Cienfuegos, A., Montañó, A., Baquero, M.J., Rubio, P.L., Candela, E., Gavilán E., Gómez, M.C. y Sánchez, G.A. (2012). ¿Es clínicamente relevante además de estadísticamente significativo? *Boletín Terapéutico Extremeño*, 3(3), [en línea], (fecha de consulta: 20 Septiembre 2014). Disponible en: <http://evalmed.es>
- Ambrosini, P., Bennett, D., Cleland, C., y Haslam, N. (2002). Taxonicity of adolescent melancholia: A categorical or dimensional construct? *Journal of Psychiatric Research*, 36(4), 247-256.
- American Medical Association (1997) *Outcomes Research Resource Guide: A Survey of Current Activities*. Chicago: American Medical Association.
- American Psychiatric Association APA<sub>2</sub> (2002). *Manual diagnóstico y estadístico de los trastornos mentales, cuarta edición texto revisad, DSM IV-TR*. Barcelona: Masson.

- American Psychiatric Association APA<sub>2</sub> (2013). *Diagnostic and statistical manual of mental disorders*. Fifth edition. DSM-V. Arlington, Virginia: American Psychiatric Association.
- American Psychological Association APA (1999). *Standards for educational and psychological tests*. Washington, DC: APA.
- American Psychological Association APA, American Educational Research Association AERA y National Council on Measurement in Education NCME. (1974, 1985, 1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychological Association APA, Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology. *American Psychologist*, 63(9), 839–851.
- American Psychological Association APA (2006). Evidence based practice in psychology. *American Psychologist*, 61(4), 271–285.
- American Psychological Association APA (2010). *Publication manual of the American Psychological Association (6th ed.)*. Washington, DC: American Psychological Association.
- American Psychological Association APA (2012). *Recognition of Psychotherapy Effectiveness*. APA [en línea], (fecha de consulta: 28 Septiembre 2014). Disponible en: <https://www.apa.org/about/policy/resolution-psychotherapy.aspx>
- American Thoracic Society ATS. (1995). Standardization of spirometry: update 1994. *American Journal of Respiratory and Critical Care Medicine*, 152(3), 1107–1136.
- Amigo, I., Fernández, C. y Pérez, M. (2010). *Manual de psicología de la salud*. 3<sup>o</sup> Edición. Madrid: Pirámide.
- Anastasi, A. y Urbina, S. (1998). *Test Psicológicos*. Octava Edición. Mexico: Prentice Hall Hispano Americana.
- Anderson, T. Ogles, B. M., Patterson, C. L., Lambert, M. J., y Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist effects. *Journal of Clinical Psychology*, 65(7), 755-768
- Andreasen, N. C. (1985). Conceptos, diagnóstico y clasificación. En E.S. Paykel (Ed) *Psicopatología de los trastornos afectivos* (pp. 50-82) Madrid: Pirámide.

- Andrews G, Brugha T, Thase M. E, Duffy FF, Rucci P, Slade T. (2007). Dimensionality and the category of major depressive episode. *The International Journal of Methods in Psychiatric Research*, 6(1), 41-51.
- Anguera, M. T. (2003). La observación. En C. Moreno Rosset (Ed.), *Evaluación psicológica. Concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia* (pp. 271-308). Madrid: Sanz y Torres.
- Anguera, M. T. (2010). Posibilidades y relevancia de la observación sistemática por el profesional de la Psicología. *Papeles del Psicólogo*, 31(1), 122-130.
- Anguera, M. T., Blanco, A., Losada, J. L. y Sánchez-Algarra, P. (1999). Análisis de la competencia en la selección de observadores. *Metodología de las Ciencias del Comportamiento*, 1 (1), 95-114.
- Antonovsky, A. (1991). The Structural Sources of Salutogenic Strengths, en C.L. Cooper, & R. Payne, (Eds.), *Personality and Stress: Individual Differences in the Stress Process* (pp. 67-104). Chichester, England: John Wiley & Sons.
- Antonuccio, D.O., Thomas, M. y Danton, W.G. (1997). A cost-effectiveness analysis of cognitive behavior therapy and fluoxetine (Prozac) in the treatment of depression. *Behavior Therapy*, 28(2), 187-210.
- Antony, M. N. y Swinson, R. P. (2000). *Phobic disorders and panic in adults: guide to assessment and treatment*. Washigton: American Psychological Association.
- Arignón, J. M. y Jiménez-Villa, J. (2013) *Métodos de investigación clínica y epidemiológica. 4º Edición*. Barcelona: Elsevier España.
- Arnau J. y Balluerka (2004). Análisis de datos longitudinales y de curvas de crecimiento. Enfoque clásico y propuestas actuales. *Psicothema*, 16(1), 156-162.
- Arnau J. y Bono R. (1993). Análisis de series temporales bivariantes: aplicación al ámbito de la investigación social. *Psicothema*, 5(2), 439-448.
- Arnau J. y Bono R. (2008). Estudios longitudinales. Modelos de diseño y análisis. *Escritos en psicología*, 2(1), 32-41.
- Asociación Española de Normalización y Certificación AENOR. (2005). *Sistemas de gestión de la calidad. Fundamentos y vocabulario. UNE-EN-ISO9000:2005*. Madrid: AENOR.

- Asociación Española de Normalización y Certificación AENOR. (2008). *Sistemas de gestión de la calidad. Requisitos. UNE-EN-ISO9000:2008*. Madrid: AENOR.
- Astrid, F., A., Scherndl, T. y Kühberger, A. (2012). A comprehensive review of reporting practices in psychological journals: Are effect sizes really enough? *Theory and Psychology*, 23(1), 98–122.
- Atkins, D. C. (2005). Using multilevel models to analyze couple and family treatment data: Basic and advanced issues. *Journal of Family Psychology*, 19(1), 98-110.
- Atkins, D., Best, D., Briss, P. A., Eccles M., Falck-Ytter Y., Flottorp S., et al. y GRADE Working Group (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal*, 328 (7454), 1490.
- Atkins, D. C., Bedics, J. D., McGlinchey, J. B. y Beauchaine, T. P. (2005). Assessing Clinical Significance: Does It Matter Which Method We Use? *Journal of Consulting and Clinical Psychology*, 73(5), 982-989.
- Augusto J. M., y Montes B. (2009). Perceived Emotional Intelligence, Health and Somatic Symptomatology in Nursing Students. *Individual Differences Research*, 7(3), 197-211.
- Auping, J. (2000). *Una revisión de la teoría psicoanalítica a la luz de la ciencia moderna*. México D.F.: Plaza y Valdés.
- Ávila A. (1990). Psicodinámica de la depresión. *Anales de psicología*, 6(1), 37-58.
- Ávila, A., Blanco, A., Calonge, I., Forns, M., Oliva, M., Martorell, M. C., Rodríguez, C., Silva, F. y Torres, M. (1994). Diagnosis and Psychological Assessment in Spain a review and new trends. *Clínica y Salud*, 5(1), 83-120.
- Badia, X. y Bigorras, J. (2000). La Investigación de Resultados en Salud: de la evidencia práctica clínica. En X. Badia (Coord.). *La Investigación de Resultados en Salud* (p. 13-33). Barcelona: Edimac.
- Badía, X. y Lizán, L. (2002). Reflexiones sobre la investigación de resultados en salud. *Atención Primaria*, 30(6), 388-391.
- Badia, X. y Lizán, L. (2007). De la medicina basada en la evidencia a la investigación de resultados en salud: aproximándonos a la realidad. *Revista Española de Economía de la Salud*, 6(1), 42-44.

- Bados, A., García, E. y Fusté, A. (2002). Eficacia y utilidad clínica de la terapia psicológica. *Revista Internacional de Psicología clínica y salud/International Journal of Clinical and Health Psychology*, 2(3), 447-502.
- Bados, A. y García-Grau, E. (2011). *La práctica de la psicología clínica: contexto, proceso y habilidades*. Saarbrücken, Alemania: Editorial Académica Española.
- Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology: Toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest*, 9(2), 67-103.
- Bakke O. M., Carné X. y García, F. (1994). *Ensayos clínicos con medicamentos. Fundamentos básicos, metodología y práctica*. Barcelona: Mosby/Doyma.
- Ballester, R. (1998). *Introducción a la psicología de la salud*. Valencia: Promolibro.
- Ballesteros, F., Fernández, P. y Labrador, F. J. (2013). Factores que influyen en la duración de los tratamientos psicológicos empíricamente apoyados. *Anales de psicología*, 29(1), 94-102.
- Balluerka, N., Vergara A. I. y Arnau J. (2009), Calculating the main alternatives to null-hypothesis-significance testing in between-subject experimental designs. *Psicothema*, 21(1), 141-151.
- Bandura, A. (1977). *Social learning theory*. New York: Holt, Rinehart and Wilson.
- Baños, J. E., Brotons, C., Farré, M. (1998). *Glosario de investigación clínica y epidemiología*. Barcelona: Fundación Dr. A. Esteve.
- Barbero, M. I. (1999b). Desarrollos recientes de los modelos psicométricos de teoría de respuesta a los ítems. *Psicothema*, 11(1), 195–210.
- Barbero, M. I. (1999a). *Psicometría*. Madrid: UNED.
- Barkham, M. (2002). Methods, outcomes and processes in the psychological therapies across four successive research generations, En W. Dryden (Ed.), *Handbook of individual therapy* (pp. 372-433). London: Sage.
- Barkham, M., Hardy, G. E. y Shapiro, D. A. (2011). The Sheffield-Leeds psychotherapy research programme. In J.C. Norcross, G.R. Vandenbos y D.K. Freedheim (Eds.), *History of psychotherapy: Continuity and change* (pp.382–388). Washington, DC: American Psychological Association.

- Barkham, M., Hardy, G.E. y Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Oxford: Wiley-Blackwell.
- Barlow, D.H. (2002). *Anxiety and its disorders. The nature and treatment of anxiety and panic*. Nueva York: Guilford Press
- Barrios, B. A. y Hartman, D. P. (1986). The contributions of traditional assessment: Concepts, issues and methodologies. En R. Nelson, y S. Hayes (Eds.), *Conceptual Foundations of Behavioral Assessment* (pp. 81-110). Nueva York: Guildford Press.
- Bauer, S., Lambert, M. J., y Nielsen, S. L. (2004). Clinical significance methods: A comparison of statistical techniques. *Journal of Personality Assessment*, 82(1), 60–70.
- Baumeister, H. y Härter, M. (2007). Prevalence of mental disorders based on general population surveys. *Social Psychiatry and Psychiatric Epidemiology*, 42(7), 537-546.
- Beach, S. R. y Amir, N. (2003). Is depression taxonomic, dimensional, or both? *Journal of Abnormal Psychology*, 112(2), 228-236.
- Beauchamp, T. (2003). Methods and principles in biomedical ethics. *Journal of Medical Ethics*, 29(5), 269-274.
- Beauchamp, T. y Childress, J. (1994). *Principles of biomedical ethics*. Oxford: Oxford University.
- Beck, A. T. (1970). *Depression, causes and treatment*. Philadelphia: University Press
- Beck, A. T. y Clark, D. A. (1988). Anxiety and depression: An information processing perspective. *Anxiety Research*, 1(1), 23-36.
- Beck, A. T. y Clark, D. A. (1997). An information processing model of anxiety: automatic and strategic processes. *Behavior Research Therapy*, 35(1), 49-58.
- Beck, A. T., Rush, A. J., Shaw, B. F., y Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- Beck, A. T., Rush, A. J., Shaw, B. F. y Emery, G. (1983). *Terapia cognitiva de la depresión. Tercera Edición*. Bilbao: Declee Ediciones.
- Beckstead, D. J., Hatch, A. L., Lambert, M. J., Eggett, D. L., Goates, M. K. y Vermeersch, D. A. (2003). Clinical significance of the Outcome Questionnaire (OQ-45.2). *The Behavior Analyst Today*, 4(1), 86–96.

- Becoña, E. (1999). La discrepancia entre la investigación y la práctica clínica de la terapia de conducta. *Revista de Psicopatología y Psicología Clínica*, 4(2), 71-103.
- Benítez, A. M., Molina, W. y Camps T. (2005). Aspectos psicológicos del asma infantil. *Revista de Pediatría en Atención Primaria*, 7(2), 137-149.
- Bennet, S., Perry, N. W. y Rozensky, R. H. (2002). *Handbook of clinical health psychology. Medical disorders and behavioral applications*. Washington: American Psychological Association.
- Bentham, J. (1836). *Deontología o Ciencia e la Moral*. Valencia: Ferrer de Orga.
- Berbesi, D.Y., García, M., Segura A., Posada, R., y Serna, P. (2013). Efectividad de un programa de terapia de familia en niños asmáticos con familias disfuncionales. *Revista Clínica de Medicina de Familia*, 6(3), 138-143.
- Berenguer, G. y Quintanilla, I. (1994). La imagen de la psicología y los psicólogos en el estado español. *Papeles del Psicólogo*, 58, 41-68.
- Beretta, V., Roten, Y. de, Drapeau, M., Kramer, U., Favre, N. y Despland, J. (2005). Clinical significance and patients perceived change in four sessions of brief psychodynamic intervention: characteristics of early responders. *Psychology and psychotherapy: Theory, Research and Practice*, 78(3), 347-362.
- Bergin, A. E. (1963). The effects of psychotherapy: Negative Results revisited. *Journal of Counseling Psychology*, 10(3), 244-250.
- Bergin, A. E. (1967). Further comments on psychotherapy research for therapeutic practise. *International Journal of Psychiatry*, 3(4), 317-323.
- Bergin, A. E. y Lambert, M. J. (1978). The evaluation of therapeutic outcomes. En S. L. Gardfield y A.E. Bergin (Eds.), *Handbook of Psychotherapy and Behavior Change* (pp. 142-143). New York: Wiley.
- Berkovits, I., Hancock, G. R. y Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement*, 60(6), 877-892.
- Bermejo, V. (2009). Motivaciones para la revisión y cambios en el código deontológico de la profesión de psicología. *Papeles del Psicólogo*, 30(3), 195-206.

- Bernaldo-de-Quirós, M., Estupiñá, F. J., Labrador, F. J., Fernández-Arias, I., Gómez L., Blanco, C., y Alonso, P. (2012). Diferencias socio demográficas, clínicas y de resultados del tratamiento entre los trastornos de ansiedad. *Psicothema*, 24(3), 396-401.
- Berrio-Costa, G. y Hernández-Zubieta, P. (2003) Tratamientos con apoyo empírico en psicología de la salud. *Cuadernos Hispanoamericanos de psicología*, 3(2), 7-26.
- Beutler, L. E. (1997). The psychotherapist as a neglected variable in psychotherapy: an illustrated by reference to the role of therapist experience and training. *Clinical Psychology*, 4(1), 44-52.
- Beutler, L. E. (2001) Comparisons of quality assurance systems: from outcome assessment to clinical utility. *Journal of Consulting and clinical psychology*, 69(2), 197-204.
- Beutler, L. E. (2009). Making science matter in clinical practice: Redefining psychotherapy. *Clinical Psychology: Science and Practice*, 16(3), 301-317.
- Beutler, L. E. y Harwood, T. M. (2000). *Prescriptive psychotherapy: A practical guide to systematic treatment selection*. New York: Oxford University Press.
- Biescad, M. y Timulak L. (2014) Measuring psychotherapy outcomes in routine practice: Examining Slovak versions of three commonly used outcome instruments. *European Journal of Psychotherapy and Counselling*, 16(2), 140-162.
- Bihlar, B. y Carlsson, A. (2001). Planned and actual goals in psychodynamic psychotherapies: do patients' personality characteristics relate to agreement?. *Psychotherapy Research*. 11(4) 383-400.
- Bjelland, I., Lie, S. A., Dahl, A., Mykletun, A., Stordal, E., y Kraemer, H. C. (2009) A dimensional versus a categorical approach to diagnosis: Anxiety and depression in the HUNT 2 study *International Journal of Methods in Psychiatric Research*, 18(2), 128-137.
- Blanco, A. (1986). *Apuntes de psicodagnóstico*. Valencia: Promolibro.
- Blanco, A. (1987). La Psicología Comunitaria: Variaciones sobre un mismo tema. *Papeles del Colegio Oficial de Psicólogos*, 5 (31), 41-46.
- Bohart, A. C., O'Hara M. y Leitner, L. M. (1998). Empirically violated treatments: Disenfranchisement of humanistic and other psychotherapies. *Psychotherapy Research*, 8(2), 141-157.

- Boletín Oficial del Estado BOE. (1974). *Disposiciones generales. Colegios profesionales*. 15 de febrero de 1974 BOE-A-1974-289, 3046-3049.
- Boletín Oficial del Estado BOE. (1982), *Ley Orgánica. De protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen*. 14 de Mayo de 1982. BOE-A-1982-11196, 12546-12548.
- Boletín Oficial del Estado BOE. (1999). *Ley Orgánica de protección de datos de carácter personal*. 13 de diciembre de 1999 BOE-A-1999-23750. 43088-43099.
- Boletín Oficial del Estado BOE. (2002). *Ley básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica*. 14 de noviembre de 2002. BOE-A-2002-2218 40126- 40132
- Boletín Oficial del Estado BOE. (2008). *Real Decreto. Disposiciones generales Se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal*. 29 de Enero de 2008. BOE-A-2008-979. 4103 a 4136.
- Bonferroni C. E. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.
- Bono R., Arnau J. y Vallejo G. (2008). Técnicas de análisis aplicadas a datos longitudinales en psicología y ciencias de la salud: periodo 1985-2005. *Papeles del Psicólogo*, 29(1):136-146.
- Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.
- Botella J. (2000) El papel de los estudios de resultados de las terapias psicológicas. *Psicothema*, 12(2), 176-179.
- Botella, C. (2009). *Tratamientos psicológicos basados en la evidencia*. Castellón de la Plana: Servicio de Comunicación y Publicaciones Universitat Jaume I.
- Bothe A. K. y Richardson J. D. (2011). Statistical, practical, clinical, and personal significance: Definitions and applications in speech-language pathology. *American Journal of Speech-Language Pathology*, 20(3), 233-242.
- Boulet L. P., Vervloet D., Magar Y., Foster J. M. (2012). Adherence : The Goal to Control Asthma. *Clinics in chest medicine*, 33(3), 405-417.
- Bouza, A. (2000). Reflexiones acerca del uso de los conceptos de eficiencia, eficacia y efectividad en el sector salud. *Revista Cubana de Salud Pública*, 26(1), 50-56.

- Bowers, M. R., Swan J. E. y Kohler, W. F. (1994). What attributes determine quality and satisfaction with health care delivery?. *Health Care Manage Review*, 19(4), 49-55.
- Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44(3), 576-578.
- Box, G., Hunter J. S. y Hunter, W. G. (2008) *Estadística para investigadores: diseño, innovación y descubrimiento (2ª edición)*. Barcelona: Editorial Reverté.
- Boyd, M., Lasserson, T.J., McKean, M.C., Gibson, P.G., Ducharme, F.M. y Haby, M. (2009) *Interventions for educating children who are at risk of asthma-related emergency department attendance. Cochrane Database of Systematic Reviews 2009, Issue 2*. Art. No.: CD001290. Wiley Online Library [en línea], (fecha de consulta: 20 Noviembre 2014). Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001290.pub2/abstract>
- Brannon, L. y Feist, J. (2009). *Psicología de la salud*. Madrid: Paraninfo.
- Breitholtz, E., Johansson, B. y Ost, L. G. (1999). Cognition in generalized anxiety disorder and panic disorder patients: a prospective approach. *Behaviour Research and Therapy*, 37(6), 533-544.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Brennan, R. L. (2006). *Educational Measurement*. Connecticut: Praeger Publishers.
- Brito, G., González A. y Cruz P. (2010). Origin, course and reflections on evidence-based medicine. *Revista Cubana de Cirugía*, 49(2), [en línea], (fecha de consulta: 20 Septiembre 2013). Disponible en: [http://www.bvs.sld.cu/revistas/cir/vol49\\_2\\_10/cir14210.htm](http://www.bvs.sld.cu/revistas/cir/vol49_2_10/cir14210.htm)
- Bronfenbrenner, U., y Morris, P. A. (2006). The bioecological model of human development. En W. Damon y R. M. Lerner (Eds), *Handbook of child psychology Vol. 1: Theoretical models of human development* 6th edition (pp. 793-828). New Jersey: John Wiley & Sons.
- Browman, G., Gómez, A., Haynes, B., Jadad, A. y Gabriel, R. (2001). Herramientas para la práctica de la medicina basada en la evidencia. Desarrollo de guías de práctica clínica basada en la evidencia: de abajo-arriba. *Medicina Clínica (Barcelona)*, 116(7), 267-270.
- Brown J.D. (2001) Statistics corner, questions and answers about language testing statistics: Point-biserial correlation coefficients. *Shiken Research Bulletin*, 5(3), 13-17.

- Brown, C., Costigan T. E., y Kendziora T. (2008). Data analytic frameworks: analysis of variance, latent growth, and hierarchical models. En A. M. Nezu y Ch. M. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 285-313). New York: Oxford University press.
- Brown, M. B. y Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*, 69(346), 364–367.
- Bryk, A. S., y Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, California: Sage Publications.
- Buela-Casal, G., Caballo, V. y Sierra, C. (1996). *Manual de evaluación en psicología clínica y de la salud*. Veintiuno: Madrid: Ediciones Siglo XXI.
- Buela-Casal, G. y Sierra, J. C. (1997). *Manual de evaluación psicológica: fundamentos, técnicas y aplicaciones*. Madrid: Siglo XXI.
- Burgos, F. (2008). *Guía práctica de la espirometría*. Barcelona: Esmonpharma.
- Caballo, V.E. (2007). *Manual para el tratamiento cognitivo-conductual de los trastornos psicológicos. Volumen 1. Trastornos por ansiedad, sexuales, afectivos y psicóticos*. Madrid: Siglo XXI Editores.
- Cabañas, R. (1994). Evaluación de la Atención de la Salud. *Revista Médica Hondureña*, 62(1), 25-29.
- Cabrero J. y Richart M. (2000). *Concepto y estado actual de la investigación en enfermería*. Alicante: Publicaciones Universidad Alicante.
- Calam, R., Gregg, I., Simpson, B., Morris, J., Woodcock, A. y Custovic, A. (2003). Childhood asthma, behaviour problems, and family functioning. *Journal of Allergy and Clinical Immunology*, 112(3), 499-504.
- Cambless D.L. (2007) Psychotherapy Research and Practice: Friends or Foes? *PsycCritiques*, 52(1), 5.
- Campbell L. F., Norcross J. Vasquez M. y Kaslow N. (2013) Recognition of Psychotherapy Effectiveness: The APA Resolution. *Psychotherapy*, 50(1), 98-101.
- Campbell, D. T. (1957). A typology of tests, projective and otherwise. *Journal of Consulting Psychology*, 21(3), 207-210.

- Campbell, D. T. (1960) Recommendations for APA standards regarding construct, trait and discriminant validity. *American psychologist*, 15(8), 546-553.
- Campbell, D. T., y Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Campbell, D. y Stanley, C. (1995). *Diseños experimentales y cuasiexperimentales en la investigación social*. Buenos Aires: Amorrortu editores.
- Campo-Arias, A. y Cassiani, M. (2008) Trastornos mentales más frecuentes: prevalencia y algunos factores sociodemográficos asociados. *Revista Colombiana de Psiquiatría*, 37(4), 598-613.
- Campo-Arias, A. y Oviedo, H. C. (2008). Psychometric properties of a scale: internal consistency. *Revista de Salud Pública*, 10(5), 831-839
- Cano, G., Dastis, C., Morales, I., Manzanares, M. L., Fernández, A. y Martín, L. (2014). Ensayo clínico aleatorio para evaluar la eficacia de una intervención educativa desarrollada en atención primaria sobre asmáticos adultos. *Atención Primaria*, 46(3), 117-139.
- Cano, J., Rodríguez, L., García, J., y Antuña, A. (2005) *Introducción a la psicología de la personalidad aplicada a las ciencias de la educación: Manual Teórico*. Sevilla: MAD SL-Eduforma.
- Cano-De la Cuerda, R., Useros-Olmo, A. E., Muñoz-Hellín, E. (2010). Eficacia de los programas de educación terapéutica y de rehabilitación respiratoria en el paciente con asma. *Archivos de Bronconeumología*, 46(11), 600-606.
- Cano-Vindel, A y Miguel-Tobal, J. J. (1999). Evaluación de la ansiedad desde un enfoque interactivo y multidimensional: El Inventario de Situaciones y Respuestas de Ansiedad (ISRA). *Psicología Contemporánea*, 6(1), 14-21.
- Caparrós-Caparrós, B., Villas-Hoz, E., Juan-Ferrer, J. y Viñas-Poch, F. (2007). Symptom Checklist-90-R: fiabilidad, datos normativos y estructura factorial en estudiantes universitarios. *International Journal of Clinical and Health Psychology*, 7 (3), 781-794.
- Carpintero, H. (2003). History of Psychological Assessment. En R. Fernández-Ballesteros (Ed.), *Encyclopedia of Psychological Assessment*, vol1 (pp.447-452) Sage Pub: Londres.
- Carretero-Dios, H., y Pérez, C. (2007). Standards for the development and the review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7(3), 863-882.

- Carrobles, J. A. (1981). Registros psicofisiológicos. En Fernández-Ballesteros y J. A. Carrobles (Eds.), *Evaluación conductual* (pp. 350-391). Madrid: Pirámide
- Carvajal-Ureña, I., García-Marcos, L., Busquets-Monge R., Morales, M., García de Andoain, N., Batlles-Garrido, J., ... Bellido-Blasco, J. (2005). Variaciones geográficas de la prevalencia de síntomas de asma en los niños y adolescentes españoles. International Study of Asthma and Allergies in Childhood (ISAAC) Fase III España. *Archivos de Bronconeumología*, 41(12), 659-666.
- Casado, M. (2006). *Psicólogos internos residentes: Manual de preparación PIR*. Sevilla: Editorial Mad.
- Castañeda, I. y Mira, J. J. (1998). Abandono de las consultas en salud mental: Estudio sobre los factores que determinan su interrupción prematura. *Cuadernos de Medicina Psicosomática y Psiquiatría de Enlace*, 11, 42-52.
- Castilla del Pino, C. (2002). *Un estudio sobre la Depresión. Fundamentos de antropología dialéctica. 10ª edición*. Barcelona: Península.
- Castonguay, L. G. (2011). Psychotherapy, psychopathology, research and practice: Pathways of connections and integration. *Psychotherapy Research*, 21(2), 125–140.
- Castonguay L. G. y Beutler L. E. (2006). *Principles of therapeutic change that work*. Nueva York: Oxford University Press.
- Castonguay, L. G., Boswell, J.F., Zack, S., Baker, S., Boutselis, M., Chiswick, N., ... Grosse, M. (2010). Helpful and hindering events in psychotherapy: A practice research network study. *Psychotherapy: Theory, Research, Practice and Training*, 47(3) 327-344.
- Cattell, J. M. (1890). Mental tests and measurements. *Mind*, 15(59), 373-381.
- Cautilli, J. (2006). Dodo-Bird or dressing up the pig and the politics of psychotherapy: a reply to Vandenberghe and de Sousa. *International Journal of Behavioral Consultation and Therapy*, 2(2), 305.
- Chambless, D. L. (2002). Beware the Dodo Bird: the dangers of overgeneralization. *Clinical Psychology: Science and Practice*, 9(1), 13–16.
- Chambless, D. L., Baker, M.J., Baucom, D.H., Beutler, L.E, Calhoun, K. S., Crits-Christoph, P., et al. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, 51(1), 3-16.

- Chambless, D. L. y Ollendick, T. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52(1), 685-716.
- Chambless, D. L., Sandersson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Cristoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, 49(2), 5-18.
- Charlton, B. G. (1995). Mega-trials: Methodological issues and clinical implications. *Journal of the Royal College of Physicians of London*, 29(2), 96-100.
- Charlton, B. G. (1997). Restoring the Balance: Evidence-based Medicine Put in Its Place. *Journal of Evaluation in Clinical Practice*, 3(2), 87-98.
- Chartier, S. y Faulkner, A. (2008). General Linear Models: An Integrated Approach to Statistics. *Tutorials in Quantitative Methods for Psychology*, 4(2): 65-78.
- Chen A. Y. y Escarce J. J. (2008). Family structure and the treatment of childhood asthma. *Medical Care*, 46(2), 174-184
- Chen, E., Bloomberrg, G. R., Fisher, E. B. y Strunk, R. C. (2003). Predictors of repeat hospitalizations in children with asthma: the role of psychosocial and sosioenviromental factors. *Health Psychology*, 22(1), 12-18.
- Chinchilla, A. (2008) *La depresión y sus máscaras: aspectos terapéuticos*. Madrid: Editorial Médica Panamericana.
- Christensen, L. y Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC Index. *Behavior Therapy*, 17(3), 305-308.
- Christiansen, R. (2005). Testing Fisher, Neyman, Pearson and Bayes. *The American Statistician*, 59(2), 121-126.
- Clancy C. y Eisenberg J. (1997). Outcomes Research at the Agency for Health Care Policy and Research. *Disease Management and Clinical Outcomes*, 1(3), 72-80.
- Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R. y Wright B. (2009). Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. *Behavior Research and Therapy*, 47(11), 910-920.
- Clarkin, J. F., y Levy, K. N. (2004). The influence of client variables on psychotherapy. En M.J. Lambert (Ed.), *Handbook of psychotherapy and behavior change 5th ed.* (pp. 194-226). New York: Wiley.

- Clay R.A. (2012). Protesting proposed changes to the DSM. *Monitor on psychology, American Psychological Association*, 43(2), 42, [en línea]. (fecha de consulta: 15 Diciembre 2014). Disponible en: <http://www.apa.org/monitor/2012/02/changes-dsm.aspx>
- Cleary, A. (1982) *Instrumentación en psicología*. México: Limusa.
- Clemente M. (1998). *Fundamentos de la psicología jurídica*. Madrid: Pirámide.
- Coan, R. W. (1968). Dimensions of psychological theory. *American Psychologist*, 23(10), 715-722.
- Coan, R. W. (1979). *Psychologists: Personal and theoretical pathways*. New York: Irvington Publishers.
- Cobo, J. V. (2005). El concepto de depresión. Historia, definición, nosología, clasificación. *Revista Psiquiatria.com*, 9(4), [en línea]. (fecha de consulta: 20 Diciembre 2014). Disponible en: <http://www.psiquiatria.com/revistas/index.php/psiquiatriacom/article/view/283/>
- Coderch, J. (1982). *Psiquiatría dinámica*. Barcelona: Herder.
- Coe, R. y Merino, C. (2003). Magnitud del efecto: Una guía para investigadores y usuarios. *Revista de psicología de la Pontificia Universidad Católica del Perú*, 21(1), 145-177.
- Coelli, T. Prasada, D. S. y Battese, G. E. (1998). *An introduction to efficiency and productivity analysis*. Boston: Kluwer Academic Publisher.
- Coffman, K. (2002). Psychiatric issues in pulmonary disease. *Psychiatric Clinics of North America*, 25(1), 89-127.
- Cohen, A. (2008). IAPT: A Brief History. *Healthcare Counselling and Psychotherapy Journal*, 5(2), 8-11.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd. Edit)*. Hillsdale, New Jersey: Erlbaum.
- Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49(12), 997-1003.
- Cohen, R. y Swerdlik, M. (2000). *Pruebas y evaluación psicológicas*. México: MacGraw Hill.
- Colegio Oficial de Psicólogos COP Madrid. (2011). *Ética y Deontología en la práctica psicológica*. Madrid: COP Madrid.
- Compas, B. E., Haaga, D. A., Keefe, F. J., Leitenberg. H. y Williams, D. (1998). Sampling of empirically supported psychological treatments from health psychology: Smoking, chronic

- pain, cancer and bulimia nervosa. *Journal of Consulting and Clinical Psychology*, 56(1), 89-112.
- Conde J. L. (2000). Estudios de eficacia y efectividad. En X. Badia (Coord.). *La Investigación de Resultados en Salud* (p. 35-44). Barcelona: Edimac.
- Conde, J. L. (2002). Eficacia y efectividad: una distinción útil para la práctica y la investigación clínicas. *Nefralogía*, 22(3): 219-222.
- Conde, C., Orozco, L., Báez, A. y Dallos, A. (2009). Aportes fisiológicos a la validez de criterio y constructo del diagnóstico de ansiedad según entrevista psiquiátrica y del State-Trait Anxiety Inventory (STAI) en una muestra de estudiantes universitarios colombianos. *Revista Colombiana de Psiquiatría*, 38(2), 263-278.
- Consejo General de Colegios Oficiales de Psicólogos CGCOP. (2009). Código Deontológico del Psicólogo. [en línea] (fecha de consulta: 1 Mayo 2013). Disponible en: <http://www.cop.es/cop/codigo.htm>
- Cook T. D., y Leviton, L. C. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. *Journal of Personality*, 48(4), 449-472.
- Cook, D. J., Mulrow, C. D. y Haynes, R. B. (1997) Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, 126(5): 376-380.
- Corbella, S. y Botella, L. (2004) *Investigación en psicoterapia: Proceso, Resultado y factores comunes*. Madrid: Editorial Vision Net
- Cordero, A. (1997). *La evaluación psicológica en el año 2000*. Madrid: TEA Ediciones.
- Cortada de Kohan, N. (1999) *Teorías Psicométricas y Construcción de Tests*. Buenos Aires: Lugar Editorial.
- Cortina, J.M. y Nouri, H. (2000). *Effect Size for ANOVA Designs. Quantitative Applications in the Social Sciences*. Thousand Oaks: Sage.
- Cosio, B.G., Fiorentino, F. y Scrimini, S. (2010). EPOC y asma. *Archivos de Bronconeumología*, 46(8), 2-7.
- Costello, C. G. (1972). Depression: Loss of reinforcers or loss of reinforcer effectiveness?. *Behavior Therapy*, 3(2), 340-347.
- Covell D. G., Uman G. C. y Manning P. R. (1985). Information needs in office practice: are they being met?. *Annals of Internal Medicine*, 103(4), 596-599.

- Coyle, Y. (2000). Southwestern internal medicine conference: developing theoretical constructs for outcomes research. *The American Journal of the Medical Sciences*, 319(4), 245-249.
- Crawford, J. R. y Garthwaite, P. H. (2006). Comparing patients' predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*, 20(3), 259-271.
- Crawford, J. R. y Howell, D. C. (1998). Regression equations in clinical neuropsychology: An evaluation of statistical methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology*, 20(5), 755-762.
- Crits-Christoph, D. L., Frank E., Chambless D. L., Brody C. y Karp J. F. (1995). Training in empirically-validated treatments: What are clinical psychology students learning? *Professional Psychological Research Practice*, 26(5), 514-522.
- Crits-Christoph, P. & Gallop, R. (2006). Therapist effects in the National Institute of Mental Health treatment of depression collaborative research programme and other psychotherapy studies. *Psychotherapy Research*, 16(2), 178-181.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Cronbach, L. J. (1985). *Fundamentos de la exploración psicológica*. Madrid: Biblioteca Nueva.
- Cronbach, L. J. (1990). *Psychological testing (5ª ed.)*. Nueva York: Harper & Row.
- Crosby, R. D., Kolotkin, R. L. y Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology*, 56(5), 395-407.
- Cruz-Morales, S. E., González-Reyes, M. R., Gómez-Romero, J. G. y Arriaga, J. P. (2003). Modelos de ansiedad. *Revista Mexicana de Análisis de la conducta*, 29(1), 93-105.
- Cuesta, G., Moreno, R. y Gutiérrez, M. (1986). *La calidad de la asistencia hospitalaria*. Valencia: Doyma, S.A.
- Cuijpers, P., van Straten, A., Hollon, S.D., y Andersson, G. (2010). The contribution of active medication to combined treatments of psychotherapy and pharmacotherapy for adult depression: A meta-analysis. *Acta Psychiatrica Scandinavica*, 121 (6), 415-423.
- Cummings, N. A., O'Donohue, W. T y Ferguson, K.E. (2002). *The impact of medical cost offset on practice and research*. Reno, NV: Context Press.

- Cureton, E. (1965). Reliability and validity: Basic assumptions and experimental designs. *Educational and Psychological Measurement*, 25(2), 327-346.
- Curran, P. J. y Willoughby, M. T. (2003). Implications of latent trajectory models for the study of developmental psychopathology. *Development and Psychopathology and Psychopathology*, 15(3), 581-612.
- Davidoff, F., Haynes B., Sackett D., y Smith, R. (1995). Evidence based medicine. *British Medical Journal*, 310 (6987), 1085-1086.
- Day, W. (1980). The historical antecedents of contemporary behaviorism. En R.W. Rieber y K. Salzinger (Eds.), *Psychology: Theoretical perspectives* (pp. 203-262). New Jersey: Academic Press.
- De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York: The Guilford Press.
- de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J. A., Giltay, E. J., van Noorden, M. S., ... Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology and Psychotherapy*, 18(1), 1-12.
- De Bruyn, E. J. (1992). A normative-prescriptive view on clinical psychodiagnostic decision making. *European Journal of Psychological Assessment*, 8(3), 163-171.
- De Irala, J., Martínez-González, M. A., Guillén-Grima, F. (2001). ¿Qué es una variable de confusión? *Medicina Clínica (Barcelona)*, 117(10), 377-385.
- De Jong, K., Nugter, M. A., Polak, M., Wagenborg, J. E., Spinhoven, y Heiser, W.J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology and Psychotherapy*, 14 (4), 288-301.
- de la Parra, G. y von Bergen, A. (2006). Medición de los resultados en psicoterapia: Uso del OQ-45, un instrumento validado en Chile. *Revista Gaceta Universitaria Universidad de Chile*, 2(2), 208-222.
- De Miguel, M., Barcina, C., Muñoz, M. y Leal, M. (2008). Control of persistent asthma in Spain: associated factors. *Journal of Ashtma*, 45(9), 740-746.
- de Miguel J., Jimenez R., Hernandez V., Noriega C., Juarez M. C., Chancafe J. Puente L., Hernandez J. y Carrasco P. (2012). Disfunción psicológica y salud mental en pacientes con asma. *Archivos de Bronconeumología*, 48(Esp. Cong.), 5.

- De Silva P., Rachman, S y Seligman M. E. (1977). Prepared phobias and obsessions: Therapeutic outcome. *Behavior Research and therapy*, 15(1), 65-77.
- De Vet, H. C., Ostelo, R. W., Terwee, C. B., van der Roer, N., Knol, D. L., Beckerman, H., Boers, M. y Bouter, L. M. (2007). Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Quality of Life Research*, 16(1): 131-142.
- Del Barrio, V. (2003) *Evaluación psicológica aplicada a diferentes contextos*. Madrid: Librería UNED.
- Del Llano-Señarís, J., Ortún, V., Martín, J., Millán, J., Gené, J. y Campella, M. (2003). *Gestión Sanitaria: Innovaciones y Desafíos*. Barcelona: Masson, S.A.
- Demoly P, Paggiaro P, Plaza V, Bolge SC, Kannan H, Sohler B, et al. (2009). Prevalence of asthma control among adults in France, Germany, Italy. Spain and UK *European Respiratory Review*, 18(112), 105-112.
- Denker, R. G. (1937). Prognosis and life expectancy in the psychoneuroses. *Procedures Association of Life Insurance Medical Directors of America*, 24, 179.
- Derogatis, L. (2002). SCL-90-R: Cuestionario de 90 síntomas. Madrid: TEA Ediciones
- Derogatis, L., y Cleary, P. (1977). Confirmation of the dimensional structure of the SCL-90: A study of construct validation. *Journal of Clinical Psychology*, 33(4), 981-989.
- Deshmukh V. M., Toelle B.G., Usherwood T., O'Grady B, Jenkins C. R. (2008). The association of comorbid anxiety and depression with asthma-related quality of life and symptom perception in adults. *Respirology*, 13(5), 695–702.
- Diario Oficial de la Unión Europea DOUE. (1995). *Relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos*. 24 de octubre de 1995. Directiva 95/46/CE del Parlamento Europeo y del Consejo. DOUE-L-1995-81678.
- Dixon-Gordon, K. L., Turner, B.J., y Chapman, A. L. (2011). Psychotherapy for personality disorders. *International Review of Psychiatry*, 23(3), 292-302.
- Doménech, J.M. (2013a). *Fundamentos de diseño y estadística Vol.8: Investigación científica y diseño de estudios*. 14ª Ed. Barcelona: Signo.

- Doménech, J. M. (2013b). *Fundamentos de diseño y estadística Vol.7: Comprobación de hipótesis: Pruebas de significación, pruebas de hipótesis y tamaño de los grupos (14ª ed.)*. Barcelona: Signo.
- Doménech, J. M. (2013c). *Fundamentos de diseño y estadística Vol.14: Medida del cambio: Análisis de diseños con medidas intrasujeto. Estudios de concordancia (14ª ed.)*. Barcelona: Signo.
- Domingo, J. (2013). La no decepción del DSM-VI. *Cuadernos de Neuropsicología.*, 7(1), 10-21.
- Donabedian, A. (1966). Evaluating the Quality of Medical Care. *Milbank Memorial Fund Quarterly*, 44(3), 166-206.
- Donabedian, A. (1984). *La calidad de la atención médica*. México: La Prensa Medica Mexicana.
- Donabedian, A. (1991). Una aproximación a la monitorización de la calidad asistencial. *Control de Calidad Asistencial*, 6(1), 1-6.
- Doward, L. C., y McKenna, S. P. (2004). Defining Patient-Reported Outcomes. *Value in Health*, 7(Suppl 1), 4-8.
- Drisko, J. W. (2004). Common factors in psychotherapy outcome: Meta-analytic findings and their implications for practice and research. *Families in Society: The Journal of Contemporary Human Services*, 85(1), 81-90.
- Drummond, M., Stoddart, G. y Torrance, G. (1991). *Métodos para la evaluación económica de los programas de atención de la salud*. Madrid: Díaz los Santos
- Du Prel, J. B., Hommel, G., Röhrig, B. y Blettner, M. (2009). Confidence interval or P-Value. Part 4 of a Series Evaluation Scientific Publications. *Deutsches Ärzteblatt International*, 106(19), 335-339.
- Duero, D. G. y Shapoff, V. (2009). El conflicto nosológico en psicopatología : notas críticas sobre el diagnóstico psiquiátrico. *Revista CES Psicología*, 2(2), 20-48.
- Duncan, B. y Miller, S. D. (2000). *The heroic client: Doing client-directed, outcome-informed therapy*. San Francisco, California: Jossey Bass.
- Duncan, B. L., Miller, S. D., Wampold, B. E., y Hubble, M. A. (2010). *The heart and soul of change: Delivering what works in therapy (2nd, Ed.)*. Washington, DC: American Psychological Association.

- Echeburúa, E. y de Corral, P. (2001). Eficacia de las terapias psicológicas: de la investigación a la práctica clínica. *Revista Internacional de Psicología Clínica y de la Salud / International Journal of Clinical and Health Psychology*, 1(1), 181-204.
- Echeburúa, E., Amor, P. y De Corral, P. (2003). Autoinformes y entrevistas en el ámbito de la psicología clínica forense: limitaciones y nuevas perspectivas. *Análisis y Modificación de la conducta*, 29, (126) 503-522.
- Echeburúa, E., Salaberría, K. y Corral P. (2010) Efectividad de las terapias psicológicas: un análisis de la realidad actual. *Revista de psicopatología y psicología clínica*, 15(2), 85-99.
- Echeburúa, E., Salaberría, K., Corral, P. y Polo-López, R. (2010). Terapias psicológicas basadas en la evidencia: Limitaciones y retos de futuro. *Revista argentina de clínica psicológica*, 19(3), 247-256.
- Echeburúa, E., Salaberría, K., y Cruz-Sáez, M. (2014). Aportaciones y Limitaciones del DSM-5 desde la Psicología Clínica. *Terapia psicológica*, 32(1), 65-74.
- Ehlers, A., Bisson, J., Clark, D.M. y Yule, W. (2010). Do all psychological treatments really work the same in posttraumatic stress disorder? *Clinical Psychological Review*, 30(2), 269-276.
- Elkins, D. N. (2009). The medical model in psychotherapy: Its limitations and failures. *Journal of Humanistic Psychology*, 49(1), 66–84.
- Elliott, R. E. (1998). A guide to the Empirically Supported Treatment controversy. *Psychotherapy Research*, 8(2), 115-126.
- Elosua, P. (2003). Sobre la validez de los test. *Psicothema*, 15(2), 315-321.
- Emilfork, M. (2009) La investigación de resultados en Salud u Outcome Research. *Acta Medica Clínica Santa María*, 3 (1), 31-34.
- Engel, G. (1977). The need for a new medical model: a challenge for biomedicine. *Science*, 196(4286), 129-136.
- Engel, G. L. (1997). From Biomedical to biopsychosocial. Being scientific in the human domain. *Psychosomatics*, 38, 528.
- Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, 196(4286), 129-136.

- Engel, G. L. (1980). The clinical application of the biopsychosocial model. *The American Journal of Psychiatry*, 137(5), 535-544.
- Epstein, R. S., y Sherwood, L. M. (1996). From outcomes research to disease management: a guide for the perplexed. *Annals of Internal Medicine*, 124(9), 832-837.
- Espada, F. J., Torres, P., Cano-Vindel, A., y Ochoa, E. F. (1994). Evaluación de trastornos de ansiedad: Diferenciación en los tres sistemas de respuesta y áreas situacionales. *Clínica y Salud*, 5(2), 209-217.
- Estévez, A. (2008). *Factores cognitivos asociados a la depresión y la ansiedad en mujeres*. Saarbrücken: VDM Verlag Dr Muller.
- Estupiñá, F. J. (2012). Hacia una práctica de la psicología clínica basada en la evidencia empírica. En Labrador, F. J. y Crespo, M. (Coords.), *Psicología basada en la evidencia* (pp. 47-59). Madrid: Pirámide
- Everitt, B. S. y Skrondal, A. (2010). *The Cambridge Dictionary of Statistics 4º edition*. Cambridge: Cambridge University Press.
- Exner, J. (2001). *A Rorschach Workbook for the comprehensive system*. Asheville, North Carolina: Rorschach Work-shop.
- Eysenck, H. J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology*, 16(5), 319-324.
- Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, 33(5), 517.
- Eysenck, H. J. (1981). Behavior therapy and the conditioning model of neurosis. *International Journal of Psychology*, 16(4), 343-370.
- Eysenck, M. W. (1997) *Anxiety and cognition: a unified theory*. Hove, UK: Psychology Press.
- Fall, R. M. (1991). Manifiesto for science of clinical psychologist. *The clinical psychologist*, 44(6), 75-88.
- Faulkner, C., Fidler, F., y Cumming, G. (2008). The value of RCT evidence depends on the quality of statistical analysis. *Behavior Research and Therapy*, 46(2), 270-281.
- Federation of National Psychology Associations EFPA. (2005). *Metacode of ethics*. [en línea] (fecha de consulta: 1 Abril 2013). Disponible en: <http://www.efpa.eu/ethics>

- Feighner, J. P., Robins, E., Guze S. B., Woodruff R. A., Winokur, G. y Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26(1), 57-63.
- Feingold, A. (2009). Effect Sizes for Growth-Modeling Analysis for Controlled Clinical Trials in the Same Metric as for Classical Analysis. *Psychol Methods*, 14(1), 43-53.
- Feinstein, A. R. (1985). *Clinical Epidemiology: The Architecture of Clinical Research 2nd edition*. Philadelphia: W. B. Saunders Company.
- Feinstein A. R. y Horwitz R. I. (1997). Problems in the evience of evidence-based medicine. *The American Journal of Medicine*, 103(6), 529-535.
- Feixas, G., Evans, C., Trujillo, A., Saúl, L. A., Botella, L., Corbella, S., González, E., Bados, A., García-Grau, E. y López-González, M. A. (2012). La versión española del CORE-OM: Clinical Outcomes in Routine Evaluation - Outcome Measure. *Revista de Psicoterapia*, 23(89), 109-135.
- Feixas, G., Pucurull, C., Roca, C. Paz, C., García, E. y A. Bados A. (2012). Escala de satisfacción con el tratamiento recibido (CRES-4): la versión en español. *Revista de psicoterapia*, 23(89), 51-58.
- Felce, D. y Symons, F. J. (2000). *Behavioral Observation. Technologyand applications in developmental disabilities*. Baltimore: Paul H. Brooks Publishing.
- Feldt, L. S. (1975). Estimation of the Reliability of a Test Divided into Two Parts of Unequal Length, *Psychometrika*, 40(4), 557-561.
- Fenichel, O. (1930). Statistischer bericht über die therapeutische Tätigkeit 1920-1930. En S. Radó, O. Fenichel y C. Müller-Braunschweig (Eds.), *Zeh jahre Berliner psychoanalytisches institute. Poliklinik und Lehranstalt* (pp.13-19). Wien: Internationale Psychoanalyse.
- Fergusson, E. (2013). Personality is of central concern to understand health: towards a theoretical model for health psychology. *Health Psychology Review*, 7(1), 32-70.
- Fernández J. R. y Pérez-Álvarez, M. (2001). Separando el grano de la paja en los tratamientos psicológicos. *Psicothema*, 13(3), 337-344.
- Fernández, P. y Díaz, P. (2001). Significancia estadística y relevancia clínica *Cuadernos de Atención Primaria*, 8(3), 191-195.

- Fernández-Álvarez, H., Garcia, F., LoBianco, J. y Corbella, S. (2003). Questionnaire on the Personal Style of the Therapist PST-Q. *Clinical Psychology and psychotherapy*, 10(2), 116-125.
- Fernández-Ballesteros, R. (1980). *Psicodiagnóstico*. Madrid: Cincel-Kaelusz.
- Fernández-Ballesteros R. (1985). Evaluación psicológica y evaluación valorativa. *Evaluación Psicológica/ Psychological Assessment*, 1(2), 7-32.
- Fernández-Ballesteros, R. (1985). Psychological assessment and evaluation. *Psychological Assessment: An International Journal*, 1(1), 5-22.
- Fernández-Ballesteros R. (1987). Ciencia, ideología y política en evaluación de programas. *Revista de Psicología Social*, 2, 191-216.
- Fernández-Ballesteros, R. (1991). Anatomía de los autoinformes. *Evaluación Psicológica / Psychological Assessment*, 7(3), 263-291.
- Fernández-Ballesteros R. (1992). *Introducción a la evaluación psicológica*. Madrid: Pirámide.
- Fernández-Ballesteros, R. (2011). *Evaluación psicológica. Conceptos, métodos y estudio de casos*. Madrid: Ediciones Pirámide.
- Fernández-Ballesteros, R. y Calero, M. D. (1992) Técnicas objetivas: instrumentación y aparatos. En R. Fernández Ballesteros (Ed.), *Introducción a la evaluación psicológica*.(pp. 183-217). Madrid: Pirámide.
- Fernández-Ballesteros, R., De Bruyn, E., Godoy, A., Hornke, L., Ter, L., Vizcarro, C. Westhoff, K., Westmeyer, H. y Zaccagnini, J. (2003). Guías del proceso de evaluación GAP: Una propuesta para debatir. *Papeles del Psicólogo*, 23(84), 58-70.
- Fernández-Ballesteros, R., Márquez, M. O., Vizcarro, C. y Zamarrón, M. D. (2011). *Buenas prácticas y competencias en evaluación psicológica*. Madrid: Pirámide.
- Fernández-Hermida J. R. (2012). El final de la inocencia: la importancia del apoyo empírico a los tratamientos. En Labrador F.J y Crespo M (Coords.) *Psicología basada en la evidencia* (pp. 21-36). Madrid: Pirámide.
- Fernández-Ríos, M. y Sánchez, J. C. (1997). *Eficacia organizacional: Concepto, desarrollo y evaluación*. Madrid: Díaz de Santos.
- Fester, C. B. (1973). A functional analysis of depression. *American Psychologist*, 28(10), 857-870.

- Feuerstein, R. (1979). *The dynamic assessment of retarded performers. The learning potential assessment device. Theory, instruments and techniques*. Baltimore: University Park Press.
- Fidler, F., y Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random-effects effect sizes. *Educational and Psychological Measurement*, 61(4), 575-604
- Field, A. (2009). *Discovering statistics using SPSS (3rd ed)*. SAGE Publications: California.
- Fierro, A. (1987). Resumen algunas intervenciones Jornadas de Trabajo para la elaboración del Proyecto de Código Deontológico. *Papeles del Psicólogo*, 32, 6-14.
- Florentino, M.T. (2008). La construcción de la resiliencia en el mejoramiento de la calidad de vida y la salud. *Suma Psicológica*, 15(1), 95-114.
- Fisher R. A. (1925a). *Statistical methods for research workers*. Edinburgo: Oliver and Boyd.
- Fisher, R. A. (1925b). Applications of 'Student's' distribution. *Metron*, 5, 90-104
- Fisher R. A. (1991) *Statistical methods experimental design and scientific inference*. Oxford: Oxford University Press.
- Fishman, D. B. (2005). From Single Case to Database: A new method for enhancing psychotherapy practice. *Pragmatic Case Studies in psychotherapy* , 1(1),1-50.
- Fitzgerald, J. (2009). Measuring change using the Outcome Questionnaire-45 and the Helping Alliance Questionnaire-II: Experiences in a psychology training clinic in New Zealand. *4<sup>th</sup> Australasian Mental Health Outcomes & Information Conference, 11-12 June 2013*.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245.
- Foa, EB. , Rothbaum, B.O.; Riggs, D.S. y Murdock, T.B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. *Journal of Consulting and clinical Psychology*, 59(5), 715-723.
- Forns, M., Abad, J., Amador, J. A., Kirchner, T., y Roig, F. (2002). *Avaluació psicològica*. Barcelona: Editorial U. O. C. Colecció Manuals.
- Forns, M., Kirchner T. y Torres M. (1991). *Principales modelos de evaluación en psicología*. Barcelona: L'estudi d'en llop.

- Forns, M., Kirchner, T. y Torres, M. (1998). *Evaluación psicológica: modelos y técnicas*. Barcelona: Paidós Ibérica.
- Freud, D., Lave, J., Clancy, C., Hawker, G, Hasselblad, V., Keller, R., ... Wright, J. (1999). Patient Outcomes Research Teams: Contribution to outcomes and effectiveness research. *The Annual Review of Public Health*, 20, 337-359.
- Freud, S. (1917). *Duelo y melancolía*. En *Sigmund Freud Obras completas Volumen XIV* (pp.83-161) Edición 1992. Buenos aires: Amorrortu Editores.
- Freud, S. (1926). Inhibición, síntoma y Angustia. En *Sigmund Freud Obras completas. Volumen XX. Edición 1992* (pp. 83-161) Buenos aires: Amorrortu Editores.
- Freud, S. (1895a). Obsesiones y fobias. En *Sigmund Freud obras completas. Volumen I. Edición 2006*. (pp.178-182). Madrid: RBA Biblioteca Nueva.
- Freud, S. (1895b). La neurastenia y la neurosis de angustia. En *Sigmund Freud obras completas. Volumen I. Edición 2006*. (pp. 183-196). Madrid: RBA Biblioteca Nueva.
- Frías M. D. y Llobell J. P. (2003). Psicología clínica basada en pruebas: efecto del tratamiento. *Papeles del Psicólogo*, 24(85), 11-18.
- Frías, M. D., Pascual, J. y García, F. (2000). Tamaño del efecto del tratamiento y significación estadística. *Psicothema*, 12(2), 236-240.
- Friendman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200), 675–701.
- Froján, M. X. (2011). ¿Por qué funcionan los tratamientos psicológicos? *Clínica y Salud*, 22(3), 201-204.
- Fundación Vasca para la Excelencia EUSKALIT. (2014). *Modelos de Gestión Avanzada*. Zamudio: Gobierno Vasco / Eusko Jaurlaritz. [en línea] (fecha de consulta: 25 Junio 2014) Disponible en: <http://www.euskalit.net/nueva/images/stories/pdfs/modelo20152.pdf>
- Gabbard, G. O., Lazar, S.G., Hornberger, J., y Spiegel, D. (1997). The economic impact of psychotherapy: a review. *The American Journal of Psychiatry*, 154(2): 147-155
- Gabbard, G., Beck, J. y Holmes, J. (2007). *Oxford Textbook of Psychotherapy*. Oxford: Oxford University Press.
- Garaigordobil, M. (1998). *Evaluación Psicológica: Bases teórico-metodológicas, situación actual y directrices de futuro*. Salamanca: Amarú.

- Garb, H. N., y Boyle, P. A. (2003). Understanding why some clinicians use pseudoscientific methods: Findings from research on clinical judgment. En S. O. Lilienfeld, S. J. Lynn, y J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 17-38). New York: The Guilford Press.
- García, E. (2007). *Validación de un modelo para medir la calidad asistencial en los hospitales.*(Tesis Doctoral). Departamento Bioquímica y Biología Molecular, Microbiología, Medicina Preventiva y Salud Pública, Cádiz: Universidad de Cádiz.
- García, E., Bados, A. y Saldaña, C. (1998). La eficacia de la intervención psicoterapéutica: Una reflexión crítica. *Avances en Psicología Clínica Latinoamericana*, 16(1), 23-48.
- García, F., Alfaro, A. y Moreno, J. (2009). Evaluación de Resultados de Salud. Panorama sobre el uso de medidas de Resultados de Salud basadas en el paciente en la práctica clínica. *Revista Clínica Médica Familiar*, 2(6), 286-293.
- García-Fernández, J. M., Inglés, C. J., Martínez-Monteagudo, M. C., Marzo J. C. y Estévez E. (2011). Inventario de Ansiedad Escolar: validación en una muestra de estudiantes de Educación Secundaria. *Psicothema*, 23(2), 301-30
- García-Montalvo, C. (1997). Clasificación de las técnicas de evaluación psicológica. En G. Buela-Casal y J. Sierra (Eds.), *Manual de evaluación psicológica: fundamentos, técnicas y aplicaciones* (pp. 251-270) Madrid: Siglo XXI.
- García-Riaño, D. (1999). *Evaluación psicológica. Fundamentos básicos.* Valencia: Promolibro.
- García-Río, F., Calle, M., Burgos, F, Casan, P., del Campo, F., Galdiz J., ... Puente L. (2013). Espirometría. *Archivos de Bronconeumología*, 49(9), 388-401.
- Garfield, S. L. (1998). Some comments on empirically supported treatments. *Journal of Consulting and Clinical Psychology*, 66(1), 121-125.
- Garzón, A. y Caro, I. (2014). Evolution and future of health psychology. *Anales de psicología*, 30(1), 1-8.
- Gatz, M., Fiske, A., Fox, L. S., Kaskie, B., Kasl-Godley, J. E., McCallum, T. J., Wetherell, J. L. (1998). Empirically validated psychological treatments for older adults. *Journal of Mental Health Aging*, 4(1), 9-46.
- Gempp, R., y Avendaño, C. (2008). Datos normativos y propiedades psicométricas del SCL-90-R en estudiantes universitarios chilenos. *Terapia Psicológica*, 26(1), 39-58.

- George, D. y Mallery, P. (2013). *IBM SPSS 21 update. Step by step: A simple guide and reference*. (13th ed.). Boston: Pearson Education.
- Gerard, A. (2012). *Depresión. La enfermedad del siglo*. Bilbao: Ediciones Mensajero.
- Gero, G. (1936). The construction of Depression. *International Journal of Psychoanalysis*, 17, 423-461.
- Gibert J. (2012). Estadística para psiquiatras o ¿cuanto mejor es un antidepresivo que otro? *Opiniones en psiquiatría*, 3, 4-15.
- Gibson, P. G., Powell, H., Wilson, A., Abramson, M. J., Haywood, P., Bauman, A., ... Roberts, J. J. (2002a). *Limited (information only) patient education programs for adults with asthma. Cochrane Database of Systematic Reviews 2002, Issue 1*. Art. No.: CD001005. Wiley Online Library [en línea]. (fecha de consulta: 20 Noviembre 2014). Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001005/abstract>
- Gibson, P. G., Powell, H., Wilson, A., Abramson, M. J., Haywood, P., Bauman, A., ... Roberts, J. J. (2002b). *Self-management education and regular practitioner review for adults with asthma. Cochrane Database of Systematic Reviews 2002, Issue 3* Art. No.: CD001117. Wiley Online Library [en línea]. (fecha de consulta: 20 Noviembre 2014). Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001117/abstract>
- Gillissen, A. (2007). Patients, adherence in asthma. *Journal of Physiology and Pharmacology*, 58(5), 205–222.
- Glass, G. V., Macgaw, B. y Smith, M.L. (1981). *Meta-analysis in social research*, Newbury Park, California: Sage.
- Global Initiative for Asthma GINA. (2014). *Estrategia global para el manejo y prevención del asma. Revisión Agosto del 2014. Versión en Español*. GINA [en línea]. (fecha de consulta: 17 Octubre 2014). Disponible en: [http://www.ginasthma.org/documents/4/documents\\_variants/16](http://www.ginasthma.org/documents/4/documents_variants/16)
- Godoy, A. (1991). El proceso de la evaluación conductual. En V. Caballo (Ed.) *Manual de técnicas de terapia y modificación de conducta* (pp. 91-108). Madrid: Siglo XXI.
- Golberg, D. y Huxley, P. (1992). *Common mental disorders. A biosocial model*. Londres: Routledge.
- Goldfried, M. R. y Wolfe, B. E. (1996). Psychotherapy practice and research: Repairing a strained alliance. *American Psychologist*, 51(10), 1007-1016.

- Goldfried, M. R. y Wolfe, B. E. (1998). Toward a more clinically valid approach to therapy research. *Journal of Consulting and Clinical Psychology*, 66(1), 143–150.
- Gol-Freixaak, J. M. y Del Llano J. E. (2012). El fenómeno de la “Medicina Basada en la Evidencia”. En J. Millán y J.E. Del Llano Señarís (Coord.). *Ser médico. Los valores de una profesión* (p. 289-307). Madrid: Unión Editorial.
- Gómez J. M. y Roussos A. (2012). ¿Cómo sabemos si nuestros pacientes mejoran? Criterios para la significación clínica en psicoterapia: Un debate que se renueva. *Revista Argentina de clínica psicológica*. 21(2), 173-190.
- Gómez, F., Vilar, J. y Tejero, M. (2003). *Seis Sigma*. Madrid: Fundación Confemetal Editorial.
- Gómez-Benito, J., Hidalgo, M.D. y Guilera, G. (2010). El sesgo de los instrumentos de medición. Tests justos. *Papeles del Psicólogo*, 31(1), 75-84.
- González, M. A. y Condón M. J. (2001) Evaluación de tecnologías médicas como parte de la Medicina Basada en la Evidencia. *Rehabilitación(Madrid)*, 35(6), 337-342.
- González, M. T. y Botella, J. (2006). Comparación en te índices de tamaño del efecto para variables dicotomizadas en Meta-análisis. *Psicológica*, 27(2), 269-293.
- González. M. (2007). *Instrumentos de Evaluación Psicológica*. La Habana: Ciencias Médicas.
- González-Pardo, H. y Pérez-Álvarez, M. (2007). *La invención de los trastornos mentales ¿Escuchando al fármaco o al paciente?* Madrid: Alianza Editorial.
- González-Pinto, A., López-Peña, P. y Zorrilla-Martínez, I. (2009). Depresión unipolar y otras depresiones. En T. Palomo y M. A. Jiménez (Eds.), *Manual de psiquiatría* (pp. 324-343). Madrid: Ene Life Publicidad S.A. y Editores.
- Goodwin R. D., Jacobi F. y Thefeld, W. (2003). Mental disorders and asthma in the community . *Archives of General Psychiatry*, 60(11), 1125-1130.
- Goodwin, L. D. y Leech, N. L. (2003). The meaning of validity in the new Standards for Educational and Psychological Testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, 36, 181-191.
- Gregory, R. J. (2001). *Evaluación psicológica. Historia, principios y aplicaciones*. México: Manual Moderno.
- Grol, R. y Grimshaw, J. (2003). From best evidence to best practice: effective implementation of change in patients care. *The Lancet*, 362(9391), 1225-1230.

- Grupo Español del Estudio Europeo del Asma. (1996). Estudio Europeo del Asma. Prevalencia de hiperreactividad bronquial y asma en adultos Jóvenes de cinco áreas españolas. *Medicina Clínica (Barcelona)*, 106 (20), 761-767.
- Guadarrama, L. Escobar A. y Zhang, L. (2006). Bases neuroquímicas y neuroanatómicas de la depresión. *Revista de la Facultad de Medicina de la Universidad Nacional Autónoma de México*, 49(2), [en línea]. (fecha de consulta: 23 Diciembre 2014). Disponible en: <http://www.revistas.unam.mx/index.php/rfm/issue/view/466/showToc>
- Guárdia J., Freixa M., Però M. y Turbany J. (2008). *Análisis de datos en psicología. 2ª edición*. Madrid: Delta Psicología.
- Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6(4), 427-439.
- Guion, R. M. (1980). On triniarian doctrin as of validity. *Professional Psychology*, 11(3), 385-398.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley.
- Guyatt, G. H., Bombardier, C., Tugwell, P. X. (1986). Measuring disease specific-quality of life in clinical trials. *Canadian Medical Association Journal*, 134 (8), 889-895.
- Guyatt, G. H., Oxman, A. D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., et al., y GRADE Working Group. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336(7650), 924-926.
- Guyatt, G. H., y Rennie, D. (2002). *Users' guides to the medical literature: A manual for evidence-based practice*. Chicago: American Medical Association
- Haby, M. M., Donnelly, M., Corry, J., y Vos, T. (2006). Cognitive behavioural therapy for depression, panic disorder and generalized anxiety disorder: A meta-regression of factors that may predict outcome. *Australian and New Zealand Journal of Psychiatry*, 40(1), 9-19.
- Haddix A. C., Teutsch, S. M. y Corso P. S. (2002) *Prevention Effectiveness. A guide to decisión analysis and economic evaluation. Second Edition*. Oxford: Oxford University Press.
- Hageman, W. J. y Arrindell, W. A. (1999a). Establishing clinically significant change: increment of precision and the distinction between individual and group level of analysis. *Behaviour Research and Therapy*, 37(12), 1169-1193.

- Hageman, W. J. y Arrindell, W. A. (1999b). Clinically significant and practical! Enhancing precision does make a difference. Reply to McGlinchey and Jacobson, Hsu, and Speer. *Behavior Research and Therapy*, 37(12), 1219-1233.
- Hankin, B., Fraley, C., Lahey, B., Waldman, I. (2005). Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology*, 114(1), 96-110
- Hansen, N., Lambert, M. y Forman, E. (2002). The psychotherapy dose-response effect and its implication for treatment delivery services. *Clinical Psychology: Science and Practice*, 9(3), 329-343.
- Hardin J. y Hilbe, J. (2003) *Generalized Estimating Equations*. London: Chapman and Hall.
- Haro, J. M., Palacín, C., Vilagut, G., Martínez, M., Bernal, M., Luque, I., Codony, M., Doiz, M., Alonso, J., y el Grupo ESEeD-España. (2006). Prevalencia de los trastornos mentales y factores asociados: resultados del estudio ESEMeD-España. *Medicina Clínica*, 126 (12) 445-451.
- Harris, R. P., Helfand, M., Woolf S. H., Lohr K. N., Mulrow, C. D., Teutsch, S. M., Atkins D., y The Methods Work Group Third U.S. Preventive Services Task Force. (2001). Current methods of the U.S. Preventive Services Task Force: a review of the process. *American Journal of Preventive Medicine*, 20(Sup. 3), 21-35.
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S. y Olds C. C. (1992). Summarizing Monte Carlo results in methodological research: The one and two factor fixed effects ANOVA cases. *Journal of Educational Statistics*, 17(4), 315-339.
- Havelka, M., Lucanin, J. y Lucanin, D. (2009). Biopsychosocial model the integrated approach to health and disease. *Collegium Antropologicum*, 33(1), 303-310.
- Hayes, W. L. (1963). *Statistics for Psychologists*. New York: Holt, Rinehart and Winston
- Hayes, A. M., Laurenceau J. P. y Cardaciotto L. A. (2008). Methods for capturing the process of change. En A.M. Nezu y Ch.M. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 335 - 358). New York: Oxford University press.
- Hayes, S. C. y Greff, J. (2001). Factors promoting and inhibiting the development and use clinical practice guidelines. *Behavior Therapy*, 32(2), 211-217.

- Haynes, R. B., Ackloo E., Sahota, N., McDonald, H. P., y Yao, X. (2008). *Interventions for enhancing medication adherence (Review)*. *The Cochrane Library, Issue 4*. Chichester, UK: John Wiley & Sons.
- Haynes, S. N. y Heiby, E. (2004). Introduction to Behavioral Assessment. En M. Hersen (Ed.), *Compressive Handbook of psychological Assessment Vol.3: Behavioral Assessment* (pp. 3-18). Nueva York: John Wiley & Son.
- Hedeker D. y Gibbons R.D. (2006). *Longitudinal Data Analysis*. New York: Wiley.
- Hedges L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of educational statistics*, 6(2), 107-128.
- Henton, I. (2012). Practice-based research and counselling psychology: A critical review and proposal . *Counselling Psychology Review*, 27(3), 11-28.
- Henton, I. y Midgley, N. (2012). ‘A path in the woods’: Child psychotherapists’ participation in a large randomised controlled trial. *Counselling and Psychotherapy Research*, 12(3), 1–10.
- Hernández, J., Santacreu, J. y Rubio, V. (1999). Evaluación de la personalidad: una alternativa teórico-metodológica. *Escritos de Psicología*, 1 (3), 20-28.
- Hernández, M. (2007). *Epidemiología. Diseño y análisis de estudios*. México: Editorial Médica Panamericana.
- Hernández-Ávila, M., Garrido, F., Salazar-Martínez, E. (2000). Sesgos en estudios epidemiológicos. *Salud Pública de México*, 42(5), 438-446.
- Hersen, M. (2004). *Comprehensive Handbook of Psychological Assessment, Vol.1*. Nueva York: Wiley.
- Hersen, M y Bellack, A. S. (1978). *Behavior Therapy in the Psychiatric Setting*, Baltimore: Williams and Wilkins.
- Hersen, M. (2004). *Comprehensive Handbook of Psychological Assessment, Vol.1*. Nueva York : John Wiley & Son.
- Hicks, C. y Hennessy, D. (1997). Mixed messages in nursing research: their contribution to the persisting hiatus between evidence and practice. *Journal of Advanced Nursing*, 25(3), 595–601.
- Hill, C. (2006) Introducton to special section on therapist effects. *Psychoterapy Research*, 16(2),143-144

- Hiller, W. y Schindler, A. (2011). Response und Remission in der Psychotherapieforschung/ Response and Remission in Psychotherapy Research. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 61 (3-4), 170-176.
- Hiller, W., Schindler, A. C. y Lambert, M. J. (2012). Defining response and remission in psychotherapy research: A comparison of the RCI and the method of percent improvement. *Psychotherapy Research*, 22(1), 1-11.
- Hilsenroth, J. (2013). Introduction to the 50th Anniversary Special Issue on Psychotherapy Outcome: A Return to the Beginning. *Psychotherapy*, 50(1), 1-2.
- Hollander M. D. y Simeon D. (2004). *Guía de los trastornos de ansiedad*. Madrid: Elvsevier España.
- Hollon, S. (1996). The efficacy and effectiveness of psychotherapy relative to medications. *American Psychologist*, 55(1), 1025-1030.
- Holmbeck G. N. y Devine K. A. (2009). Editorial an author's checklist for measure development and validation manuscripts. *Journal Pediatric Psychology*, 34(7), 691-696.
- Horney, K (1964). *New ways in psychoanalysis*. Oxford: W. W. Norton and Company.
- Horvath, A. (2005). The therapeutic relationship: Research and theory. *Psychotherapy Research*, 15(1), 3-7.
- Hotopf, M., Churchill, R. y Lewis, G. (1999) Pragmatic randomised controlled trials in psychiatry. *The British Journal of Psychiatry*, 175(3), 217-223.
- Howlin, P. (1997). *When is a significant change not significant?* [Carta Editor]. *Journal of Autism and Developmental Disorders*, 27(3), 347-348.
- Hsu, L. M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment*, 11(4), 459- 467.
- Hsu, L. M. (1995). Regression toward the mean associated with measurement error and identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology*, 63(1), 141-144.
- Huber, M., Knottnerus, J. A., Geen, L. van der Horst, H., Jadad, A. R., Kromhout, D., Leonard, B., Loring, K. et al. (2011). How we define health?. *British Medical Journal*, 343, d4163.

- Hunsley, J., Dobson, K. S., Johnston, C., y Mikail, S. F. (1999). Empirically supported treatments in psychology: Implications for Canadian professional psychology. *Canadian Psychology / Psychologie canadienne*, 40(4), 289-301.
- Hunt E. (1996). Errors in Seligman's "The effectiveness of psychotherapy: The Consumer Reports study". *American Psychologist*, 51(10) 1082-1082.
- Hurtado, J. (2007). *El proyecto de Investigación. Quinta edición*. Venezuela Caracas: Ediciones Quirón-Sypal.
- Imel, Z., y Wampold, B. (2008). The Importance of Treatment and the Science of Common Factors. En T.R. Bronw S. D. y R. W. Cuaresma (eds.), *Handbook of counseling Psychology, 4th ed.* (pp. 249-262). Hoboken, New Jersey: John Wiley & Sons.
- International Committee of Medical Journal Editors. (2013). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. [en línea], (fecha de consulta: 12 Septiembre 2014). Disponible en: <http://www.icmje.org/icmje-recommendations.pdf>
- Ioannidis, J.P. (2005). Why Most Published Research Findings Are False. *Public Library of Science Medicine*, 2(8), 696-701.
- Iraurgi, I. (2000). Cuestiones metodológicas en la evaluación de resultados terapéuticos. *Trastornos Adictivos*, 2(2), 99-113.
- Iraurgi, I. (2009a). Evaluación de resultados clínicos: Entre la significación estadística y la relevancia clínica. *Norte de Salud Mental*, 8(33), 94-108.
- Iraurgi, I. (2009b). Evaluación de resultados clínicos II: Las medidas de la significación clínica o los tamaños del efecto. *Norte de Salud Mental*, 8(34), 94-110.
- Iraurgi, I. (2010). Evaluación de resultados clínicos III: Índices de Cambio Fiable (ICF) como estimadores del cambio clínicamente significativo. *Norte de Salud Mental*, 8(36), 105-122.
- Iraurgi, I., Laespada, M.T., (2011). *Nuevos enfoques en el tratamiento de las drogodependencias*. Bilbao: Publicaciones Universidad de Deusto.
- Iraurgi, J. A., Hormaechea, J. M., Jiménez-Lerma, J. y Landabaso, M. (2009). Valoración de resultados en psicoterapia: datos preliminares de la adaptación y características del OQ-45 *Revista de Psiquiatría y Salud Mental*, 2(NºEsp.2009), 4. [en línea], (fecha de consulta: 5 Octubre 2013). Disponible en: [www.sepsiq.org/descargas/2/8](http://www.sepsiq.org/descargas/2/8)

- Ishikawa, K. (1986). *¿Qué es el control total de la calidad? La modalidad japonesa*. Bogotá: Norma.
- Jackson, S. (2010) *Research Methods: A Modular Approach. 2º Edition*. Belmont, California: Wadsworth Cengage Learning.
- Jacobson, E. (1971). *Depression*. New York: International University Press.
- Jacobson, N. S., Follette, W. C., y Revenstorf, D. (1986). Toward a standard definition of clinically significant change [Carta Editor]. *Behavior Therapy*, 17(3), 309-311.
- Jacobson, N. S., Roberts, L. J., Berns, S. B. y McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67(3), 300-307.
- Jacobson, N. S., y Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19.
- Jacobson, N. S., Follette, W. C., y Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15(4), 336-352.
- Jáuregui, I. y Tejedor, M. A. (2004). Asma y factores psicosociales. *Alergología e Inmunología Clínica*, 19(3), 101-109.
- Jiménez, C. R. (2012). The origin, truths, fallacies of the evidence-based medicine and its acceptance in Cuba. *Revista de Salud Pública*, 35(5), 702-713.
- Jiménez, J. (2002). *Manual de gestión para jefes de servicios clínicos*. Madrid: Ediciones Díaz de Santos.
- Jiménez-Navarro, M. F., Cabrera-Bueno, F., Muñoz-García, A. J. y Teresa-Galván, E. (2009). Ensayos clínicos y práctica clínica en el “mundo real”. ¿Conocemos los factores de confusión entre la eficacia y la efectividad) (Carta al Editor). *Revista Española de Cardiología*, 62(3), 328-336.
- Joint Commission on Accreditation of Healthcare Organizations. (2005) *Comprehensive Accreditation Manual for Hospitals: The Official Handbook*. Oakbrook Terrace, Illinois: Joint Commission Resources.

- Jovell A. J. y Navarro-Rubio, M. D. (1995). Evaluación de la evidencia científica. *Medicina Clínica (Barcelona)*, 105(19):740-743.
- Juárez, F. (2011). El concepto de salud: Una explicación sobre su unicidad, multiplicidad y los modelos de salud. *International Journal of Psychological Research*, 4(1), 70-79.
- Kahn, R. L., y Juster, F. (2002). Well-Being: Concepts and Measures. *Journal of Social Issues*. 58(4), 627-644
- Kane, M. T. (2006). Validation. En R.L. Brennan (Ed.), *Educational Measurement*. (pp. 17-64). Westport, Connecticut: Praeger.
- Kanfer, F. H., y Saslow, G. (1965). Behavior analysis. *Archives of General Psychiatry*, 12(6), 529-538.
- Kazdin, A. E. (1978). *History of behavior modification: experimental foundations of contemporary research*. Baltimore: University Park Press.
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology*, 67(3), 332–339.
- Kazdin, A. E. (2001). Almost Clinically Significant ( $p < .10$ ): Current Measures May Only Approach Clinical Significance. *Clinical Psychology: Science and Practice*, 8(4), 455-462.
- Kazdin, A. E. (2001). *Métodos de investigación en Psicología Clínica* (3ª ed.). México: Prentice Hall.
- Kazdin, A. E. (2003). *Evidence-based psychotherapies for children and adolescents*. New York: Guilford Press.
- Kazdin, A. E. (2008). Evidence-Based Treatment and Practice new Opportunities to bridge clinical research and practice, enhance the knowledge base, and improve Patient Care. *American Psychologist*, 63(3), 146–159.
- Kazdin A. E. (2011). Evidence-based treatment research: Advances, limitations, and next steps. *American Psychologist*, 66(8), 685-98.
- Kazdin, A. E. y Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44(8), 1116–1129.

- Kazdin, A. E., y Whitley, M. K. (2006). Comorbidity, case complexity, and effects of evidence-based treatment for children referred for disruptive behavior. *Journal of Consulting and Clinical Psychology*, 74(3), 455-467.
- Kazis L. E., Anderson, J. J., Meenan, R. S. (1989). Effect sizes for interpreting changes in health status. *Medical Care*, 27(3), 178-189.
- Kelly, G. (1955). *The psychology of personal construct*. Nueva York: Norton.
- Kendall, P. C, Hollon, S. D., Beck, A. T., Hammen, C. L. e Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research*, 11 (3), 289-299.
- Kendall, P. C., Butcher, J. N. y Holmbeck, G. N. (1999). *Handbook of research methods in clinical psychology (2nd ed.)*. New York: Wiley.
- Kenny, D. A. y Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112(1), 165-172.
- Kernberg, O. (1976). *Object Relations Theory and Clinical Psychoanalysis*. New York: Jason Aronson.
- Kernick, D. P. (2003). Introduction to health economics for the medical practitioner. *Postgraduate Medical Journal*, 79(929), 147-150.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R. y Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6), 593-602.
- Kessler, R. C., Chiu, W. T., Demler, O., y Walters, E. E. (2005). Prevalence, severity and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6), 617-627.
- Kimmel, A. J. (1996) *Ethical issues in behavioral research*. Cambridge: Blackwell.
- Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, 61(2), 213-218.
- Klein, M. (1940). Mourning and its relation to manicdepressive states. En *Melanie Klein Contributions to Psychoanalysis, 1921-1945, Edición 1948*. (pp. 311-338) London: The Hogarth Press.

- Kleinbaum, D. G., Kupper, L., Morggenstern, H., (1982). *Epidemiologic Research. Principles and Quantitative Methods*. Belmont, California: John Wiley & Sons.
- Klennert M. D., Nelson H. S., Price M. R., Adinoff A.D., Leung D.Y. y Mrazek D.A. (2001) Onset and persistence of childhood asthma: predictors from infancy. *Pediatrics*, 108(4), 69.
- Kolbe J., Fergusson W., Vamos M. y Garrett J. (2000). Case-control study of severe life-threatening asthma in adults: demographics, health care, and management of the acute attack. *Thorax*, 55(12), 1007-1015.
- Korta, J., Valvere, J., Praena, M., Figuerola, M., Rodriguez, C. R., Rueda, S., Neira, A., Vazquez, C., Matínez, M., Román, J. M. (2007). La educación terapéutica del asma. *Anales de Pediatría (Barcelona)*, 66(5), 496-517.
- Kraus, D. R., Seligman, D. A., y Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology*, 61(3), 285–314.
- Krumholz, H. M. (2009). Outcomes Research: Myths and Realities [Editorial]. *Circulation Cardiovascular Quality Outcomes*, 2(1), 1-3.
- Kruskal, W. H. y Wallis W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621.
- Kuder, G. F. y Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160.
- Labrador, F. L., Echeburúa, E. y Becoña, E. (2000). *Guía para la elección de tratamientos psicológicos efectivos: hacia una nueva psicología clínica*. Madrid: Dykinson.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-test and ANOVAs. *Frontiers in Psychology*, 4(863), [en línea], (fecha de consulta: 15 Noviembre 2014). Disponible en: [www.journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/full](http://www.journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/full)
- Lambdin, C. (2012). Significance tests as sorcery: Science is empirical, significance tests are not. *Theory & Psychology*, 22(1), 67-90.
- Lambert M. J. y Hawkins E. J. (2004). Measuring Outcome in Professional Practice: Considerations in Selecting and Using Brief Outcome Instruments. *Professional Psychology: Research and Practice*, 35(5), 492-499.

- Lambert, M. J. (1980). The effects of psychotherapy: A professional update. *Journal of the association of mormon counselors and psychotherapist AMCAP*, 7(2), 19-36.
- Lambert, M. J. (1992). Implications of psychotherapy outcome research for eclectic and integrative psychotherapies. En J. C. Norcross y M. V. Goldfried (Eds.), *Handbook of Psychotherapy Integration* (pp. 143-189). New York: Basic Books.
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington, DC: American Psychological Association.
- Lambert, M. J. (2013). Outcome in Psychotherapy: The Past and Important Advances. *Psychotherapy*, 50(1), 42–51.
- Lambert, M. J. y Asay, T. P. (1984). Patient characteristics and their relationship to psychotherapy outcome. En M. Herson, L. Michelson y A. S. Bellack (Eds.), *Issues in psychotherapy research*, (pp 313-359). New York: Plenum Press.
- Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C. y Yanchar S.C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 3(4), 249–258.
- Lambert, M. J., Hansen, N. B. y Bauer, S. (2008). Assessing the clinical significance of outcome results. En A.M. Nezu y Ch.M. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 359-378) New York: Oxford University press.
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J. y Burlingame, G. M. (1996). Administration and Scoring Manual for the OQ-45.2. Brigham: American Professional Credentialing Services LLC
- Lambert, M. J. y Ogles, B. M. (2009). Using clinical significance in psychotherapy outcome research: the need for a common procedure and validity data. *Psychotherapy research*, 19(4-5), 493-501.
- Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S. y Shimokawa, K. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire) (3rd ed.)*. Wilmington, DE: American Professional Credentialing Services LLC.
- Lampropoulos, G. K., Goldfried, M. R., Castonguay, L. G., Lambert, M. J., Stiles, W. B., y Nestoros, J. N. (2002). What kind of research can we realistically expect from the practitioner? *Journal of Clinical Psychology*, 58 (10), 1241-1264.

- Landis, C. (1938). Statistical evaluation of psychotherapeutic methods. En S. E. Hinsie (Ed.), *Concepts and problems of psychotherapy* (pp. 155-165). London: Heineman.
- Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. En J. M. Shleien (Ed.), *Research in psychotherapy 3* (pp. 105-116). Washington, DC, USA: American Psychological Association.
- Lantz, B. (2013). The impact of sample non-normality on ANOVA and alternative methods. *British Journal of Mathematical and Statistical Psychology*, 66(2), 224-244.
- Laporte, J. R. (2001). *Principios básicos de investigación clínica. 2ª ed.* Barcelona: AstraZeneca.
- Larenas-Linnemann, D. (2009). Tratamiento del asma. *Neumología y Cirugía de Tórax*, 68(2), 149-162.
- Larner, G. (2004). Family therapy and the politics of evidence. *Journal of Family Therapy*, 26(1), 17-39.
- Last, J. M. (2007). *A dictionary of public health*. Oxford: Oxford University Press.
- Lavoie, K. L., Cartier, A., Labrecque, M., Bacon, S. L., Lemièrre, C., Malo, J. L., ... Ditto, B. (2005). Are psychiatric disorders associated with worse asthma control and quality of life in asthma patients?. *Respiratory Medicine*, 99(10), 1249-1257.
- Layard, R. (2006a). The case for psychological treatment centres. *British Medical Journal*, 332(7548): 1030-1032.
- Layard, R. (2006b). Happiness and Public Policy: a Challenge to the Profession. *The Economic Journal*, 116(510), 24-33.
- Layard, R., Bell, S., Clark, D. M., Knapp, M., Meacher, M. y Priebe, S. (2006). *The depression report: A new deal for depression and anxiety disorders*. London: School of Economics, Economic Performance Report.
- Layard, R., Clark, D.M., Knapp, M., y Mayraz, G. (2007). Cost-benefit analysis of psychological therapy. *National Institute Economic Review*, 202(1), 90-98.
- Lazarus R. S. y Folkman, S. (1986). *Estrés y procesos cognitivos*. Barcelona: Martínez Roca.
- Lee, S. J., Earle, C. C., Jane C. y Weeks, J. C. (2000). Outcomes Research in Oncology: History, Conceptual Framework, and Trends in the Literature. *Journal of the National Cancer Institute*, 92(3), 195-204.

- Lemos-Giráldez, S. (2010). *Psicopatología general*. Madrid: Síntesis.
- Lenfant, C. (2003). Clinical research to clinical practice. Lost in translation? *The New England Journal of Medicine*, 349(9), 868-874.
- Lenhard J. (2006). Models and Statistical Inference: The Controversy between Fisher and Neyman–Pearson. *British Journal for the Philosophy of Science*, 57(1), 69-91.
- León, O. y Montero, I. (2003). *Métodos de investigación en Psicología y Educación*. McGraw-Hill: Madrid.
- Leonhard, K. (1999). *Classification of Endogenous Psychoses and their Differentiated Etiology. Second Edition*. New York: Springer Verlag
- León-Rubio, J. M. (2004). Fundamentos de la psicología de la salud. En J. M. León, S. Medina, A. Ballesteros y I. M. Herrera (Eds.), *Psicología de la salud y de la calidad de vida* (pp. 15-39). Barcelona: Universitat Oberta de Catalunya.
- Lerner, J., Safren, S. A., Henin, A., Warman, M., Heimberg, R. G. y Kendall, P. C. (1999). Differentiating anxious and depressive self-statements in youth referred to an anxiety disorders clinic. *Journal of Clinical Child Psychology*, 28 (1), 82-93.
- Levene, H. (1960). Robust tests for equality of variances. En I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow y H. B. Mann (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. (pp. 278–292). Palo Alto California: Stanford University Press.
- Levine, T. R., Weber, R., Hullett, C., Park H. S. y Lindsey L. M. (2008). A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research. *Human Communication Research*, 34(2), 171-187
- Levine, T. y Hullet, C. (2002). Eta squared, partial eta squared, and misreporting of effect size in Communication Research. *Human Communication Research*, 28(4), 612-625.
- Levy, M.L., Robb, M., Allen, J., Doherty, C., Bland, J. M. y Winter R.J. (2000). A randomized controlled *evaluation* of specialist nurse education following accident and emergency department attendance for acute asthma. *Respiratory Medicine*, 94(9), 900-908.
- Lewinsohn, P. M. (1974). A behavioral approach to depression. En R. J. Friedman y M .M Katz (Eds), *The psychology of depression: Contemporary research* (pp. 157-185). New York: Wiley.

- Lewinsohn, P. M., Solomon, A., Seely J. R. y Zeiss, A. (2000). Clinical implications of subthreshold depressive symptoms. *Journal of abnormal psychology*, 109(2), 345-351.
- Lewis, M. S., Maruff, P., Silbert, B. S., Evered L. A., Scott D. A. (2007) The influence of different error estimates in the detection of postoperative cognitive dysfunction using reliable change índices with correction for practice effects. *Archives of clinical Neuropsychology*, 22 (2), 249-257
- Lilienfeld, S. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2(1), 53-70.
- Lindemann, H. (1990). *Entrenamiento autógeno. El mejor sistema de relajación*. Barcelona: Plaza y Janes Editores.
- Lindsay, G. (2009). Ética profesional y psicología. *Papeles del Psicólogo*, 30(3), 184-194.
- Lindsay, G., Koene, C., Ovreide, H. y Lang, F. (2008). *Ethics for European psychologists*. Gottingen, Germany and Cambridge: Hogrefe.
- Linehan, M.M., Comtois, K.A., Murray, A.M., Brown, M.Z., Gallop, R.J., Heard, H.L., et al. (2006). Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Archives of General Psychiatry*, 63 (7) 757-766
- Lipsey M. W. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. London: Sage.
- Lizondo G. y Llanque C. (2008). Depresión. *Revista Pacea de Medina Familiar*, 5(7), 24-31.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports* 3(Sup.9), 635-694.
- López, M. J., Marí-Dell'Olmo, M., Pérez-Giménez A. y Nebot, M. (2011) Diseños evaluativos en salud pública: aspectos metodológicos. *Gaceta Sanitaria*, 25(Sup.1), 9-16.
- López-Viña, A. (2005). Actitudes para fomentar el cumplimiento terapéutico en asma. *Archivos de Bronconeumología*, 41(6), 334-340.
- López-Viña, A. (2006) Asma grave y resistente al tratamiento: conceptos y realidades *Archivos de Bronconeumología*, 42(1), 20-25.
- Lora-Espinosa, A. (2005). Adherencia al tratamiento del asma en el paciente pediátrico y sus cuidadores. *Revista Pediátrica de Atención Primaria*, 7(2), 97-105.

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Losada, J. L. (1999). *Metodología observacional*. A Coruña: Penta.
- Lubin, B. y Zuckerman, M. (1999). *The Multiple Affect Adjective Check List—Revised* (3rd ed.). San Diego, California: Educational and industrial Testing Service.
- Luborsky, L. y DeRubeis, R. J. (1984). The use of psychotherapy treatment manuals: A small revolution in psychotherapy research style. *Clinical Psychology Review*, 4(1), 5-14.
- Luborsky L., Rosenthal R., Diguier L., Andrusyna T. P., Berman J. S., Levitt J. T., ... Krause E.D. (2002). The Dodo bird verdict is alive and well – mostly. *Clinical Psychology: Science and Practice*, 9(1), 2-12
- Luborsky, L., Singer, B. y Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that “everybody has won and all must have prizes?”. *Archives of General Psychiatry*, 32(8), 995-1008.
- Lueger, R. (2006). Technology to support the clinical management of psychotherapy cases: commentary on the Polaris- MH System. *Pragmatic Case Studies in Psychotherapy*, 2(3), 1-7.
- MacPherson, H. (2004) Pragmatic clinical trials. *Complementary Therapies in Medicine*, 12(2-3), 136-140
- Maloney, M. P. y Ward, M. P. (1976) *Psychological Assessment*. Nueva York: Oxford University.
- Manterola, C. y Pineda, V. (2008). El valor de “p” y la “significación estadística”. Aspectos generales y su valor en la práctica clínica. Interpretation of medical statistics. *Revista Chilena de Cirugía*, 60(1), 86-89.
- Manterola, C. y Zavado, D. (2009). Como interpretar los niveles de evidencia en los diferentes escenarios clínicos. *Revista Chilena de Cirugía*, 61(6), 582-595.
- Margni, A., Molina, M. F., Sartori, N., Rebull, M., Vazquez, L. (2012). Estudio descriptivo de resultados de tratamientos psicoterapéuticos en contexto natural en la ciudad autónoma de Buenos Aires, Argentina. *Revista argentina de clínica psicológica*, 21(1), 15-24.
- Marín, F. y Sánchez, J. (1996). Estimadores del tamaño del efecto en meta-análisis: un estudio Monte Carlo del sesgo y la eficiencia. *Psicología*, 17(3), 467-482.

- Marín, M. A. y Buisán, E. (1986). *Tendencias actuales en el diagnóstico pedagógico*. Barcelona: Laertes.
- Marks, D. F., Murray, M., Evans, B., Willig, C., Woodall, C. y Sykes, C. (2005). *Health psychology: Theory, research and practice*. London, UK: Sage.
- Marsá, B. M. y Aragona, M. (2009.) De la categoría a la dimensión: una mirada crítica a la evolución de la nosografía psiquiátrica. *Revista de la asociación española de neuropsiquiatría*, 29(103), 217-228.
- Martínez, E. (2010) Control del asma: un objetivo lejano. *Archivos de Bronconeumología*, 46(7), 347-348.
- Martínez, R. (1996) *Teoría de los tests psicológicos y educativos*. Madrid: Síntesis Psicológica.
- Martínez, S., Saiz, P. A., García-Portilla, M. P. (2008). Trastornos de ansiedad en atención primaria. *Jano Medicina y Humanidades*, 1714, 9-14.
- Martínez-Arias, M. R. (2007). *Psicometría*. Madrid: Alianza
- Martínez-Arias, M. R., Hernández, M. V., Hernández-Lloreda, M. J. (2006). *Psicometría*. Madrid: Alianza.
- Martínez-Monteagudo M. C., Ingles, J., Cano-Vindel, A. y García-Fernández, J. M. (2012). Estado actual de la investigación sobre la teoría tridimensional de la ansiedad de Lang. *Ansiedad y Estrés*, 18(2-3), 201-219.
- Martorell, M. C. (1988). *Técnicas de evaluación psicológica*. Valencia: Promolibro
- Matarazzo, J. D. (1980). Behavioral health and behavioral medicine: Frontiers for a new health psychology. *American Psychologist*, 35(9), 807-817.
- Matarazzo, J. D. (1982). Behavioral health challenge to academic, scientific, and professional psychology. *American Psychologist*, 37(1), 1-14.
- Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom. *American Psychologist*, 45(9), 999-1017.
- Matt, G. E. y Navarro, A. M. (1997). What meta-analysis have and have not taught us about psychotherapy effects: A review and future directions. *Clinical Psychology Review*, 17(1), 1-32.

- Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The annals of mathematical statistics*, 29, 204-209.
- McGlinchey, J. B., Atkins, D. C., y Jacobson, N. S. (2002). Clinical significance methods: Which one to use and how useful are they? *Behavior Therapy*, 33(4), 529-550.
- McGraw, K O. y Wong, S. P. (1992) A common language effect size statistic. *Psychological Bulletin*, 111(2), 361-365.
- McMillan, J. H. y Foley, J. (2011). Reporting and Discussing Effect Size: Still the Road Less Traveled? *Practical Assessment, Research and Evaluation*, 16(14), 1-12.
- Meeus, M. y Nijs, J. (2007). Central sensitization: a biopsychosocial explanation for chronic widespread pain in patients with fibromialgia and chronic fatigue síndrome. *Clinical Rheumatology*, 26(4), 103-115.
- Meherens, W. A. (1997). The consequences of consequential validity. *Educational measurement: Issues and Practice*, 16(2), 16-19.
- Mehrens, W. A. y Lehmann, I. J. (1984). *Measurement and Evaluation in Education and Psychology (3rd edition)*. New York: Holt, Rinehart and Winston.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1.012-1.027.
- Messick, S. (1989). Validity. En R.L. Linn (Ed.). *Educational measurement 3°Ed* (pp. 13-103). New York: American Council on Education.
- Messick, S. (1995) Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Meyer, G. P. y Archer, R. P. (2001) The hard science of Rorschach research: what do we go?. *Psychological Assessment*, 13(4), 486-502.
- Miller, M. R., Hankinson J., Brusasco V., Burgos F., Casaburi R., Coates A., ... ATS/ERS Task Force (2005). Standardisation of spirometry. *The European Respiratory Journal*, 26(2), 319-338.
- Millon, T., Krueger, R. y Simonsen, E. (2008). *Contemporary Directions in Psychopathology: Toward the DSM-V and ICD-11*. New York: Guilford Press.

- Minami, T., Davies, D. R., Tierney, S. C., Bettman, J. E., McAward, S. M., Averill, L. A., ... Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology*, 56(2), 309-320.
- Mira-Engo, J. M. (1990). *Psicología y salud. Un marco conceptual*. Valencia: Promolibro.
- Mitchell K., Truax Ch., Bozarth J. y Krauft C. (1973). *Antecedents to psychotherapeutic Outcome. NIMH Grant Report (12306)*. Arkansas: University of Arkansas, Rehabilitation Services Hot Spring.
- Mitjans, M. y Arias, B. (2012). La genética de la depresión: ¿ qué información aportan las nuevas aproximaciones metodológicas? *Actas Españolas de psiquiatría*, 40(2), 70-83.
- Moder, K. (2010) Alternatives to *F*-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study) *Psychological Test and Assessment Modeling*, 52(4), 343-353.
- Mogg, K., Mathew, A. y Weinman, J. (1999). Selective processing of threat cues in anxiety states: a replication. *Behaviour Research and Therapy*, 27(4), 317-323.
- Molina M. (2013) El significado de los intervalos de confianza. *Revista Pediatría de Atención Primaria* ,15(57), 91-94.
- Mollon, P. (2009). The NICE guidelines are misleading, unscientific, and potentially impede good psychological care and help. *Psychodynamic Practice: Individuals, Groups and Organisations*, 15(1), 9–24.
- Montero, I. y León, O. (2007) A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology*, 7(3), 847-862.
- Monterrey, P. (2012).  $p < 0,05$ , ¿Criterio mágico para resolver cualquier problema o leyenda urbana? *Universitas Scientiarum*, 17(2), 203-215.
- Montgomery, W. (2007). Ciencia conductual y aplicaciones clínicas en el presente: ¿por qué su eficacia?. *Revista Psicología Científica.com*, 9(13),[en línea], (fecha de consulta: 22 Junio 2013). Disponible en: <http://www.psicologiacientifica.com/ciencia-conductual-aplicaciones-clinicas>
- Moos, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.

- Morales, P. (2012). *El tamaño del efecto (effect size): análisis complementarios al contraste de medias*. Madrid: Universidad Pontificia Comillas. [en línea], (fecha de consulta: 12 Octubre 2014). Disponible en: <http://www.upcomillas.es/personal/peter/investigacion/Tama%F1oDelEfecto.pdf>
- Moreno, A., Moreno, L. y García, G. (2012). *Epidemiología y estadística en salud pública*. México: Mcgraw Hill.
- Moreno, C. (2003). *Evaluación psicológica. Concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia*. Madrid: Sanz y Torres.
- Morgan, P. y Canadian Task Force on the Periodic Health Examination. (1979). The periodic health examination. *Canadian Medical Association Journal* 121(9):1193-1254.
- Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364-386.
- Morrison, V. y Bennett, P. (2008). *Psicología de la salud*. Madrid: Pearson Prentice Hall.
- Mowrer, O. H. (1940). Anxiety-reduction and learning. *Journal of Experimental Psychology*, 27(5), 497-516.
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Muñoz-Solomando, A., Kendall, T., y Whittington, C. J. (2008). Cognitive behavioural therapy for children and adolescents. *Current Opinion in Psychiatry*, 21(4), 332-337.
- Murali, R., Hanson, M. D. y Chen, E. (2006). Psychological stress and its relationship to cytokines and inflammatory diseases. En R. E. Faith, A. J. Murgo, R. A. y Good, N. P. Plonikoff (Eds.), *Cytokines: Stress and immunity, Second Edition* (pp. 29-49). Boca Ratón, Florida: CERC Press, Taylor and Francis Group.
- Murphy, K. R. (1997). Editorial. *Journal of Applied Psychology*, 82(1), 3-5.
- Mustaca, A. E. (2004). Tratamiento psicológico eficaces y ciencia básica. *Revista Latinoamericana de Psicología*, 36(1), 11-20.
- Mustaca, A. E. y Bentosela, M. (1995). Estados psicológicos, salud y enfermedad. *Avances en Psicología Clínica Latino americana*, 13(1), 53-57.

- Myers, J. K., Weissman, M. M., Tischler, G. L., Holzer, C. E., Leaf, P. J., Orvaschel, H., ... Stoltzman, R. (1984). Six month prevalence of psychiatric disorders in three communities: 1980–1982. *Archives of General Psychiatry*, 41(10), 959–970.
- Naberan, K. (2002). Diferencias entre el asma y la enfermedad obstructiva crónica. *Atención Primaria*, 30(10), 648-654.
- Nakagawa, S. y Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, 82(4), 591-605.
- Nathan, P. E. y Gorman, J. M. (1998). *Treatment that work*. Oxford: Oxford University Press.
- National Center for Education Statistics NCES. (2002). *NCES statistical standards*. Washington, DC: U.S. Department of Education.
- National Center for Health Statistics, NCHS. y Centers for Medicare & Medicaid Services, CMS. (2011). ICD-9-CM stands for International Classification of Diseases, Ninth Revision, and Clinical Modification. [en línea]. (fecha de consulta: 15 Diciembre 2014). Disponible en: <http://www.cdc.gov/nchs/icd/icd9cm.htm>
- National Health, Lung and Blood Institute, NHLBI. (2007). *Guidelines for the Diagnosis and Management of Asthma* (EPR-3). [en línea]. (fecha de consulta: 20 Octubre 2014). Disponible en: <http://www.nhlbi.nih.gov/>
- National Institute for Clinical Excellence. NICE (2004a). *Depression: Management of depression in primary and secondary care (clinical guideline 23)*. London, UK: National Institute for Clinical Excellence.
- National Institute for Clinical Excellence. (2004b). *Anxiety: Management of anxiety (panic disorder, with and without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care (clinical guideline 22)*. London, UK: National Institute for Clinical Excellence.
- National Institute for Clinical Excellence. NICE (2005a). *Obsessive-compulsive disorder: Core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder (clinical guideline 31)*. London, UK: National Institute for Clinical Excellence.
- National Institute for Clinical Excellence. NICE (2005b). *Post-traumatic stress disorder (ptsd): The management of ptsd in adults and children in primary and secondary care (clinical guideline 26)*. London, UK: National Institute for Clinical Excellence.

- National Institute for Clinical excellence. NICE (2006). *Computerized cognitive behaviour therapy for depression and anxiety: Technology appraisal 97*. London, UK: National Institute for Clinical Excellence.
- National Institute of Mental Health NIMH e Insel T. (2013). *Transforming Diagnosis. 29 Abri 2013 Report*. [en línea]. (fecha de consulta: 15 Diciembre 2014). Disponible en: <http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml>
- Navarro, V. (1998). Concepto actual de la salud pública. En F. Martínez, P.L. Castellanos y V. Navarro (Eds.) *Salud Publica* (pp. 49-54). México D.F.: Mc Graw-Hill
- Nebot, C., Baqués, T., Crespo, A., Valverde, I. y Canela-Soler, J. (2004). La opinión de los usuarios como oportunidad de mejora en atención primaria. *Atención Primaria*, 35(6), 290-294.
- Nebot, M. y Victoria, C. (2007). Evaluación en salud pública: ¿todo vale? [Editorial]. *Gaceta Sanitaria*, 21(2), 95-96.
- Nebot, M. y Victoria, C. (2011). Valorando la efectividad de las intervenciones en salud pública: la fuerza de los «diseños evaluativos débiles». *Gaceta Sanitaria*, 25(Sup. 1), 1-2.
- Nebot, M., López, M. J. Ariza C., Villalbi J. R., García-Altes, A. (2011) Evaluación de la efectividad en salud pública: fundamentos conceptuales y metodológicos. *Gaceta Sanitaria*, 25(Sup.1), 3-8.
- Newbold, P., Carlson W.L. y Thorne, B.N. (2008). *Estadística para administración y economía*. Madrid: Pearson Educación.
- Newton J. y West E. (1996). Evidence-based Medicine and Compassion. Correspondence. *The Lancet*, 347(9018), 1839.
- Neyman, J, y Pearson, E. S. (1928a). On the use and interpretation certain test criteria for purposes of statistical inference (Part I). *Biometrika*, 20(1-2), 175-240.
- Neyman, J, y Pearson, E. S. (1928b). On the use and interpretation certain test criteria for purposes of statistical inference (Part II). *Biometrika*, 20A(1-2), 263-294.
- Neyman, J, y Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Science*, 231(694-706), 289-337.

- Nielsen, S. L., Smart, D., Isakson, R., Worthen, V., Gregersen, A. y Lambert, M. (2004). The Consumer Reports effectiveness ecore: What did consumers report? *Journal of Counseling Psychology*, 51(1), 25-37.
- Norcross, J. C. (2011). *Psychotherapy relationships that work: Evidence-based responsiveness (2nd ed.)*. New York: Oxford University Press.
- Norcross, J. C. y Wampold, B. E. (2011). What Works for Whom: Tailoring Psychotherapy to the Person. *Journal of clinical psychology in sesión*, 67(2), 127-132.
- Nunnally, J. C. y Bernstein, H. (1994). *Psychometric Theory, 3º. Ed.* New York: McGraw-Hill.
- Nunnally, J. C. y Kotsch, W. E. (1983). Studies of individual subjects: Logic and methods of analysis. *British Journal of Clinical Psychology*, 22(2), 83-93.
- Nuzzo, R. (2014). Statistical Errors. P value, the “gold standard” of statistical validity, are not as reliable as many scientists assume. *Nature*, 506 (7487), 150-152.
- Oblitas L. A. (2008). Psicología de la salud: una ciencia del bienestar y la felicidad. *Revista Avances en Psicología*, 16(1), 9-38.
- Oblitas, L. A. y Becoña E. (2000). *Psicología de la salud*. México D.F: Plaza y Valdés.
- Ochoa, C. (2010). Evaluación de la importancia de los resultados de estudios clínicos. Importancia clínica frente a significación estadística. *Evidencias en Pediatría*, 6(2), e40, [en línea], (fecha de consulta: 20 Septiembre 2014). Disponible en: <http://www.evidenciasenpediatria.es>
- Office of Strategic Services OSS Assessment Staff. (1948). *Assessment of men: Selection of personnel for the Office of Strategic Services*. New York: Rinehart & Company.
- Ogles, B. M., Lunnen, K. M., y Bonesteel, K. (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review*, 21(3), 421–446.
- O'Leary, K. D., y Wilson, G. T. (1987). *Behavior therapy: Application and outcome. Second Edition*. New Jersey: Prentice-Hall.
- Olejnik, S. y Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434-447.
- Olfson, M., Sing, M. y Schlesinger, H. G. (1999). Mental health/medical care cost offsets: opportunities for managed care. *Health Affairs (Millwood)*, 18(2), 79-90.

- Oliva, J., López J., Montejo, A., Osuna, R. y Duque, B. (2009). The socioeconomic costs of mental illness in Spain. *The European Journal of Health Economics*, 10(4), 361-369.
- Ollendick, T., Allen, B., Benoit, K., y Cowart, M. (2011). The tripartite model of fear in children with specific phobias: Assessing concordance and discordance using the behavioral approach test. *Behaviour Research and Therapy*, 49(8), 459- 465.
- Organización Mundial de la Salud / World Health Organization OMS/WHO (1948). *Constitución de la Organización Mundial de la Salud* [en línea]. (fecha de consulta: 20 Noviembre 2014). Disponible en: [http://www.who.int/gb/bd/PDF/bd46/s-bd46\\_p2.pdf](http://www.who.int/gb/bd/PDF/bd46/s-bd46_p2.pdf)
- Organización Mundial de la Salud / World Health Organization OMS/WHO (1985). *The principles of quality assurance, report on a WHO meeting*. Copenhagen: OMS.
- Organización Mundial de la Salud / World Health Organization OMS/WHO (1998). *Therapeutic patient education. Continuing education programmes for healthcare providers in the field of prevention of chronic diseases*. Copenhagen: WHO Regional Office for Europe.
- Organización Mundial de la Salud / World Health Organization OMS/WHO (2001). *Informe sobre la salud en el mundo, 2001: Salud mental: nuevos conocimientos, nuevas esperanzas*. Ginebra: OMS. [en línea], (fecha de consulta: 23 Diciembre 2014). Disponible en: <http://www.who.int/whr/2001/es/>.
- Organización Mundial de la Salud / World Health Organization OMS/WHO (2003). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la salud: Décima revisión, CIE-10*. México: Correo de la Unesco.
- Organización Mundial de la Salud / World Health Organization OMS/WHO (2005). *Mental health: facing the challenges, building solutions*. Copenhagen: WHO Regional Office for Europe. [en línea]. (fecha de consulta: 23 Diciembre 2014). Disponible en: [http://www.euro.who.int/\\_data/assets/pdf\\_file/0008/96452/E87301.pdf](http://www.euro.who.int/_data/assets/pdf_file/0008/96452/E87301.pdf)
- Organización Mundial de la Salud. OMS (2014). *Adolescents: health risks and solutions. Fact sheet N° 345 May 2014*. [en línea]. (fecha de consulta: 23 Diciembre 2014). Disponible en: <http://www.who.int/mediacentre/factsheets/fs345/en/>
- Orlinsky, D. y Russell L. (1994). Tradition and change in psychotherapy research: Notes on the fourth generation. En R. Russell (Eds.), *Reassessing Psychotherapy Research* (pp. 185-214). New York: Guilford.

- Osakidetza Servicio Vasco de Salud (2005) *Guía de práctica clínica sobre asma*. Gobierno Vasco Departamento de Sanidad. [en línea]. (fecha de consulta: 13 Octubre 2014). Disponible en: <http://www.avpap.org/gtvr/GPCasma.pdf>
- Osgood, C., Suci, G., Tannenbaum, P. (1957). *The measurement of meaning*. Champaign: University of Illinois Press.
- Otto, M. W., Pollack, M. H. y Maki, K. M. (2000). Empirically supported treatments for panic disorder: Costs, benefits, and stepped care. *Journal of Consulting and Clinical Psychology*, 68(4), 556-563.
- Oxford Centre for Evidence-based Medicine CEBM. (2011). *The Oxford 2011 Levels of Evidence*. [en línea], (fecha de consulta: 15 Enero 2013). Disponible en: <http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf>
- Oxman, A. D., Cook D. J., Guyatt, G. H. y Evidence-Based Medicine Working Group (1994) User's Guides to The Medical Literature VI. How to use an overview. *Journal of the American Medical Association*, 272 (17), 1367-1371.
- Pailhez G. y Bulbena, A. (2009). Trastornos de ansiedad, fobias y crisis de angustia. En T. Palomo y M. A. Jiménez (Eds.), *Manual de psiquiatría* (pp. 377-390). Madrid: Ene Life Publicidad S. A. y Editores.
- Pallares E. (2002). *La Ansiedad*. Bilbao: Mensajero.
- Pardo, A. Ruiz M.A. y San Martín, R. (2009). Análisis de datos en ciencias sociales y de la salud. Madrid: Síntesis.
- Pardo, A. y Ferrer, R. (2013). Significación clínica: falsos positivos en la estimación del cambio individual *Anales de psicología*, 29(2), 301-310
- Pardo, A. y San Martín, R. (2010) *Análisis de datos en ciencias sociales y de la salud. Volumen 2*. Madrid: Editorial Síntesis.
- Pardo, M. A. y Fuster, R. (2007). Medidas de resultados en las UFPE. En: R. Fuster, M.A. Pardo y J.M. Ventura (Eds.). *Manual para las unidades de atención farmacéutica a pacientes externos (UFPE) de servicios de farmacia hospitalaria* (p 53-70). Valencia: Sociedad Española de Farmacia Hospitalaria.
- Parker, I. (2011). Psicoanálisis lacaniano y terapia cognitivo-conductual: vacíos. *Teoría y crítica de la psicología*, 1, 83-94.

- Parry, G. (1996). *NHS Psychotherapy services in England*. London: Department of Health. National Health Service.
- Parry, G. D., Cooper, C. L., Moore, J. M., Yadegarfar, G., Campbell, M. J., Esmonde, L., Morice, A. H. y Hutchcroft, B. J. (2012). Cognitive behavioural intervention for adults with anxiety complications of asthma: prospective randomised trial. *Respiratory Medicine*, 106(6), 802-810.
- Parry, G., Castonguay, L.G., Borkovec, T.D. y Wolf, A. W. (2010). Practice research networks and psychological services research in the UK and the USA. En M. Barkham, G.E. Hardy y J. Mellor- Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp.311–325). Oxford: Wiley-Blackwell.
- Pascoe, G. (1983). Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and Program Planning*, 6(3-4), 185-210.
- Pascual, J., Frias, D. y García, F. (1996). *Manual de psicología experimental. Metodología de investigación*. Barcelona: Editorial Ariel
- Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, 31(2), 109–118.
- Pawlik, K. (1979). *Diagnosis del diagnostico*. Barcelona: Herder.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine (5th series)*, 50, 157–175.
- Pearson, K. (1911). On a correction needful in the case of the correlation ratio. *Biometrika*, 8 (1/2), 254-256.
- Peiró, S. (2002) La calidad de vida relacionada con la salud en la investigación y la práctica clínica. *Documentación Social*, 127, 67-95.
- Pekarik, G. (1993). Beyond effectiveness: Uses of consumer-oriented criteria en defining treatment success. En T.R. Giles (Ed), *Handbook of effective psychotherapy* (pp. 409-436). Nueva York: Plenum.
- Pelechano, V. (1976.) *Psicodiagnóstico* Madrid: Publicaciones UNED.
- Pelechano, V. (1988). Del Psicodiagnóstico clásico al análisis ecopsicológico. Valencia: Alfaplus.

- Peña, D. (2000). *Estadística. Modelos y métodos 1. Fundamentos. 2ª Edición*. Madrid: Alianza Universidad Textos.
- Peña, G., Cañoto, Y. y Santalla de Banderali, Z. (2006). *Una Introducción a la Psicología: La dinámica del Comportamiento: Motivación y Emoción*. Venezuela, Caracas: Casa Editora Publicaciones UCAB.
- Peñate, W. y Matud, P. (1996). Técnicas Subjetivas. En G. Buela-Casal, y J.C. Sierra (Eds.). *Manual de Evaluación Psicológica: Fundamentos, técnicas y aplicaciones* (pp. 455-491). Madrid: Siglo XXI.
- Perepletchikova, F., y Kazdin, A. E. (2005). Oppositional defiant disorder and conduct disorder. In K. Cheng & K. M. Myers (Eds.), *Child an adolescent psychiatry: The essentials* (pp. 73-88). Philadelphia: Lippincott Williams & Wilkins.
- Pérez, M. (1996). *Tratamientos psicológicos*. Madrid: Editorial Universitas,
- Pérez, M. (2012). La invención de los trastornos psicológicos y sus implicaciones para la validación de tratamiento. En Labrador F.J y Crespo M (Coords.), *Psicología basada en la evidencia* (pp. 37-45). Madrid: Pirámide
- Pérez, M. (2013). Anatomía de la psicoterapia: El diablo no está en los detalles. *Clínica Contemporánea*, 4(1), 5-28.
- Pérez, M., Fernández, J. R., Fernández, C. y Amigo, I. (2003a). *Guía de tratamientos psicológicos eficaces I: Adultos*. Madrid: Pirámide.
- Pérez, M, Fernández J. R., Fernández, C. y Amigo, I. (2003b). *Guía de tratamientos psicológicos eficaces II: Psicología de la Salud*. Madrid: Pirámide.
- Pérez-Gil, J.A., Chacón, S. y Moreno, R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez *Psicothema*, 12(2), 442-446.
- Perpiña, C. (2012). *Manual de la entrevista psicológica: saber escuchar, saber preguntar*. Madrid: Pirámide
- Persons, J. B. y Silberschatz, G. (1998). Are results of randomised controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology*, 66(1), 126–135.
- Pervin, L. A. (1980). *Personality: Theory, assessment, and research*. Nueva York: John Wiley & Son.

- Pervin, L. A. (1988). Personalidad: controversias, problemas y tendencias actuales. *Revista de Psiquiatría y Psicología Humanista*, 19(20), 73-99.
- Pichot, P. (1971). *Los Tests Mentales*. Buenos Aires: Paidós.
- Pineault, R. y Daveluy, C. (1988) *La planificación sanitaria: conceptos, métodos y estrategias*. Barcelona: Masson.
- Piqueras, J. A., Ramos, V., Martínez, A. E. y Oblitas, L. (2010). Emociones negativas y salud. *Avances en Psicología*, 18(1), 33-57.
- Plaza V., Álvarez, F.J., Casan P., Cobos, N., López, A., Llauger M.A. y Quintano J.A. y Comité Ejecutivo Comité Ejecutivo de la Guía Española para Manejo del Asma –GEMA- (2003) Guía Española para Manejo del Asma. *Archivos de Bronconeumología* ,39(5), 3-42.
- Pope, B. (1979). *The mental health interview: Research and application*. New York: Pergamon Press.
- Popham, W. J. (1997). Consequential validity: Right concern-wrong concept. *Educational measurement: Issues and practice*, 16(2), 9-13.
- Porta-Serra, M. Álvarz-Dardet, C., Bolúmar, F., Plasencia A., Velilla E. (1987) La calidad de la información clínica (I). validez. *Medicina Clínica (Barcelona)*, 89(17), 741-747.
- Posada A. y de Zubiría E. (2004) Aspectos psicológicos asociados con el asma. En E. de Zubiria, E. de Zubiria y A. de Zubiría-Salgado (Eds) *Asma Bronquial. 2ª Edición* (pp. 346-367) Bogotá D.C.: Editorial Medica Panamericana.
- Preacher, K. J. y Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, 16(2), 93-115.
- Preacher, K. J. y Kelley, K. (2012). On Effect Size. *Psychological Methods*, 17(2), 137-152.
- Prieto, G. y Delgado, A. (2010). Fiabilidad y validez. *Papeles del Psicólogo*, 31(1), 67-74.
- Primero, G. y Moriana J.A. (2011) Tratamientos con apoyo empírico. Una aproximación a sus ventajas, limitaciones y propuestas de mejora. *Revista de investigación en psicología*, 14(2), 190-207.
- Primo, J. (2003). Niveles de evidencia y grados de recomendación. *Enfermedad Inflamatoria Intestinal al día*, 2(2), 39-42.

- Pryzwansky, W. B. y Wendt, R. N. (1999). *Profesional and ethical issues in psychology*. London: William Warder Norton.
- Pyne, J. M., Rost, K. M., Farahati, F., Tripathi, S., Smith, J., Williams, D. K., ... Coyne, J. C. (2005). One size fits some: The impact of patient treatment attitudes on the cost-effectiveness of a depression primary-care intervention. *Psychological Medicine*, 35(6), 839-854.
- Rachman, S. J. y Wilson, G. T. (1980). *The effects of psychologicaltherapy: Second edition*. New York: Pergamon Press.
- Ramos-Cejudo, J. y Cano-Vindel, A. (2008). Tratamiento cognitivo-conductual de un caso de trastorno por ansiedad generalizada: El componente metacognitivo. *Ansiedad y Estrés*, 14(2-3), 305-319.
- Ramsey P. G., Carline J. D., Inui T. S., Larson E. B., LoGerfo J. P., Norcini J. J., Wenrich M. D. (1992). Changes over time in the knowledge base of practicing internists. *Journal of the American Medical Association*, 266(8), 1103–1107.
- Rapier, C.M. (1996). *An introduction to outcomes research*. Brookwood, Surrey UK: Brookwood Medical Publications.
- Rebasa, P. (2003). Entendiendo la “ $p < 0,001$ ”. *Cirugía española*, 73(6), 361-365.
- Regier, D. A, Narrow, W. E, Kuhl, E. A. y Kupfer, D. J. (2009), The conceptual development of DSM-V. *The American Journal of Psychiatry*, 166(6), 645-650.
- Rehm, L. P. (1977). A self-control modelo of depression. *Behavior Therapy*, 8(5), 787-804.
- Reig, A. (2005). ¿Qué debería saber un médico de psicología?. *Análisis y Modificación de Conducta*, 31(136), 273-312
- Reisner, A. D. (2005) The common factors, empirically validated treatments and recovery models of therapeutic change. *The Psychological Record*, 55(3), 377-399.
- Reiss, S. y Bootzin, R. R. (1985). *Theoretical issues in behavior therapy*. Orlando: Academic Press.
- Relman, A. (1998). Assessment and Accountability the third revolution in Medical Care. *New England Journal of Medicine*, 319(18), 1220-1222.
- Revicki, D. A. (2002). Consistent Patient-Reported Outcomes. *Value in Health*, 5(4), 295-296.

- Revicki, D. A. (2004) Patient assessment of treatment satisfaction: methods and practical issues. *Gut Journal*, 53(4), 40–44.
- Richards, D.A., y Suckling, R. (2008). Improving access to psychological therapy: the Doncaster demonstration site organizational model. *Clinical Psychology Forum*, 181, 9-16.
- Richards, D.A., y Suckling, R. (2009). Improving access to psychological therapies: phase IV prospective cohort study. *British Journal of Clinical Psychology*, 48(4), 377-396.
- Ries, F., Castañeda, C., Campos, M. C. y Del Castillo. O. (2012). Relaciones entre ansiedad-rasgo y ansiedad-estado en competiciones deportivas. *Cuadernos de Psicología del Deporte*, 12(2), 9-16
- Ritz, T., Wilhelm, F. H., Meuret, A. E., Gerlach, A. L. y Roth W. T.(2011). Airway response to emotion- and disease-specific films in asthma, blood phobia, and health. *Psychophysiology*, 48(1), 121-135.
- Rivera C. L., y Bernal, G. (2008). La medición en el cambio psicoterapéutico: hacia una práctica psicológica basada en la evidencia y la medición. *Revista Puertorriqueña de psicología*, 19, 223-244.
- Robles, J. C., Andreu, J. M. y Peña, M. E. (2002). SCL 90-R: Aplicación y análisis de sus propiedades psicométricas en una muestra de sujetos clínicos españoles. *Psicopatología Clínica, Legal y Forense*, 2(1), 5-19.
- Roca, J., Sanchis, J., Agusti-Vidal, A., Segarra, F., Navajas, D., Rodríguez-Roisin, R., ... Sans, S. (1986). Spirometric reference values for a Mediterranean population. *Bulletín Eurpéen Physiopathologie Respiratoire*, 22(3), 217-224.
- Rodríguez, A. (2004). La investigación de resultados y el futuro de la psicoterapia: alternativas a los tratamientos empíricamente validados. *Papeles del Psicólogo*, 25(87), 45-53.
- Rodríguez, E. (2005). Estadística y psicología: Análisis histórico de la inferencia estadística. *Revista Psicología Científica.com*, 7(4), [en línea], (fecha de consulta: 5 Agosto 2014). Disponible en: <http://www.psicologiacientifica.com/estadistica-y-psicologia>
- Rodríguez-Sutil, C. (1996). La ética de la devolución en el psicodiagnóstico clínico. *Papeles del Psicólogo*, 66, 91-94.
- Rodríguez-Trigo, G., Plaza, V., Picado, C. y Sanchis, J. (2008). El tratamiento según la guía de la Global Initiative for Asthma (GINA) reduce la morbimortalidad de los pacientes con asma de riesgo vital. *Archivos de Bronconeumología*, 44(4), 192-196.

- Rofail, D., Abetz, L., Viala, M., Gait, C., Baladi, J. F. y Payne, K. (2009). Satisfaction and adherence in patients with iron overload receiving iron chelation therapy as assessed by a newly developed patient instrument. *Value in Health*, 12(1), 109-117.
- Rogers, A., Maidman, J. y House, R. (2011). The bad faith of evidence-based practice: Beyond counsels of despair. *Therapy Today*, 22(6), 26–29.
- Roland, M. y Torgerson, D. J. (1998) Understanding controlled trials: What are pragmatic trials? *British Medical Journal*, 316(7127), 285.
- Romero, A. F. (2008) Factores atribucionales de la efectividad psicoterapéutica *Anales de psicología*, 24(1), 88-99.
- Ronk F. R., Hooke G. R. y Page A .C. (2012). How Consistent Are Clinical Significance Classifications. When Calculation Methods and Outcome Measures Differ? *Clinical Psychology: science and practice*, 19(2), 167-179.
- Rosander, A. C. (1992). *La búsqueda de la calidad en los servicios*. Madrid: Ediciones Díaz de Santos.
- Rosenberg, W. y Donald, A. (1995). Evidence based medicine: an approach to clinical problem-solving. *British Medical Journal*, 310 (6987), 1122-1126.
- Roshental, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research*, 21(4), 37-59.
- Rosnow, R. L. y Roshental, R. (2008). Assessing the effect size of outcome research. En A.M. Nezu y Ch.M. Nezu (Eds.), *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions* (pp. 335 -358) New York: Oxford University press
- Rosser, W. (2008). Bringing important research evidence into practice: Canadian developments. *Family Practice: The international Journal for Research*.;25 (Sup.1):38-43.
- Rossi, P. y Freeman, H. (1989). *Evaluation: A Systematic Approach*. 4th editon. California: Sage Publications.
- Rothman, K. J. y Greenland, S. (1998). *Modern Epidemiology*. Washington: Lippincott- Raven.
- Roussos A. J. (2007). El diseño de caso único en investigación en psicología clínica. Un vínculo entre la investigación y la práctica clínica. *Revista Argentina de Clínica Psicológica*, 16(3), 261-270.

- Rubio-Terrés, C., Sacristán, J. A., Badía, X., Cobo, E. y García-Alonso, F. (2004). Métodos utilizados para realizar evaluaciones económicas de intervenciones sanitarias. *Medicina Clinica (Barcelona)*, 122 (15), 578-583.
- Rudd, M. D. y Rajab, M. H. (1995). Specificity of the Beck Depression Inventory and the confounding role of comorbid disorders in a clinical sample. *Cognitive Therapy and Research*, 19(1), 51-68.
- Ruscio, J., Brown, T. A., y Ruscio, A. M. (2009). A taxometric investigation of DSM-IV major depression in a large outpatient sample: Interpretable structural results depend on the mode of assessment. *Assessment*, 16(2), 127-144.
- Ruscio, J., y Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology*, 109 (3), 473-487.
- Sackett, D. L (1997). Evidence-based medicine. *Seminars in Perinatology*, 21(1), 3-5.
- Sackett, D. L. y Rosenberg, W. M. (1995). The need for evidence-based medicine. *Journal of the Royal Society of Medicine*, 88(11), 620-624.
- Sackett, D. L. y Straus, S. E. (1998). Finding and Applying Evidence During Clinical Rounds. The "Evidence Cart". *Journal of the American Medical Association*, 280(15):1336-1338.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes R. B. y Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't [editorial]. *British Medical Journal*, 312(7023), 71-72.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., y Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2 ed.). New York: Churchill Livingstone.
- Sacristán, J.A., Badía, X. y Rovira, J. (1995). *Farmacoeconomía: evaluación económica de medicamentos*. Madrid: Editores Médicos.
- Safran, J. D. y Muran, J. C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide*. New York: The Guilford Press.
- Sakashita, C., Slade, T., y Andrews, G. (2007). An empirical analysis of two assumptions in the diagnosis of DSM-IV major depressive episode. *Australian and New Zealand Journal of Psychiatry*, 41(1), 17-23.

- Salgueiro, M., Buesa, I., Aira, Z., Montoya, P., Bilbao, J., y Azkue, J. (2009). Valoración de factores sociales y clínicos en el síndrome de fibromialgia. *Revista de Sociedad Española del Dolor*, 16(6), 323-329
- Sánchez R. O. y Ledesma R.D. (2009) Análisis psicométrico del inventario de Síntomas Revisado SCL 90-R en población clínica. *Revista Argentina de Clínica Psicológica*, 18 (3), 265-274.
- Sánchez-Bahillo, M., García-Marcos, L., Pérez-Fernández, V., Martínez-Torres, A. E. y Sánchez-Solís, M. (2009). Evolución de la mortalidad por asma en España, 1960–2005. *Archivos de Bronconeumología*, 45(3), 123-128.
- Sánchez-Meca, J., López-Pina J. A. y López J. A. (2009). Generalización de la fiabilidad: un enfoque metaanalítico aplicado a la fiabilidad. *Fisioterapia*, 31(6), 262-270.
- Sánchez-Meca, J. y Botella, J. (2010). Revisiones sistemáticas y meta-análisis: Herramientas para la práctica profesional. *Papeles del Psicólogo*, 31(1), 7-17.
- Sandín, B. (2008). El estrés. En A. Belloch, B. Sandín y F. Ramos (Eds.), *Manual de Psicopatología, Volumen 2*, (pp. 3-42). Madrid: McGrawHill.
- Sandín, B. y Chorot, P. (2008). Concepto y categorización de los trastornos de ansiedad. En A. Belloch, B. Sandín y F. Ramos (Eds.), *Manual de Psicopatología, Volumen 2*, (pp. 43-63). Madrid: McGrawHill.
- Sandler, J. y Joffe, W. G. (1965). Notes on Childhood Depression. *International Journal of Psychoanalysis*, 46, 88-96.
- Santacreu, J. (1991). Psicología clínica y psicología de la salud. *Revista de Psicología de la Salud*, 3(1), 3-20.
- Santibáñez, P. M., Roman, M. F. y Vinet, E. V. (2009) Efectividad de la psicoterapia y su relación con la alianza terapéutica. *Interdisciplinaria*, 26(2), 267-287.
- Santibáñez, P. M., Román M. F., Chenevard, C. L., Espinoza, A. E. Irribarra D. L. y Müller, A. (2008). Variables inespecíficas en psicoterapia. *Terapia psicológica*, 26(1), 89-98.
- Santiestaban, C. (1990) *Psicometría: teoría y práctica en la construcción de tests*. Madrid: Ediciones Norma.
- Sanz, J. y Vázquez, C. (1998). Fiabilidad, validez y datos normativos del inventario para la depresión. *Psicothema*, 10(2), 303-318.

- Sanz, J. y Vázquez, C. (2011). *Manual Inventario de Depresión de Beck-II. BDI-II*. Madrid: Pearson.
- Sanz, J., Perdigon, A. L. y Vázquez, C. (2003). The Spanish adaptation of Beck's Depression Inventory: Psychometric properties in the general population. *Clínica y Salud*, 14(3), 249-280.
- Satterfield, J. M., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., Walker, B. B. y Whitlock, E. P. (2009). Toward a Transdisciplinary Model of Evidence-Based Practice. *Milbank Quarterly*, 87(2), 368-390.
- Scheffé, H. (1953). A Method for Judging all Contrasts in the Analysis of Variance. *Biometrika*, 40(1-2), 87-104.
- Schlaepfer-Pedrazzini, L. y Infante-Castañeda, C. (1990). La medición de salud: perspectivas teóricas y metodológicas. *Salud Publica de México*, 32(2), 141-155.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, 1(2), 115-129.
- Schmitt, N. (1996). Uses and abuses of Coefficient Alpha. *Psychological Assessment*, 8 (4), 350-353.
- Schmitt, N., y Stults, D. N. (1986). Methodology review: Analysis of multitrait - multimethod matrices. *Applied Psychological Measurement*, 10(1), 1-22.
- Schulz, K. F., Altman, D. G., y Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340(c332), 698-702.
- Scottish Intercollegiate Guidelines Network SIGN y British Thoracic Society BTS. (2012). *British Guideline on the Management of Asthma: A national clinical guideline*. Edinburgh: Scottish Intercollegiate Guidelines Network.
- Segarra, G., Farriols N. y Palma C. (2011). Prevalencia, impacto y tratamiento psicológico grupal de los trastornos de ansiedad: una revisión en el ámbito de la atención primaria. *Aloma Revista de psicología, Ciències de l'Educació i de l'Esport*, 28, 209-224.
- Segerstrom, S. C. y Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin*, 130(4), 601-630.

- Seligman, M. E. (1975). *Helplessness: On depression, development, and death*. New York: Freeman.
- Seligman, M. E. (1981). *Indefensión*. Madrid: Debate.
- Seligman, M. E. (1995). The effectiveness of psychotherapy. The Consumer Reports Study. *American Psychologist*, 50(12), 965-974.
- Shadish, W. R., Matt, G. E., Navarro, A. M. y Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, 126(4), 512-529.
- Shafran, R. (2011). The appliance of science. *The Psychologist*, 24(11), 816–818.
- Shapiro, D. A., y Shapiro, D. (1982) Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin*, 92(3), 581-604.
- Shapiro, S. S. Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52 (3-4), 591–611.
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *American Psychologist*, 6 (2), 98–109.
- Sheldrick, R. C., Kendall, P. C. y Heimberg, R. G. (2001). The clinical significance of treatments: A comparison of three treatments for conduct disorder children. *Clinical Psychology: Science and Practice*, 8(4), 418-430 Jacobson,
- Sidney S., y Braman M. D. (2006). The global burden of asthma. *Chest Journal*, 130(1), 4-12.
- Sierra, J. C., Ortega V. y Zubeidat I. (2003). Ansiedad, angustia y estrés: tres componentes a diferenciar *Revista Mal-estar e Subjetividade/ Fortaleza*, 3(1), 10-59.
- Silva, F. (1986). La entrevista. En R. Fernández Ballesteros y J.A. Carrobbles (Eds), *Evaluación Conductual* (pp.185-212). Madrid: Pirámide.
- Silva, F. (1988). La evaluación psicológica como proceso. *Evaluación psicológica/Psychological Assessment*, 4(1), 31-50.
- Silva, F. (1989). *Evaluación conductual y criterios psicométricos*. Ediciones Piramide: Madrid.
- Silva, F. (1995). *Evaluación Psicológica en niños y adolescentes*. Madrid: Síntesis Psicológica.
- Silva, H. (2002). Nuevas perspectivas en la biología de la depresión. *Revista Chilena de neuro-psiquiatría*. 40 (1), 9-20.

- Singer, J. D. y Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Sinha, S., Peach, G., Poloniecki J. D., Thompson, M. M., Holt, P. J. (2012). Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes--systematic review and recommendations for reporting. *The European Journal of Public Health*, 23(1), 86-92.
- Sireci, S. G. (1998). The construct of content validity. En B.D. Zumbo, (Ed.), *Validity Theory and the Methods Used in Validation: Perspectives from the Social and Behavioral Sciences* (pp. 83-117). The Netherlands: Kluwer Academic Press.
- Skinner, B. F. (1957). The experimental analysis of behavior. *American Scientist*, 45, 344-371.
- Slade, T. y Watson, D. (2005). The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population, *Psychological Medicine*, 36(11), 1590-1600.
- Smith, M. L., y Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32(9), 752-760
- Smith, M. L., Glass, G. V., y Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Snedecor, G. W. (1934). *Calculation and Interpretation of Analysis of Variance and Covariance*. Ames, Iowa: Collegiate Press.
- Sociedad Española de Neumología y Cirugía Torácica SEPAR, Sociedad Española de Alergología e Inmunología Clínica SEAIC, Sociedad Española de Otorrinolaringología y Patología Cérvico-Facial SEORL PCF, Sociedad Española de Medicina de Familia y Comunitaria SEMFYC, Sociedad Española de Médicos de Atención Primaria SEMERGEN, et al. (2009). *Guía Española para el Manejo del Asma GEMA*. Madrid: Luzán 5 Ediciones
- Sociedad Española de Neumología y Cirugía Torácica, SEPAR. (2002). *Manual de procedimientos. Modulo 3. Procedimientos de la evaluación de la función pulmonar*. Madrid: Novartis.
- Somers, J. M., Goldner, E. M., Waraich, P., y Hsu, L. (2006). Prevalence and incidence studies of anxiety disorders: Review of the literature. *Canadian Journal of Psychiatry*, 51(2), 100-113.
- Soto, J. (2006). Valoración de los resultados en salud percibidos y comunicados por los pacientes: ¿necesidad o lujo?. *Anales de Medicina Interna (Madrid)*, 23(3), 136-138.

- Soto, J. (2007). Implicación de la investigación de resultados en salud en la mejora continua de la calidad asistencial del Sistema Nacional de Salud. *Anales de Medicina Interna (Madrid)*, 24(11), 517-519.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271-295.
- Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology*, 5(4), 417-426.
- Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. New York: Macmillan.
- Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology*, 60(3), 402-408
- Speer, D. C. y Greenbaum, P. (1995). Five methods for computing significant individual client change and improvement rates: Support for an individual growth curve approach. *Journal of Consulting and Clinical Psychology*, 63(6), 1044-1048.
- Spielberger, C. D. (1976.) *Anxiety Behavior*. New York: Academic Press.
- Spielberg, C. D., Gorsuch, R. L. y Lushene, R. E. (1982). *Cuestionario de ansiedad estado-rasgo: Manual*. Madrid: TEA.
- Stein, D. J. (2012). Dimensional or categorical: different classifications and measures of anxiety and depression *Medicographia*, 34(3), 270-275.
- Stemler, S. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4).[en línea] (fecha de consulta: 25 Abril 2013). Disponible en: <http://pareonline.net/getvn.asp?v=9&n=4>
- Stiles, W. B. (2013). The Variables Problem and Progress in Psychotherapy Research. *Psychotherapy*, 50(1), 33-41
- Stiles, W. B., Barkham, M., Mellor-Clark, J., y Connell, J. (2008). Effectiveness of cognitive-behavioral, person-centred, and psychodynamic therapies in UK primary-care routine practice: Replication in a larger sample. *Psychological Medicine*, 3(5), 677-688.

- Stiles, W. B., Honos-Webb, L. y Knobloch, L. M. (1999). Treatment process research methods. En P. Kendall (Ed.), *Handbook of research methods in clinical psychology* (pp.364-402). New York: Wiley.
- Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P. y Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology*, 71(6), 963-972.
- Strauss, B. M. y Kaechele, H. (1998). The writing on the wall: comments on the current discussion about empirically validated treatments in Germany. *Psychotherapy Research*, 8(2), 158-177.
- Streiner, D. (2003). Staring at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, 80 (1), 99-103.
- Strupp, H. H. (1963). The outcome problem in psychotherapy revisited. *Psychotherapy: Theory, Research and Practice*, 1(1), 1-13.
- Strupp, H. H. (1964). The outcome problem in psychotherapy: A rejoinder. *Psychotherapy: Theory, Research & Practice*, 1(3), 97-100.
- Student / Gosset, W. (1925). New table for testing the significance of observations. *Metron*, 5, 105-108.
- Stucki, G., Liang, M., Fossel, A. y Katz, J. (1995). Relative responsiveness of condition specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of Clinical Epidemiology*, 48(11), 1369-1378.
- Suárez, L., Guadalupe, J. y del Olmo, H. (2010). Aspectos psicológicos de asma. *Alergia, Asma e Inmunología*, 19(1), 18-22.
- Sullivan, S. H. (1968). *The Interpersonal Theory of Psychiatry*. Nueva York: W. W. Norton and Company.
- Szklo, M. y Nieto, J. (2003). *Epidemiología intermedia. Conceptos y aplicaciones*. Madrid: Díaz de Santos.
- Talavera, J. O. y Rivas-Ruiz R. (2011). Investigación clínica. Relevancia clínica. *Revista Médica del Instituto Mexicano del Seguro Social*, 49(6), 631-635.

- Tannenbaum, S. (1995). Getting there from here: evidentiary quandaries of the US outcomes movement. *Journal of Evaluation in Clinical Practice*, 1(2), 97-103.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatment: Report and recommendations. *The Clinical Psychologist*, 48(1), 3-23.
- Task Force on Promotion and Dissemination of Psychological Procedures. (1995). Training in and dissemination of empirically-validated psychological treatment: Report and recommendations. *The Clinical Psychologist*, 48(1), 3-23.
- Taylor, S. E. (2007). *Psicología de la salud*. México D.F: McGraw-Hill.
- Teachman, B. A., Drabick, D., Hershenberg, R., L., Vivian, D., Wolfe, B. E. y Goldfried, M. R. (2013). Bridging the gap between clinical research and clinical practice: Introduction to the special section. *Psychotherapy*, 49(2), 97-100.
- ten Brinke, A., Ouverkerk, M. E., Zwinderman, A. H., Spinhoven, P., Bel, E. H. (2001). Psychopathology in patients with severe asthma is associated with increased health care utilization. *American Journal of Respiratory and Critical Care Medicine*, 163(5), 1093-1096.
- Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology*, 30(1), 47-54.
- Thomas, M., Butron, A., Moffat M. y Cleland, J. (2011). Asthma and psychological dysfunction. *Primary Care Respiratory Journal*, 20(3), 250-256.
- Thorpe, K. E., Zwarenstein, M., Oxman, A. D., Treweek, S., Furberg, C. D., Altman, D. G., ... Chalkidou, K. (2009). A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Canadian Medical Association Journal*, 180(10), 47-57.
- To, T., Stanojevic, S., Ginette, M., Gershon A. S., Bateman, E. D., Cruz, A. A. y Boulet, L. P. (2012). Global asthma prevalence in adults: findings from the cross-sectional world health survey. *BioMed Central's Public Health*, 12(204), fecha de consulta: 17 Octubre 2014). Disponible en: [www.biomedcentral.com/1471-2458/12/204](http://www.biomedcentral.com/1471-2458/12/204)
- Tobón, S., Sandín, B., Vinaccia S. y Núñez A. C. (2008). La dispepsia funcional: aspectos biopsicosociales, evaluación y terapia psicológica. *Suma Psicológica*, 15(1), 199-306.
- Tomás, J. M., Oliver A. y Hontangas, P. (2000). Análisis factorial confirmatorio de segundo orden y matrices multirrasgo-multimétodo. *Psicothema*, 12(2), 534-539.

- Tortosa, F. y Civera, C. (2006). *Historia de la psicología*. Madrid: McGraw-Hill Interamericana.
- Towse, A. (1999). The National Institute for Clinical Excellence (NICE). *Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 5(5), 1-2.
- Traub, R. (1994) *Reliability for the Social Sciences: Theory and Applications*. Newbury: SAGE Publications.
- Trull, T. y Jerry, P. (2003). *Psicología clínica: Conceptos, métodos y aspectos prácticos de la profesión*. México: International Thomson Editores.
- Tukey, J. W. (1953). The problem of multiple comparisons. In *The Collected Works of John W. Tukey VIII. Multiple Comparisons* (pp. 1- 300). New York: Chapman and Hall.
- Turner, D., Schünemann, H., Griffith, L., Beaton, D., Griffiths, A., Critch, J. y Guyatt, G. (2010). The minimal detectable change cannot reliably re-place the minimal important difference. *Journal of Clinical Epidemiology*, 63(1): 28-36.
- U.S. Preventive Services Task Force USPSTF (2014). *The guide to clinical preventive services*. Agency of Healthcare Research and Quality. [en línea], (fecha de consulta: 20 Enero 2014). Disponible en: <http://www.uspreventiveservicestaskforce.org/Page/Name/tools-and-resources-for-better-preventive-care>
- Urrutia, I., Aguirre, U., Sunyer, J., Plana, E., Maldonado, J., Anto, J., ... Grupo Español ECRHS. (2007). Cambios en la prevalencia de asma en la población española del Estudio de Salud Respiratoria de la Comunidad Europea (ECRHS-II). *Archivos de Bronconeumología*, 43(8), 425-430.
- Vacha-Haase, T. y Ness, C. M. (1999). Statistical significance testing as it relates to practice: Use within Professional Psychology: Research and Practice. *Professional Psychology: Research and Practice*, 30(1), 104-105.
- Vacha-Haase, T., y Nilsson, J. E. (1998). Statistical significance reporting: Current trends and usages within MECD. *Measurement and Evaluation in Counseling and Development*, 31(1), 46-57.
- Vacha-Haase, T., y Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473-481.
- Valderas, J., Ferrer, M. y Alonso, J. (2005). Instrumentos de medida de calidad de vida relacionada con la salud y de otros resultados percibidos por los pacientes. *Medina Clínica (Barcelona)*, 125(Supl. 1), 56-60.

- Vallejo, J. (2011). *Introducción a la psicopatología y a la psiquiatría*. Barcelona: Masson.
- Vallejo, J. y Gastó, C. (2000) *Trastornos afectivos: ansiedad y depresión. Segunda edición*. Barcelona: Masson.
- Van Weel, C. (2002). General practice-research networks: Gateway to primary care evidence. *Medical Journal of Australia*, 177(2), 62–63.
- Vandenbroucke, I. P. (1996). Evidence based medicine and “Medicine d’Observation”. *The Journal of Clinical Epidemiology*, 49(12), 1335-1338.
- Varela, M. (2010). El concepto de Salud. Una nueva perspectiva. *Encuentros Multidisciplinares*, 12(34), 48-55.
- Varo, J. (1995). La calidad de la atención médica. *Medicina clínica*, 14 (104), 538-540.
- Vaske, J. J., Gliner, J. A. y Morgan, G. A. (2002). Communicating judgments about practical significance: Effect size, confidence intervals and odds ratios. *Human Dimensions of Wildlife*, 7(4), 287-300.
- Vázquez, C. y Nieto, M. (2005). Psicología (clínica) basada en la evidencia (PBE): una revisión conceptual y metodológica. En J.L. Romero (Ed.) *Psicópolis: paradigmas actuales y alternativos en la psicología contemporánea* (pp. 465-485). Barcelona: Kairos.
- Vázquez, C. y Sanz, J. (1997). Reliability and normative data of Spanish version Beck’s Depression Inventory. *Clinica y Salud*, 8(3), 403-422.
- Vázquez, C. y Sanz, J. (2008). Trastornos del estado de ánimo: Aspectos clínicos. En A. Belloch, B. Sandín y F. Ramos (Eds.), *Manual de Psicopatología, Volumen 2* (pp. 233-269). Madrid: McGrawHill.
- Vázquez, C., Hervás, G., Hernangómez L. y Romero N. (2010). Modelos cognitivos de la depresión: Una síntesis y una nueva propuesta basada en 30 años de investigación. *Behavioral Psychology/Psicología Conductual*, 18(1), 139-165.
- Vázquez-Cordero, C. (2007). Adherencia al tratamiento de mantenimiento en el asma. En Sociedad Española de Pediatría (Eds), *Curso para educadores de en asma* (pp. 13-18). La Laguna: Litomaype
- Ventegodt, S., Andersen, N. J., y Merrick, J. (2003). Holistic Medicine: Scientific Challenges. *The scientific world Journal*, 3, 1108-1116.

- Villalba L. (2010). Tratamiento de los trastornos depresivos desde una perspectiva psicodinámica. *Revista de psiquiatría de Uruguay*, 74(2), 169-178.
- Villalbí, J. R. y Tresserras, R. (2011). Evaluación de políticas y planes de salud. *Gaceta Sanitaria*, 25(Sup. 1), 17-24.
- Villar, J., Lizán, L., Soto, J. y Peiró, S. (2009). La satisfacción con el tratamiento Treatment satisfaction. *Atención Primaria*, 41(11), 637–645.
- Virués-Ortega, J. y Moreno-Rodríguez, R. (2008). Guidelines for clinical case reports in behavioral clinical Psychology. *International Journal of Clinical and Health Psychology*, 8(3), 765-777.
- Von-Bertalanffy, L. (1976). *La teoría general de los sistemas: fundamentos, desarrollo y aplicaciones*. Madrid: Ediciones Fondo de Cultura Económica.
- Wadeley, A. y Blasco, T. (1995). *La Ética en la investigación y en la práctica psicológica*. Barcelona. Ariel Psicología.
- Wagner, R., Silove, D., Joukhador, J., Manicavasagar, V., Marnane, C., y Kirsten, L. T. (2005). Characteristics of the first 1,000 clients attending an anxiety clinic in South West Sydney. *Australian Journal of Psychology*, 57(3), 180-185.
- Wales, J. A., Palmer, R. L., y Fairburn, C. G. (2009). Can treatment trial samples be representative? *Behaviour Research & Therapy*, 47(10) 893-896.
- Wampold, B. E. (2001a). Contextualizing psychotherapy as a healing practice: Culture, history, and methods. *Applied and Preventive Psychology*, 10(2), 69-86.
- Wampold, B. E. (2001b). *The great psychotherapy debate: Model, methods, and findings*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Wampold, B. E. (2010). *The basics of psychotherapy: An introduction to theory and practice*. Washington, DC: American Psychological Association.
- Wampold B. E. (2011). *Qualities and actions of effective therapists. Research suggests that certain psychotherapist characteristics are key to successful treatment*. American Psychological Association APA Education Directorate [en línea], (fecha de consulta: 4 Octubre 2014). Disponible en: <https://www.apa.org/education/ce/>
- Wampold, B. E. (2013). The Good, the Bad, and the Ugly: A 50-Year Perspective on the Outcome Problem. *Psychotherapy*, 50(1), 16-24.

- Ware, J. H. y Hamel M. B. (2011). Pragmatic trials guides to better patient care?. *The New England Journal of Medicine*, 364(18), 1685-1687.
- Weiner, B. K. (2008). Spine update. The biopsychosocial model and spine care. *Spine*, 33(2), 219-223.
- Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38(3-4), 330-336.
- Werbart, A., Levin, L., Andersson H., y Sandell, R. (2013). Everyday Evidence: Outcomes of Psychotherapies in Swedish Public Health Services. *Psychotherapy*, 50(1), 119-130.
- Westbrook, D. y Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy*, 43(10), 1243-1261.
- Westen, D., Novotny, C. M. y Thompson-Brenner, H. (2004). The Empirical Status of Empirically Supported Psychotherapies: Assumptions, Findings, and Reporting in Controlled Clinical Trials. *Psychological Bulletin*, 130(4), 631-663.
- Westen, D., y Morrison, K. (2001). A multi-dimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported treatments. *Journal of Consulting and Clinical Psychology*, 69(6), 875-889.
- Westmeyer, H. y Hageböck, J. (1992). Computer-assisted assessment: A normative perspective. *European Journal of Psychological Assessment*, 8(1), 1-16.
- Widakowich, Ch. (2012). El enfoque dimensional vs el enfoque categórico en psiquiatría: aspectos históricos y epistemológicos. *Revista Argentina de Clínica Neuropsiquiátrica*, 17(4), 365-374.
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics*, 1, 80-83.
- Wilder, J. (1957). The law of initial value in neurology and psychiatry, facts and problems. *Journal of Nervous and Mental Disease*, 125(1), 73-86.
- Wilhelm, F. y Roth, W. (2001). The somatic symptom paradox in DSM-IV anxiety disorders: Suggestions for a clinical focus in psychophysiology. *Biological Psychology*, 57(1-3), 105-140.

- Wilkinson, L., y American Psychological Association APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.
- Wilson, S. R., Strub, P., Buist, A. S., Knowles, B., Lavori, P. W., Lapidus, J., ... Better Outcomes of Ashtma Treatment (BOAT) Study Group (2010). Shared treatment decision making improves adhrece and outcomes in poorly controlled asthma. *American Journal of Respiratory adn Critical Care Medicine*, 181(6), 566-577.
- Wilson, G. T. (1996). Manual-based treatments: the clinical application of research findings. *Behavior Research and Therapy*, 34(4), 295-314.
- Wilton, N. K. y Slim, A. M. (2012). Application of the Principles of Evidence-Based Medicine to Patient Care. *Southern Medical Journal*, 105(3), 136-143.
- Windsor, R., Baranowski, T., Clark N., y Cutter, G. (1994). *Evaluation of health promotion, health education and disease prevention programs. Second edition*. California: Mayfield Co.
- Wise, E. A. (2004). Methods for analyzing psichotherapy outcomes: A review of clinical significance, reliable change and recommendations for future directions. *Journal of Pesonality Assessment*, 82(1), 50-49.
- Womack, J. (1992). *La máquina que cambió el mundo*. Madrid: McGraw-Hill.
- Woody, S. R., Weisz, J. y McLean, C. (2005). Empirically Supported Treatments: 10 years later. *The clinical Psychologist*, 5(4), 5-11.
- Wright J. G. (2000). Evaluating the outcome of treatment: shouldn't we be asking patients if they are better? *The Journal of Clinical Epidemiology*, 53(6), 549-553.
- Wyrwich, K. (2004). Minimal important difference thresholds and the stan-dard error of measurement: Is there a connection? *Journal of Biopharmaceutical Statistics*, 14(1), 97-110.
- Wyrwich, K. W., Tierney, W. M. y Wolinsky, F. D. (1999). Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *Journal of Clinical Epidemiology*, 52(9), 861-873.
- Yela, M. (1996). Unidad y diversidad de la psicología. *Psicothema*, 8(sup.1), 327-351.
- Yorke, J., Fleming, S. L. y Shuldham, C. M. (2009). *Psychological interventions or adults with asthma (Review)*. *The Cochrane Library, 2009 Issue 1*. Chichester, UK: John Wiley y Sons.

Yorke, J. y Shuldham, C. (2005). *Family therapy for asthma in children*. *Cochrane Database of Systematic Reviews Issue 2*. Art. No.: CD000089. Wiley Online Library, [en línea] (fecha de consulta: 20 Noviembre 2014). Disponible en: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD000089.pub2/abstract>

Yule, G. U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, 75(6), 579–652.

Zumbo, B. D. (2007). Validity: Foundational Issues and Statistical Methodology. En C.R. Rao y S. Sinharay (Eds.) *Handbook of Statistics, Vol. 26: Psychometrics*, (pp. 45-79). The Netherlands: Elsevier Science.



## Anexos

---

### Anexo 1. Sintaxis SPSS para el cálculos de los Índices de cambio Fiable

\* BDI - Estimadores de la media, DE y fiabilidad en el grupo normativo (Sanz y Vázquez, 1997).

```
COMPUTE BDI_Fiabilidad=.83.  
COMPUTE BDI_nor_M=7.5.  
COMPUTE BDI_nor_SD=6.7.  
COMPUTE BDI_clin_M= 30.93.  
COMPUTE BDI_clin_SD= 8.11.
```

\* *Método de JACOBSON Y TRUAX.*

```
COMPUTE BDI_PC= ((BDI_nor_M * BDI_clin_SD) + (BDI_clin_M * BDI_nor_SD)) / (BDI_clin_SD + BDI_nor_SD).  
COMPUTE BDI_EEMjt = BDI_nor_SD * SQRT(1 - BDI_Fiabilidad).  
COMPUTE BDI_EEDif = SQRT(2* (BDI_EEMjt*BDI_EEMjt)).  
COMPUTE BDI_RCI = (BDI2 - BDI1) / BDI_EEDif.  
RECODE BDI_RCI (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO BDI_RCIsig.  
COMPUTE BDI_PC_Cambio = BDI2 - BDI_PC.  
RECODE BDI_PC_Cambio (Lowest thru 0=1) (0.1 thru Highest=2) INTO BDI_PC_Cambio2.  
VALUE LABELS BDI_RCIsig      1 'Empeora'  
                        0 'No Cambio'  
                        -1 'Mejora'  
                / BDI_PC_Cambio2 1 'Funcional'  
                        2 'Disfuncional'.
```

```
EXECUTE.
```

```
COMPUTE JT_ICScat= 0.  
DO IF ((BDI_RCIsig = - 1) & (BDI_PC_Cambio2 = 1)).  
RECODE JT_ICScat (0=5).  
END IF.  
DO IF ((BDI_RCIsig = - 1) & (BDI_PC_Cambio2 = 2)).  
RECODE JT_ICScat (0=4).  
END IF.  
DO IF ((BDI_RCIsig = 0) ).  
RECODE JT_ICScat (0=3).  
END IF.  
DO IF ((BDI_RCIsig = 1) & (BDI_PC_Cambio2 = 1)).  
RECODE JT_ICScat (0=2).  
END IF.  
DO IF ((BDI_RCIsig = 1) & (BDI_PC_Cambio2 = 2)).  
RECODE JT_ICScat (0=1).  
END IF.
```

EXECUTE.

VALUE LABELS JT\_ICScat 1 'Empeorado'

2 'Empeora dentro de la funcionalidad'

3 'No cambio'

4 'Mejora dentro de la disfuncionalidad'

5 'Recuperado'.

EXECUTE.

\* *Método GLM - GULLISKEN-LORD-NOVICK.*

COMPUTE BDI\_GLNdif=(BDI2 - BDI\_nor\_M) - (BDI\_Fiabilidad \* (BDI1 - BDI\_nor\_M)).

COMPUTE BDI\_GLNerror= (BDI\_nor\_SD \* SQRT(1 - (BDI\_Fiabilidad \* BDI\_Fiabilidad))).

COMPUTE BDI\_GLN= BDI\_GLNdif/ BDI\_GLNerror.

RECODE BDI\_GLN (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO BDI\_GLNsig.

VALUE LABELS BDI\_GLNsig 1 'Empeora' 0 'No Cambio' -1 'Mejora'.

EXECUTE.

\* *Método EN - EDWARD-NUNNALLY.*

IF (Grupo = 1) DEpre = 8.11.

IF (Grupo = 1) DEpos = 7.41.

IF (Grupo = 2) DEpre = 4.15.

IF (Grupo = 2) DEpos = 4.54.

IF (Grupo = 1) Mpre = 30.93.

IF (Grupo = 1) Mpos = 18.66.

IF (Grupo = 2) Mpre = 5.25.

IF (Grupo = 2) Mpos = 5.20.

COMPUTE EN\_II = ((BDI\_Fiabilidad \* (BDI1 - Mpre)) + Mpre) - (2 \* DEpre \* SQRT(1 - BDI\_Fiabilidad)).

COMPUTE EN\_IS = ((BDI\_Fiabilidad \* (BDI1 - Mpre)) + Mpre) + (2 \* DEpre \* SQRT(1 - BDI\_Fiabilidad)).

COMPUTE EN=0.

DO IF (BDI2 <= EN\_II).

RECODE EN (0=-1).

END IF.

DO IF (BDI2 >= EN\_IS).

RECODE EN (0=1).

END IF.

VALUE LABELS EN -1 'Mejora'

0 'No Cambio'

1 'Empeora'.

EXECUTE.

\* Método HA - HAGEMANN-ARRINDELL.

IF (Grupo = 1) DEpre = 8.11.

IF (Grupo = 1) DEpos = 7.41.

IF (Grupo = 2) DEpre = 4.15.

IF (Grupo = 2) DEpos = 4.54.

IF (Grupo = 1) Mpre = 30.93.

IF (Grupo = 1) Mpos = 18.66.

IF (Grupo = 2) Mpre = 5.25.

IF (Grupo = 2) Mpos = 5.20.

IF (Grupo = 1) Rxxprepost = 0.754.

IF (Grupo = 2) Rxxprepost = 0.698.

EXECUTE.

\* Índice de cambio fiable.

COMPUTE Rxx\_pre = ((DEpre \* DEpre) - (BDI\_EEMjt\*BDI\_EEMjt)) / (DEpre \* DEpre).

COMPUTE Rxx\_pos = ((DEpos \* DEpos) - (BDI\_EEMjt\*BDI\_EEMjt)) / (DEpos \* DEpos).

COMPUTE Rdd = (((DEpre \* DEpre) \* Rxx\_pre) + ((DEpos \* DEpos) \* Rxx\_pos) - (2 \* DEpre \* DEpos \* 0.754)) /  
((DEpre \* DEpre) + (DEpos \* DEpos) - (2 \* DEpre \* DEpos \* Rxxprepost)).

COMPUTE HA\_ICFdif= ((BDI2 - BDI1) \* Rdd) + ((Mpos - Mpre) \* (1 - Rdd)).

COMPUTE HA\_ICFerror= SQRT(Rdd) \* BDI\_EEDif.

COMPUTE HA\_ICF = HA\_ICFdif / HA\_ICFerror.

\* Índice de cambio significativo.

COMPUTE VCP = (Mpre \* (BDI\_nor\_SD \* SQRT(BDI\_Fiabilidad))) +  
(BDI\_nor\_M \* (DEpre \* SQRT(BDI\_Fiabilidad))) /  
((BDI\_nor\_SD \* SQRT(BDI\_Fiabilidad)) + (DEpre \* SQRT(BDI\_Fiabilidad))).

COMPUTE HA\_ICS = ((Mpos + ((BDI2 - Mpos) \* Rxx\_pos)) - VCP) /  
( (SQRT(Rxx\_pos) \* BDI\_EEMjt)).

RECODE HA\_ICF (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO HA\_ICFsig196.

RECODE HA\_ICF (-1.65 thru 1.65=0) (1.651 thru Highest=1) (Lowest thru -1.651=-1) INTO HA\_ICFsig165.

VALUE LABELS HA\_ICFsig196 HA\_ICFsig165 1 'Empeora'

0 'No Cambio'

-1 'Mejora'.

RECODE HA\_ICS (Lowest thru 0=1) (0.1 thru Highest=2) INTO HA\_PC\_Cambio2.

EXECUTE.

DO IF ((HA\_ICFsig165 = - 1) & (HA\_PC\_Cambio2 = 1)).

RECODE HA\_ICScat165 (0=5).

END IF.

DO IF ((HA\_ICFsig165 = - 1) & (HA\_PC\_Cambio2 = 2)).

RECODE HA\_ICScat165 (0=4).

END IF.

DO IF ((HA\_ICFsig165 = 0) ).

RECODE HA\_ICScat165 (0=3).

END IF.

```

DO IF ((HA_ICFsig165 = 1) & (HA_PC_Cambio2 = 1)).
RECODE HA_ICScat165 (0=2).
END IF.
DO IF ((HA_ICFsig165 = 1) & (HA_PC_Cambio2 = 2)).
RECODE HA_ICScat165 (0=1).
END IF.
EXECUTE.
DO IF ((HA_ICFsig196 = -1) & (HA_PC_Cambio2 = 1)).
RECODE HA_ICScat196 (0=5).
END IF.
DO IF ((HA_ICFsig196 = -1) & (HA_PC_Cambio2 = 2)).
RECODE HA_ICScat196 (0=4).
END IF.
DO IF ((HA_ICFsig196 = 0) ).
RECODE HA_ICScat196 (0=3).
END IF.
DO IF ((HA_ICFsig196 = 1) & (HA_PC_Cambio2 = 1)).
RECODE HA_ICScat196 (0=2).
END IF.
DO IF ((HA_ICFsig196 = 1) & (HA_PC_Cambio2 = 2)).
RECODE HA_ICScat196 (0=1).
END IF.
EXECUTE.
VALUE LABELS HA_ICScat165 HA_ICScat196
          1 'Empeorado'
          2 'Empeora dentro de la funcionalidad'
          3 'No cambio'
          4 'Mejora dentro de la disfuncionalidad'
          5 'Recuperado'.
EXECUTE.

```

\* *Método DTW - DIFERENCIA TIPIFICADA DE WYRWICH.*

```

COMPUTE DTWnorm = (BDI2 - BDI1) / (BDI_nor_SD * SQRT(1-BDI_Fiabilidad)).
RECODE DTWnorm (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO DTW_ICFsig196norm.
VALUE LABELS DTW_ICFsig196norm 1 'Empeora' 0 'No Cambio' -1 'Mejora'.
COMPUTE DTWpre = (BDI2 - BDI1) / (DEpre * SQRT(1-BDI_Fiabilidad)).
RECODE DTWpre (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO DTW_ICFsig196pre.
VALUE LABELS DTW_ICFsig196pre 1 'Empeora'
          0 'No Cambio'
          -1 'Mejora'.
EXECUTE.

```

*\* Método TEI - TAMAÑO DEL EFECTO INDIVIDUAL.*

```
COMPUTE TEI_Den_prev = (BDI1 - Mpre) * (BDI1 - Mpre).
IF (Grupo = 1) TEI_Denominador = 6518.51 / 99.
IF (Grupo = 2) TEI_Denominador = 1735.39 / 99.
EXECUTE.
COMPUTE TEI = (BDI2 - BDI1) / SQRT(TEI_Denominador).
RECODE TEI (-0.20 thru 0.20=0) (0.20 thru Highest=1) (Lowest thru -0.20=-1) INTO TEI_ICFsig0.20.
VALUE LABELS TEI_ICFsig0.20 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
RECODE TEI (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO TEI_ICFsig1.96.
VALUE LABELS TEI_ICFsig1.96 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
EXECUTE.
```

*\* Método DIT - DIFERENCIA INDIVIDUAL TIPIFICADA.*

```
COMPUTE DIT_Den_prev = ((BDI2 - BDI1) - (Mpos - Mpre)) * ((BDI2 - BDI1) - (Mpos - Mpre)).
IF (Grupo = 1) DIT_Denominador = 2979.71 / 99.
IF (Grupo = 2) DIT_Denominador = 1120.00 / 99.
COMPUTE DIT = (BDI2 - BDI1) / SQRT(DIT_Denominador).
RECODE DIT (-0.20 thru 0.20=0) (0.20 thru Highest=1) (Lowest thru -0.20=-1) INTO DIT_ICFsig0.20.
VALUE LABELS DIT_ICFsig0.20 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
RECODE DIT (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO DIT_ICFsig1.96.
VALUE LABELS DIT_ICFsig1.96 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
EXECUTE.
```

*\* Método ESG - ESTADISTICO DE SENSIBILIDAD DE GUYATT.*

```
COMPUTE ESG = (BDI2 - BDI1) / SQRT(1120.00 / 99).
RECODE ESG (-0.20 thru 0.20=0) (0.20 thru Highest=1) (Lowest thru -0.20=-1) INTO ESG_ICFsig020.
VALUE LABELS ESG_ICFsig020 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
RECODE ESG (-1.96 thru 1.96=0) (1.961 thru Highest=1) (Lowest thru -1.961=-1) INTO ESG_ICFsig196.
VALUE LABELS ESG_ICFsig196 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
EXECUTE.
```

\* *Método ICCP / ICPI - INTERVALO PRONOSTICO DE CRAWFORD.*

```
COMPUTE Xpron = 1.202 + BDI1*0.768 + 0.078.
COMPUTE Craw = ( (BDI1 - Mpre) * (BDI1 - Mpre)) / 1735.39.
COMPUTE ICCP_II = Xpron - (1.96 * SQRT(10.478 *(0.01 + Craw))).
COMPUTE ICCP_IS = Xpron + (1.96 * SQRT(10.478 *(0.01 + Craw))).
COMPUTE ICCI_II = Xpron - (1.984 * SQRT(10.478 *(1.01 + Craw))).
COMPUTE ICCI_IS = Xpron + (1.984 * SQRT(10.478 *(1.01 + Craw))).
EXECUTE.
COMPUTE ICCP=0.
DO IF (BDI2 <= ICCP_II).
RECODE ICCP (0=-1).
END IF.
DO IF (BDI2 >= ICCP_IS).
RECODE ICCP (0=1).
END IF.
COMPUTE ICCI=0.
DO IF (BDI2 <= ICCI_II).
RECODE ICCI (0=-1).
END IF.
DO IF (BDI2 >= ICCI_IS).
RECODE ICCI (0=1).
END IF.
VALUE LABELS ICCP ICCI 1 'Empeora'
                0 'No Cambio'
                -1 'Mejora'.
EXECUTE.
```

\* *Método CPM\_50- CRITERIO DE PORCENTAJA DE MEJORÍA 50%*

```
COMPUTE PM_50Ran = BDI1-18.10.
COMPUTE PM_50dif = BDI1-BDI2.
COMPUTE PM_50 = PM_50dif / PM_50Ran.
COMPUTE PM_25= PM_50dif / BDI1.
COMPUTE PM_50sig=0.
DO IF ((PM_50 >= .50) & (PM_25 >= .25)).
RECODE PM_50sig (0=-1).
END IF.
DO IF ((PM_50 <0) & (PM_25 >= .25)).
RECODE PM_50sig (0=-1).
END IF.
VALUE LABELS PM_50sig 0 'No cambio'
                    1 'Empeora'
                    -1 'Mejora'.
EXECUTE.
```

## Anexo 2. Instrumento de Evaluación de Resultados OQ-45

**Instrucciones:** Nos gustaría conocer como se ha sentido durante la última semana, incluyendo el día de hoy. Lea cuidadosamente cada situación de las referidas más abajo y marque con una X el cuadradito que refleje mejor la ocurrencia de cada situación en el periodo mencionado (última semana).

**Nota.-** En este cuestionario se entiende por 'trabajo' al realizado por empleo, estudios, tareas del hogar, trabajo voluntario o no remunerado.

Nombre: \_\_\_\_\_

Fecha de Nacimiento: \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Sesión: \_\_\_\_\_ Día de hoy: \_\_\_\_ / \_\_\_\_ / 20\_\_

	Nunca	Rara-mente	Algunas veces	Frecuente-mente	Casi Siempre	S.	R.I.	R.S.
1. Me llevo bien con los demás .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0		<input type="checkbox"/>	
2. Me canso rápidamente .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
3. No siento interés por las cosas .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
4. Me siento estresado/a en el colegio/trabajo .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4			<input type="checkbox"/>
5. Me culpo a mí mismo/a por cosas .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
6. Me siento enfadado/a .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
7. Me siento infeliz en mi matrimonio/ con mis relaciones importantes .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input type="checkbox"/>	
8. Tengo pensamientos sobre acabar con mi vida .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input checked="" type="checkbox"/>		
9. Me siento débil .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
10. Me siento temeroso/a .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
11. Después de beber en exceso necesito beber a la mañana siguiente para volver a funcionar. (Si no bebe, marca "nunca") .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input checked="" type="checkbox"/>		
12. Encuentro mi trabajo/colegio satisfactorio .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0			<input checked="" type="checkbox"/>
13. Soy una persona feliz .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0	<input checked="" type="checkbox"/>		
14. Trabajo/estudio demasiado .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4			<input type="checkbox"/>
15. Siento que no valgo nada .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
16. Estoy preocupado/a por problemas familiares .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input type="checkbox"/>	
17. Tengo una vida sexual insatisfactoria .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input type="checkbox"/>	
18. Me siento solo/a .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input type="checkbox"/>	
19. Tengo discusiones frecuentemente .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input type="checkbox"/>	
20. Me siento querido/a .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0		<input checked="" type="checkbox"/>	
21. Disfruto de mi tiempo libre .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0			<input checked="" type="checkbox"/>
22. Tengo dificultades para concentrarme .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
23. Me siento desesperanzado/a ante el futuro .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
24. Me gusta cómo soy .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0	<input checked="" type="checkbox"/>		
25. Me asaltan a mi cabeza pensamientos molestos de los que no puedo deshacerme ..	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
26. Me siento molesto/a con quien critica mis hábitos de consumo de alcohol (o de drogas) (Si no consumes alcohol o drogas, marca "nunca") .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input checked="" type="checkbox"/>	
27. Me siento mal del estómago .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
28. No estoy trabajando/estudiando tan bien como suelo hacer .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		<input type="checkbox"/>
29. Mi corazón palpita demasiado .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
30. Tengo problemas para llevarme bien con amigos y conocidos .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4		<input type="checkbox"/>	
31. Estoy satisfecho/a con mi vida .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0	<input checked="" type="checkbox"/>		
32. Tengo problemas en el trabajo/colegio por mi consumo de alcohol o drogas (Si no consumes alcohol o drogas, marca "nunca") .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		<input checked="" type="checkbox"/>
33. Siento que algo malo va a ocurrir .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
34. Tengo agujetas / los músculos doloridos .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
35. Siento miedo a los espacios abiertos, a conducir, a viajar en autobús, metro u otros transportes públicos. ....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
36. Me siento nervioso/a .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
37. Siento que mis relaciones románticas son plenas y completas .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0		<input checked="" type="checkbox"/>	
38. Siento que en el trabajo/colegio no lo estoy haciendo bien .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		<input type="checkbox"/>
39. Tengo muchos desacuerdos en el trabajo/colegio .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		<input type="checkbox"/>
40. Siento que algo en mi cabeza no funciona bien .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
41. Tengo problemas para quedarme dormido/a o para estar despierto/a .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
42. Me siento triste .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		
43. Estoy satisfecho/a con mis relaciones interpersonales .....	<input type="checkbox"/> 4	<input type="checkbox"/> 3	<input type="checkbox"/> 2	<input type="checkbox"/> 1	<input type="checkbox"/> 0		<input checked="" type="checkbox"/>	
44. Me siento tan enfadado/a en el trabajo/colegio como para hacer algo de lo que pueda arrepentirme .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		<input checked="" type="checkbox"/>
45. Tengo dolores de cabeza .....	<input type="checkbox"/> 0	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/>		

+ +  
Total =

### Anexo 3. Tablas de Resultados de Cambio Individual

#### A3.E2 – Tablas adicionales del Estudio 2

Tabla A3E2.1. Resultados individuales “OQ 45 Total” con Índice de Cambio Fiable

Caso	Grupo Diagnóstico	Post –Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
1	Trast. Ánimo	-32	-3,31	15,32	Mejorado en disfuncionalidad
2	Trast. Ánimo	-32	-3,31	7,32	Mejorado en disfuncionalidad
3	Trast. Ánimo	10	1,04	-4,68	No cambio fiable en funcionalidad
4	Trast. Ánimo	8	0,83	-5,68	No cambio fiable en funcionalidad
5	Trast. Ánimo	4	0,41	-8,68	No cambio fiable en funcionalidad
6	Trast. Ánimo	6	0,62	-4,68	No cambio fiable en funcionalidad
7	Trast. Ánimo	0	0,00	10,32	No cambio fiable en disfuncionalidad
8	Trast. Ánimo	-6	-0,62	11,32	No cambio fiable en disfuncionalidad
9	Trast. Ánimo	-22	-2,28	5,32	Mejorado en disfuncionalidad
10	Trast. Ánimo	-38	-3,94	20,32	Mejorado en disfuncionalidad
11	Trast. Ánimo	-40	-4,14	19,32	Mejorado en disfuncionalidad
12	Trast. Ánimo	-42	-4,35	23,32	Mejorado en disfuncionalidad
13	Trast. Ánimo	-10	-1,04	7,32	No cambio fiable en disfuncionalidad
14	Trast. Ánimo	-32	-3,31	12,32	Mejorado en disfuncionalidad
15	Trast. Ánimo	-37	-3,83	12,32	Mejorado en disfuncionalidad
16	Trast. Ánimo	-39	-4,04	19,32	Mejorado en disfuncionalidad
17	Trast. Ánimo	-27	-2,80	6,32	Mejorado en disfuncionalidad
18	Trast. Ánimo	-12	-1,24	22,32	No cambio fiable en disfuncionalidad
19	Trast. Ánimo	-44	-4,56	3,32	Mejorado en disfuncionalidad
20	Trast. Ánimo	-36	-3,73	18,32	Mejorado en disfuncionalidad
21	Trast. Ánimo	-36	-3,73	30,32	Mejorado en disfuncionalidad
22	Trast. Ánimo	-9	-0,93	-0,68	No cambio fiable en funcionalidad
23	Trast. Ánimo	-15	-1,55	0,32	No cambio fiable en disfuncionalidad
24	Trast. Ánimo	-15	-1,55	13,32	No cambio fiable en disfuncionalidad
25	Trast. Ánimo	-29	-3,00	21,32	Mejorado en disfuncionalidad
26	Trast. Ánimo	-44	-4,56	15,32	Mejorado en disfuncionalidad
27	Trast. Ánimo	-23	-2,38	10,32	Mejorado en disfuncionalidad
28	Trast. Ánimo	-24	-2,49	13,32	Mejorado en disfuncionalidad
29	Trast. Ánimo	-39	-4,04	27,32	Mejorado en disfuncionalidad
30	Trast. Ánimo	-39	-4,04	12,32	Mejorado en disfuncionalidad
31	Trast. Ánimo	-15	-1,55	1,32	No cambio fiable en disfuncionalidad
32	Trast. Ánimo	-49	-5,07	11,32	Mejorado en disfuncionalidad
33	Trast. Ansiedad	19	1,97	-16,68	Empeorado en funcionalidad
34	Trast. Ansiedad	-10	-1,04	-3,68	No cambio fiable en funcionalidad
35	Trast. Ansiedad	-19	1,97	-9,68	Empeorado en funcionalidad
36	Trast. Ansiedad	-11	1,14	-4,68	No cambio fiable en funcionalidad
37	Trast. Ansiedad	-7	0,72	-5,68	No cambio fiable en funcionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Total = 58,68)

**Tabla A3E2.1. Resultados individuales “OQ 45 Total” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Post –Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> +Punto de Corte Funcional)
38	Trast. Ansiedad	1	0,10	-1,68	No cambio fiable en funcionalidad
39	Trast. Ansiedad	4	0,41	16,32	No cambio fiable en disfuncionalidad
40	Trast. Ansiedad	-16	-1,66	-1,68	No cambio fiable en funcionalidad
41	Trast. Ansiedad	-14	-1,45	14,32	No cambio fiable en disfuncionalidad
42	Trast. Ansiedad	-20	-2,07	11,32	Mejorado en disfuncionalidad
43	Trast. Ansiedad	-21	-2,17	11,32	Mejorado en disfuncionalidad
44	Trast. Ansiedad	-26	-2,69	11,32	Mejorado en disfuncionalidad
45	Trast. Ansiedad	13	1,35	-7,68	No cambio fiable en funcionalidad
46	Trast. Ansiedad	19	1,97	0,32	Empeorado en disfuncionalidad
47	Trast. Ansiedad	3	0,31	0,32	No cambio fiable en disfuncionalidad
48	Trast. Ansiedad	5	0,52	2,32	No cambio fiable en disfuncionalidad
49	Trast. Ansiedad	3	0,31	1,32	No cambio fiable en disfuncionalidad
50	Trast. Ansiedad	-6	-0,62	8,32	No cambio fiable en disfuncionalidad
51	Trast. Ansiedad	-20	-2,07	12,32	Mejorado en disfuncionalidad
52	Trast. Ansiedad	-23	-2,38	10,32	Mejorado en disfuncionalidad
53	Trast. Ansiedad	-19	-1,97	20,32	Mejorado en disfuncionalidad
54	Trast. Ansiedad	-18	-1,86	22,32	No cambio fiable en disfuncionalidad
55	Trast. Ansiedad	-37	-3,83	22,32	Mejorado en disfuncionalidad
56	Trast. Ansiedad	-31	-3,24	10,99	Mejorado en disfuncionalidad
57	Trast. Ansiedad	-11	-1,14	8,32	No cambio fiable en disfuncionalidad
58	Trast. Ansiedad	22	2,28	-14,68	Empeorado en funcionalidad
59	Trast. Ansiedad	4	0,41	8,32	No cambio fiable en disfuncionalidad
60	Trast. Ansiedad	-12	-1,24	9,32	No cambio fiable en disfuncionalidad
61	Trast. Ansiedad	-21	-2,17	5,32	Mejorado en disfuncionalidad
62	Trast. Ansiedad	-23	-2,38	9,32	Mejorado en disfuncionalidad
63	Trast. Ansiedad	-21	-2,17	15,32	Mejorado en disfuncionalidad
64	Trast. Ansiedad	-30	-3,11	10,32	Mejorado en disfuncionalidad
65	Trast. Ansiedad	-28	-2,90	16,32	Mejorado en disfuncionalidad
66	Trast. Ansiedad	-3	-0,31	-12,68	No cambio fiable en funcionalidad
67	Trast. Ansiedad	6	0,62	2,32	No cambio fiable en disfuncionalidad
68	Trast. Ansiedad	3	0,31	10,32	No cambio fiable en disfuncionalidad
69	Trast. Ansiedad	-11	-1,14	9,32	No cambio fiable en disfuncionalidad
70	Trast. Adaptativos	8	0,83	-13,68	No cambio fiable en funcionalidad
71	Trast. Adaptativos	12	1,24	-8,68	No cambio fiable en funcionalidad
72	Trast. Adaptativos	-2	-0,21	6,32	No cambio fiable en disfuncionalidad
73	Trast. Adaptativos	-9	-0,93	12,32	No cambio fiable en disfuncionalidad
74	Trast. Adaptativos	-17	-1,76	16,32	No cambio fiable en disfuncionalidad
75	Trast. Adaptativos	11	1,14	-19,68	No cambio fiable en funcionalidad
76	Trast. Adaptativos	21	2,17	-8,68	Empeorado en funcionalidad
77	Trast. Adaptativos	-3	-0,31	2,32	No cambio fiable en disfuncionalidad
78	Trast. Adaptativos	-1	-0,10	12,32	No cambio fiable en disfuncionalidad
79	Trast. Adaptativos	-21	-2,17	14,32	Mejorado en disfuncionalidad
80	Trast. Adaptativos	9	0,93	-2,68	No cambio fiable en funcionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Total = 58,68)

**Tabla A3E2.1. Resultados individuales “OQ 45 Total” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> + Punto de Corte Funcional)
81	Trast. Adaptativos	3	0,31	-7,68	No cambio fiable en funcionalidad
82	Trast. Adaptativos	27	2,80	-9,68	Empeorado en funcionalidad
83	Trast. Adaptativos	6	0,62	-15,68	No cambio fiable en funcionalidad
84	Trast. Adaptativos	12	1,24	-8,68	No cambio fiable en funcionalidad
85	Trast. Adaptativos	17	1,76	0,32	No cambio fiable en disfuncionalidad
86	Trast. Adaptativos	-11	-1,14	25,32	No cambio fiable en disfuncionalidad
87	Trast. Control Impulsos	-3	-0,31	-3,68	No cambio fiable en funcionalidad
88	Trast. Control Impulsos	6	0,66	10,66	No cambio fiable en disfuncionalidad
89	Trast. Control Impulsos	13	1,35	-8,68	No cambio fiable en funcionalidad
90	Trast. Control Impulsos	-12	-1,24	-0,68	No cambio fiable en funcionalidad
91	Trast. Control Impulsos	4	0,41	15,32	No cambio fiable en disfuncionalidad
92	Trast. Control Impulsos	-21	-2,17	14,32	Mejorado en disfuncionalidad
93	Trast. Control Impulsos	-20	-2,07	21,32	Mejorado en disfuncionalidad
94	Trast. Control Impulsos	-2	-0,21	1,32	No cambio fiable en disfuncionalidad
95	Trast. Control Impulsos	-38	-3,94	12,32	Mejorado en disfuncionalidad
96	Trast. Control Impulsos	-42	-4,35	22,32	Mejorado en disfuncionalidad
97	Trast. Control Impulsos	-42	-4,35	24,32	Mejorado en disfuncionalidad
98	Trast. Control Impulsos	1	0,10	-3,68	No cambio fiable en funcionalidad
99	Trast. Control Impulsos	-30	-3,11	11,32	Mejorado en disfuncionalidad
100	Trast. Control Impulsos	-47	-4,87	14,32	Mejorado en disfuncionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Total = 58,68)

**Tabla A3E2.1. Resultados individuales “OQ 45 Síntomas” con Índice de Cambio Fiable**

Caso	Grupo Diagnóstico	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> + Punto de Corte Funcional)
1	Trast. Ánimo	-26	-3,94	5,41	Mejorado en disfuncionalidad
2	Trast. Ánimo	-19	-2,88	4,41	Mejorado en disfuncionalidad
3	Trast. Ánimo	4	0,61	-6,59	No cambio fiable en funcionalidad
4	Trast. Ánimo	6	0,91	-4,59	No cambio fiable en funcionalidad
5	Trast. Ánimo	7	1,06	-7,59	No cambio fiable en funcionalidad
6	Trast. Ánimo	5	0,76	-2,59	No cambio fiable en funcionalidad
7	Trast. Ánimo	-1	-0,15	6,41	No cambio fiable en disfuncionalidad
8	Trast. Ánimo	-1	-0,15	9,41	No cambio fiable en disfuncionalidad
9	Trast. Ánimo	-21	-3,18	1,41	Mejorado en disfuncionalidad
10	Trast. Ánimo	-26	-3,94	9,41	Mejorado en disfuncionalidad
11	Trast. Ánimo	-33	-5,00	12,41	Mejorado en disfuncionalidad
12	Trast. Ánimo	-32	-4,84	14,41	Mejorado en disfuncionalidad
13	Trast. Ánimo	-6	-0,91	4,41	No cambio fiable en disfuncionalidad
14	Trast. Ánimo	-17	-2,57	4,41	Mejorado en disfuncionalidad
15	Trast. Ánimo	-30	-4,54	7,41	Mejorado en disfuncionalidad
16	Trast. Ánimo	-34	-5,15	5,41	Mejorado en disfuncionalidad
17	Trast. Ánimo	-21	-3,18	5,41	Mejorado en disfuncionalidad
18	Trast. Ánimo	-5	-0,76	9,41	No cambio fiable en disfuncionalidad
19	Trast. Ánimo	-31	-4,69	0,41	Mejorado en disfuncionalidad
20	Trast. Ánimo	-27	-4,09	6,41	Mejorado en disfuncionalidad
21	Trast. Ánimo	-28	-4,24	15,41	Mejorado en disfuncionalidad
22	Trast. Ánimo	-2	-0,30	-1,59	No cambio fiable en funcionalidad
23	Trast. Ánimo	-16	-2,42	0,41	Mejorado en disfuncionalidad
24	Trast. Ánimo	-1	-0,15	8,41	No cambio fiable en disfuncionalidad
25	Trast. Ánimo	-24	-3,63	6,41	Mejorado en disfuncionalidad
26	Trast. Ánimo	-34	-5,15	10,41	Mejorado en disfuncionalidad
27	Trast. Ánimo	-19	-2,88	9,41	Mejorado en disfuncionalidad
28	Trast. Ánimo	-15	-2,27	5,41	Mejorado en disfuncionalidad
29	Trast. Ánimo	-22	-3,33	16,41	Mejorado en disfuncionalidad
30	Trast. Ánimo	-36	-5,45	9,41	Mejorado en disfuncionalidad
31	Trast. Ánimo	-12	-1,82	4,41	No cambio fiable en disfuncionalidad
32	Trast. Ánimo	-30	-4,54	4,41	Mejorado en disfuncionalidad
33	Trast. Ansiedad	8	1,21	-14,59	No cambio fiable en funcionalidad
34	Trast. Ansiedad	2	-0,30	-3,59	No cambio fiable en funcionalidad
35	Trast. Ansiedad	16	2,42	-3,59	Empeorado en funcionalidad
36	Trast. Ansiedad	9	1,36	-7,59	No cambio fiable en funcionalidad
37	Trast. Ansiedad	7	1,06	-5,59	No cambio fiable en funcionalidad

PCF – Punto de Corte Funcional (OQ-45 Síntomas = 34,59)

**Tabla A3E2.1. Resultados individuales “OQ 45 Síntomas” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Post –Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> +Punto de Corte Funcional)
38	Trast. Ansiedad	7	1,06	-1,59	No cambio fiable en funcionalidad
39	Trast. Ansiedad	7	1,06	5,41	No cambio fiable en disfuncionalidad
40	Trast. Ansiedad	-2	-0,30	-5,59	No cambio fiable en funcionalidad
41	Trast. Ansiedad	-3	-0,45	6,41	No cambio fiable en disfuncionalidad
42	Trast. Ansiedad	-9	-1,36	3,41	No cambio fiable en disfuncionalidad
43	Trast. Ansiedad	-4	-0,61	-0,59	No cambio fiable en funcionalidad
44	Trast. Ansiedad	-12	-1,82	7,41	No cambio fiable en disfuncionalidad
45	Trast. Ansiedad	12	1,82	-7,59	No cambio fiable en funcionalidad
46	Trast. Ansiedad	11	1,67	-1,59	No cambio fiable en funcionalidad
47	Trast. Ansiedad	6	0,91	-4,59	No cambio fiable en funcionalidad
48	Trast. Ansiedad	1	0,15	2,41	No cambio fiable en disfuncionalidad
49	Trast. Ansiedad	11	1,67	-3,59	No cambio fiable en funcionalidad
50	Trast. Ansiedad	0	0,00	4,41	No cambio fiable en disfuncionalidad
51	Trast. Ansiedad	-11	-1,67	5,41	No cambio fiable en disfuncionalidad
52	Trast. Ansiedad	-4	-0,61	3,41	No cambio fiable en disfuncionalidad
53	Trast. Ansiedad	-20	-3,03	8,41	Mejorado en disfuncionalidad
54	Trast. Ansiedad	-12	-1,82	10,41	No cambio fiable en disfuncionalidad
55	Trast. Ansiedad	-29	-4,39	16,41	Mejorado en disfuncionalidad
56	Trast. Ansiedad	-12	-1,87	1,08	No cambio fiable en disfuncionalidad
57	Trast. Ansiedad	-12	-1,82	1,41	No cambio fiable en disfuncionalidad
58	Trast. Ansiedad	12	1,82	-10,59	No cambio fiable en funcionalidad
59	Trast. Ansiedad	-2	-0,30	-2,59	No cambio fiable en funcionalidad
60	Trast. Ansiedad	-2	-0,30	1,41	No cambio fiable en disfuncionalidad
61	Trast. Ansiedad	-11	-1,67	1,41	No cambio fiable en disfuncionalidad
62	Trast. Ansiedad	-12	-1,82	5,41	No cambio fiable en disfuncionalidad
63	Trast. Ansiedad	-14	-2,12	8,41	Mejorado en disfuncionalidad
64	Trast. Ansiedad	-21	-3,18	6,41	Mejorado en disfuncionalidad
65	Trast. Ansiedad	-17	-2,57	8,41	Mejorado en disfuncionalidad
66	Trast. Ansiedad	-1	-0,15	-11,59	No cambio fiable en funcionalidad
67	Trast. Ansiedad	6	0,91	3,41	No cambio fiable en disfuncionalidad
68	Trast. Ansiedad	3	0,45	3,41	No cambio fiable en disfuncionalidad
69	Trast. Ansiedad	-14	-2,12	-0,59	Recuperado
70	Trast. Adaptativos	12	1,82	-10,59	No cambio fiable en funcionalidad
71	Trast. Adaptativos	8	1,21	-13,59	No cambio fiable en funcionalidad
72	Trast. Adaptativos	3	0,45	-1,59	No cambio fiable en funcionalidad
73	Trast. Adaptativos	-5	-0,76	6,41	No cambio fiable en disfuncionalidad
74	Trast. Adaptativos	-13	-1,97	5,41	Mejorado en disfuncionalidad
75	Trast. Adaptativos	5	0,76	-12,59	No cambio fiable en funcionalidad
76	Trast. Adaptativos	12	1,82	-7,59	No cambio fiable en funcionalidad
77	Trast. Adaptativos	-5	-0,76	1,41	No cambio fiable en disfuncionalidad
78	Trast. Adaptativos	-4	-0,61	1,41	No cambio fiable en disfuncionalidad
79	Trast. Adaptativos	-10	-1,51	3,41	No cambio fiable en disfuncionalidad
80	Trast. Adaptativos	-3	-0,45	-4,59	No cambio fiable en funcionalidad

PCF – Punto de Corte Funcional (OQ-45 Síntomas = 34,59)

**Tabla A3E2.1.- Resultados individuales “OQ 45 Síntomas” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> + Punto de Corte Funcional)
81	Trast. Adaptativos	-1	-0,15	-10,59	No cambio fiable en funcionalidad
82	Trast. Adaptativos	10	1,51	-13,59	No cambio fiable en funcionalidad
83	Trast. Adaptativos	11	1,67	-11,59	No cambio fiable en funcionalidad
84	Trast. Adaptativos	9	1,36	-9,59	No cambio fiable en funcionalidad
85	Trast. Adaptativos	7	1,06	-5,59	No cambio fiable en funcionalidad
86	Trast. Adaptativos	-3	-0,45	17,41	No cambio fiable en disfuncionalidad
87	Trast. Control Impulsos	0	0,00	-7,59	No cambio fiable en funcionalidad
88	Trast. Control Impulsos	9	1,36	2,41	No cambio fiable en disfuncionalidad
89	Trast. Control Impulsos	11	1,67	-7,59	No cambio fiable en funcionalidad
90	Trast. Control Impulsos	-2	-0,30	-2,59	No cambio fiable en funcionalidad
91	Trast. Control Impulsos	7	1,06	6,41	No cambio fiable en disfuncionalidad
92	Trast. Control Impulsos	-12	-1,82	8,41	No cambio fiable en disfuncionalidad
93	Trast. Control Impulsos	-19	-2,88	12,41	Mejorado en disfuncionalidad
94	Trast. Control Impulsos	7	1,06	-0,59	No cambio fiable en funcionalidad
95	Trast. Control Impulsos	-23	-3,48	5,41	Mejorado en disfuncionalidad
96	Trast. Control Impulsos	-32	-4,84	8,41	Mejorado en disfuncionalidad
97	Trast. Control Impulsos	-22	-3,33	9,41	Mejorado en disfuncionalidad
98	Trast. Control Impulsos	2	0,30	-5,59	No cambio fiable en funcionalidad
99	Trast. Control Impulsos	-12	-1,82	4,41	No cambio fiable en disfuncionalidad
100	Trast. Control Impulsos	-31	-4,69	7,41	Mejorado en disfuncionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Síntomas = 34,59)

**Tabla A3E3.1.- Resultados individuales “OQ 45 Interpersonal” con Índice de Cambio Fiable**

Caso	Grupo Diagnóstico	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
1	Trast. Ánimo	-2	-0,45	4,93	No cambio fiable en disfuncionalidad
2	Trast. Ánimo	1	0,22	2,93	No cambio fiable en disfuncionalidad
3	Trast. Ánimo	6	1,34	3,93	No cambio fiable en disfuncionalidad
4	Trast. Ánimo	2	0,45	0,93	No cambio fiable en disfuncionalidad
5	Trast. Ánimo	1	0,22	0,93	No cambio fiable en disfuncionalidad
6	Trast. Ánimo	3	0,67	-0,07	No cambio fiable en funcionalidad
7	Trast. Ánimo	1	0,22	-0,07	No cambio fiable en funcionalidad
8	Trast. Ánimo	0	0,00	2,93	No cambio fiable en disfuncionalidad
9	Trast. Ánimo	2	0,45	0,93	No cambio fiable en disfuncionalidad
10	Trast. Ánimo	-6	-1,34	6,93	No cambio fiable en disfuncionalidad
11	Trast. Ánimo	-6	-1,34	5,93	No cambio fiable en disfuncionalidad
12	Trast. Ánimo	-1	-0,22	5,93	No cambio fiable en disfuncionalidad
13	Trast. Ánimo	-1	-0,22	1,93	No cambio fiable en disfuncionalidad
14	Trast. Ánimo	-13	-2,89	3,93	Mejorado en disfuncionalidad
15	Trast. Ánimo	2	0,45	5,93	No cambio fiable en disfuncionalidad
16	Trast. Ánimo	-1	-0,22	5,93	No cambio fiable en disfuncionalidad
17	Trast. Ánimo	-3	-0,67	0,93	No cambio fiable en disfuncionalidad
18	Trast. Ánimo	-3	-0,67	8,93	No cambio fiable en disfuncionalidad
19	Trast. Ánimo	-6	-1,34	0,93	No cambio fiable en disfuncionalidad
20	Trast. Ánimo	0	0,00	5,93	No cambio fiable en disfuncionalidad
21	Trast. Ánimo	0	0,00	9,93	No cambio fiable en disfuncionalidad
22	Trast. Ánimo	-3	-0,67	1,93	No cambio fiable en disfuncionalidad
23	Trast. Ánimo	5	1,11	3,93	No cambio fiable en disfuncionalidad
24	Trast. Ánimo	-6	-1,34	5,93	No cambio fiable en disfuncionalidad
25	Trast. Ánimo	0	0,00	10,93	No cambio fiable en disfuncionalidad
26	Trast. Ánimo	-8	-1,78	4,93	No cambio fiable en disfuncionalidad
27	Trast. Ánimo	1	0,22	-0,07	No cambio fiable en funcionalidad
28	Trast. Ánimo	2	0,45	4,93	No cambio fiable en disfuncionalidad
29	Trast. Ánimo	-6	-1,34	4,93	No cambio fiable en disfuncionalidad
30	Trast. Ánimo	4	0,89	3,93	No cambio fiable en disfuncionalidad
31	Trast. Ánimo	3	0,67	-1,07	No cambio fiable en funcionalidad
32	Trast. Ánimo	13	-2,89	0,93	Mejorado en disfuncionalidad
33	Trast. Ansiedad	7	1,56	-1,07	No cambio fiable en funcionalidad
34	Trast. Ansiedad	-5	-1,11	1,93	No cambio fiable en disfuncionalidad
35	Trast. Ansiedad	2	0,45	-5,07	No cambio fiable en funcionalidad
36	Trast. Ansiedad	-2	-0,45	2,93	No cambio fiable en disfuncionalidad
37	Trast. Ansiedad	-1	-0,22	-1,07	No cambio fiable en funcionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Interpersonal = 13.07)

**Tabla A3E3.1.- Resultados individuales “OQ 45 Interpersonal” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Post – Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> + Punto de Corte Funcional)
38	Trast. Ansiedad	0	0,00	2,93	No cambio fiable en disfuncionalidad
39	Trast. Ansiedad	-3	-0,67	3,93	No cambio fiable en disfuncionalidad
40	Trast. Ansiedad	-9	-2,00	5,93	Mejorado en disfuncionalidad
41	Trast. Ansiedad	-8	-1,78	7,93	No cambio fiable en disfuncionalidad
42	Trast. Ansiedad	-8	-1,78	2,93	No cambio fiable en disfuncionalidad
43	Trast. Ansiedad	-10	-2,23	5,93	Mejorado en disfuncionalidad
44	Trast. Ansiedad	-6	-1,34	3,93	No cambio fiable en disfuncionalidad
45	Trast. Ansiedad	1	0,22	-0,07	No cambio fiable en funcionalidad
46	Trast. Ansiedad	4	0,89	0,93	No cambio fiable en disfuncionalidad
47	Trast. Ansiedad	-6	-1,34	0,93	No cambio fiable en disfuncionalidad
48	Trast. Ansiedad	4	0,89	-0,07	No cambio fiable en funcionalidad
49	Trast. Ansiedad	-3	-0,67	5,93	No cambio fiable en disfuncionalidad
50	Trast. Ansiedad	-2	-0,45	1,93	No cambio fiable en disfuncionalidad
51	Trast. Ansiedad	-6	-1,34	4,93	No cambio fiable en disfuncionalidad
52	Trast. Ansiedad	-13	-2,89	4,93	Mejorado en disfuncionalidad
53	Trast. Ansiedad	4	0,89	5,93	No cambio fiable en disfuncionalidad
54	Trast. Ansiedad	-4	-0,89	6,93	No cambio fiable en disfuncionalidad
55	Trast. Ansiedad	-2	-0,45	5,93	No cambio fiable en disfuncionalidad
56	Trast. Ansiedad	-9	-2,00	10,93	Mejorado en disfuncionalidad
57	Trast. Ansiedad	1	0,22	3,93	No cambio fiable en disfuncionalidad
58	Trast. Ansiedad	8	1,78	-1,07	No cambio fiable en funcionalidad
59	Trast. Ansiedad	2	0,45	3,93	No cambio fiable en disfuncionalidad
60	Trast. Ansiedad	-6	-1,34	2,93	No cambio fiable en disfuncionalidad
61	Trast. Ansiedad	-4	-0,89	1,93	No cambio fiable en disfuncionalidad
62	Trast. Ansiedad	-4	-0,89	2,93	No cambio fiable en disfuncionalidad
63	Trast. Ansiedad	0	0,00	-0,07	No cambio fiable en funcionalidad
64	Trast. Ansiedad	-5	-1,11	2,93	No cambio fiable en disfuncionalidad
65	Trast. Ansiedad	-10	-2,23	5,93	Mejorado en disfuncionalidad
66	Trast. Ansiedad	3	0,67	1,93	No cambio fiable en disfuncionalidad
67	Trast. Ansiedad	3	0,67	2,93	No cambio fiable en disfuncionalidad
68	Trast. Ansiedad	4	0,89	0,93	No cambio fiable en disfuncionalidad
69	Trast. Ansiedad	1	0,22	6,93	No cambio fiable en disfuncionalidad
70	Trast. Adaptativos	-2	-0,45	-3,07	No cambio fiable en funcionalidad
71	Trast. Adaptativos	4	0,89	2,93	No cambio fiable en disfuncionalidad
72	Trast. Adaptativos	-2	-0,45	5,93	No cambio fiable en disfuncionalidad
73	Trast. Adaptativos	1	0,22	-1,07	No cambio fiable en funcionalidad
74	Trast. Adaptativos	0	0,00	8,93	No cambio fiable en disfuncionalidad
75	Trast. Adaptativos	5	1,11	-4,07	No cambio fiable en funcionalidad
76	Trast. Adaptativos	6	1,34	-1,07	No cambio fiable en funcionalidad
77	Trast. Adaptativos	6	1,34	-0,07	No cambio fiable en funcionalidad
78	Trast. Adaptativos	3	0,67	5,93	No cambio fiable en disfuncionalidad
79	Trast. Adaptativos	-8	-1,78	4,93	No cambio fiable en disfuncionalidad
80	Trast. Adaptativos	7	1,56	1,93	No cambio fiable en disfuncionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Interpersonal = 13.07)

**Tabla A3E3.1.- Resultados individuales “OQ 45 Interpersonal” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> + Punto de Corte Funcional)
81	Trast. Adaptativos	4	0,89	1,93	No cambio fiable en disfuncionalidad
82	Trast. Adaptativos	11	2,45	1,93	Empeorado en disfuncionalidad
83	Trast. Adaptativos	-2	-0,45	-3,07	No cambio fiable en funcionalidad
84	Trast. Adaptativos	4	0,89	-0,07	No cambio fiable en funcionalidad
85	Trast. Adaptativos	7	1,56	2,93	No cambio fiable en disfuncionalidad
86	Trast. Adaptativos	-1	-0,22	5,93	No cambio fiable en disfuncionalidad
87	Trast. Control Impulsos	-2	-0,45	-0,07	No cambio fiable en funcionalidad
88	Trast. Control Impulsos	-1	-0,22	2,93	No cambio fiable en disfuncionalidad
89	Trast. Control Impulsos	3	0,67	-0,07	No cambio fiable en funcionalidad
90	Trast. Control Impulsos	-1	-0,22	3,93	No cambio fiable en disfuncionalidad
91	Trast. Control Impulsos	-1	-0,22	3,93	No cambio fiable en disfuncionalidad
92	Trast. Control Impulsos	1	0,22	1,93	No cambio fiable en disfuncionalidad
93	Trast. Control Impulsos	3	0,67	5,93	No cambio fiable en disfuncionalidad
94	Trast. Control Impulsos	-5	-1,11	2,93	No cambio fiable en disfuncionalidad
95	Trast. Control Impulsos	-6	-1,34	6,93	No cambio fiable en disfuncionalidad
96	Trast. Control Impulsos	-3	-0,67	8,93	No cambio fiable en disfuncionalidad
97	Trast. Control Impulsos	-10	-2,23	9,93	Mejorado en disfuncionalidad
98	Trast. Control Impulsos	1	0,22	1,93	No cambio fiable en disfuncionalidad
99	Trast. Control Impulsos	-8	-1,78	6,93	No cambio fiable en disfuncionalidad
100	Trast. Control Impulsos	-3	-0,67	5,93	No cambio fiable en disfuncionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Interpersonal = 13.07)

**Tabla A3E2.1.- Resultados individuales “OQ 45 Rol Social” con Índice de Cambio Fiable**

Caso	Grupo Diagnóstico	Post –Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> +Punto de Corte Funcional)
1	Trast. Ánimo	-4	-0,84	5,15	No cambio fiable en disfuncionalidad
2	Trast. Ánimo	14	-2,94	0,15	Mejorado en disfuncionalidad
3	Trast. Ánimo	0	0,00	-1,85	No cambio fiable en funcionalidad
4	Trast. Ánimo	0	0,00	-1,85	No cambio fiable en funcionalidad
5	Trast. Ánimo	-4	-0,84	-1,85	No cambio fiable en funcionalidad
6	Trast. Ánimo	-2	-0,42	-1,85	No cambio fiable en funcionalidad
7	Trast. Ánimo	0	0,00	4,15	No cambio fiable en disfuncionalidad
8	Trast. Ánimo	-5	-1,05	-0,85	No cambio fiable en funcionalidad
9	Trast. Ánimo	-3	-0,63	3,15	No cambio fiable en disfuncionalidad
10	Trast. Ánimo	-6	-1,26	4,15	No cambio fiable en disfuncionalidad
11	Trast. Ánimo	-1	-0,21	1,15	No cambio fiable en disfuncionalidad
12	Trast. Ánimo	-9	-1,89	3,15	No cambio fiable en disfuncionalidad
13	Trast. Ánimo	-3	-0,63	1,15	No cambio fiable en disfuncionalidad
14	Trast. Ánimo	-2	-0,42	4,15	No cambio fiable en disfuncionalidad
15	Trast. Ánimo	-9	-1,89	-0,85	No cambio fiable en funcionalidad
16	Trast. Ánimo	-4	-0,84	8,15	No cambio fiable en disfuncionalidad
17	Trast. Ánimo	-3	-0,63	0,15	No cambio fiable en disfuncionalidad
18	Trast. Ánimo	-4	-0,84	4,15	No cambio fiable en disfuncionalidad
19	Trast. Ánimo	-7	-1,47	2,15	No cambio fiable en disfuncionalidad
20	Trast. Ánimo	-9	-1,89	6,15	No cambio fiable en disfuncionalidad
21	Trast. Ánimo	-8	-1,68	5,15	No cambio fiable en disfuncionalidad
22	Trast. Ánimo	-4	-0,84	-0,85	No cambio fiable en funcionalidad
23	Trast. Ánimo	-4	-0,84	-3,85	No cambio fiable en funcionalidad
24	Trast. Ánimo	-8	-1,68	-0,85	No cambio fiable en funcionalidad
25	Trast. Ánimo	-5	-1,05	4,15	No cambio fiable en disfuncionalidad
26	Trast. Ánimo	-2	-0,42	0,15	No cambio fiable en disfuncionalidad
27	Trast. Ánimo	-5	-1,05	1,15	No cambio fiable en disfuncionalidad
28	Trast. Ánimo	-11	-2,31	3,15	Mejorado en disfuncionalidad
29	Trast. Ánimo	-11	-2,31	6,15	Mejorado en disfuncionalidad
30	Trast. Ánimo	-7	-1,47	-0,85	No cambio fiable en funcionalidad
31	Trast. Ánimo	-6	-1,26	-1,85	No cambio fiable en funcionalidad
32	Trast. Ánimo	-6	-1,26	6,15	No cambio fiable en disfuncionalidad
33	Trast. Ansiedad	4	0,84	-0,85	No cambio fiable en funcionalidad
34	Trast. Ansiedad	-3	-0,63	-1,85	No cambio fiable en funcionalidad
35	Trast. Ansiedad	1	0,21	-0,85	No cambio fiable en funcionalidad
36	Trast. Ansiedad	4	0,84	0,15	No cambio fiable en disfuncionalidad
37	Trast. Ansiedad	1	0,21	1,15	No cambio fiable en disfuncionalidad

PCF – Punto de Corte Funcional (OQ-45 Rol Social = 10.85)

**Tabla A3E2.1.- Resultados individuales “OQ 45 Rol Social” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final (ICF <sub>JT</sub> + Punto de Corte Funcional)
38	Trast. Ansiedad	-6	-1,26	-2,85	No cambio fiable en funcionalidad
39	Trast. Ansiedad	0	0,00	7,15	No cambio fiable en disfuncionalidad
40	Trast. Ansiedad	-5	-1,05	-1,85	No cambio fiable en funcionalidad
41	Trast. Ansiedad	-3	-0,63	0,15	No cambio fiable en disfuncionalidad
42	Trast. Ansiedad	-3	-0,63	5,15	No cambio fiable en disfuncionalidad
43	Trast. Ansiedad	-7	-1,47	6,15	No cambio fiable en disfuncionalidad
44	Trast. Ansiedad	-8	-1,68	0,15	No cambio fiable en disfuncionalidad
45	Trast. Ansiedad	0	0,00	0,15	No cambio fiable en disfuncionalidad
46	Trast. Ansiedad	4	0,84	1,15	No cambio fiable en disfuncionalidad
47	Trast. Ansiedad	3	0,63	4,15	No cambio fiable en disfuncionalidad
48	Trast. Ansiedad	0	0,00	0,15	No cambio fiable en disfuncionalidad
49	Trast. Ansiedad	-5	-1,05	-0,85	No cambio fiable en funcionalidad
50	Trast. Ansiedad	-4	-0,84	2,15	No cambio fiable en disfuncionalidad
51	Trast. Ansiedad	-3	-0,63	2,15	No cambio fiable en disfuncionalidad
52	Trast. Ansiedad	-6	-1,26	2,15	No cambio fiable en disfuncionalidad
53	Trast. Ansiedad	-3	-0,63	6,15	No cambio fiable en disfuncionalidad
54	Trast. Ansiedad	-2	-0,42	5,15	No cambio fiable en disfuncionalidad
55	Trast. Ansiedad	-6	-1,26	0,15	No cambio fiable en disfuncionalidad
56	Trast. Ansiedad	-10	-2,10	-0,85	Recuperado
57	Trast. Ansiedad	0	0,00	3,15	No cambio fiable en disfuncionalidad
58	Trast. Ansiedad	2	0,42	-2,85	No cambio fiable en funcionalidad
59	Trast. Ansiedad	4	0,84	7,15	No cambio fiable en disfuncionalidad
60	Trast. Ansiedad	-4	-0,84	5,15	No cambio fiable en disfuncionalidad
61	Trast. Ansiedad	-6	-1,26	2,15	No cambio fiable en disfuncionalidad
62	Trast. Ansiedad	-7	-1,47	1,15	No cambio fiable en disfuncionalidad
63	Trast. Ansiedad	-7	-1,47	7,15	No cambio fiable en disfuncionalidad
64	Trast. Ansiedad	-4	-0,84	1,15	No cambio fiable en disfuncionalidad
65	Trast. Ansiedad	-1	-0,21	2,15	No cambio fiable en disfuncionalidad
66	Trast. Ansiedad	-5	-1,05	-2,85	No cambio fiable en funcionalidad
67	Trast. Ansiedad	-3	-0,63	-3,85	No cambio fiable en funcionalidad
68	Trast. Ansiedad	-4	-0,84	6,15	No cambio fiable en disfuncionalidad
69	Trast. Ansiedad	2	0,42	3,15	No cambio fiable en disfuncionalidad
70	Trast. Adaptativos	-2	-0,42	0,15	No cambio fiable en disfuncionalidad
71	Trast. Adaptativos	0	0,00	2,15	No cambio fiable en disfuncionalidad
72	Trast. Adaptativos	-3	-0,63	2,15	No cambio fiable en disfuncionalidad
73	Trast. Adaptativos	-5	-1,05	7,15	No cambio fiable en disfuncionalidad
74	Trast. Adaptativos	-4	-0,84	2,15	No cambio fiable en disfuncionalidad
75	Trast. Adaptativos	1	0,21	-2,85	No cambio fiable en funcionalidad
76	Trast. Adaptativos	3	0,63	0,15	No cambio fiable en disfuncionalidad
77	Trast. Adaptativos	-4	-0,84	1,15	No cambio fiable en disfuncionalidad
78	Trast. Adaptativos	0	0,00	5,15	No cambio fiable en disfuncionalidad
79	Trast. Adaptativos	-3	-0,63	6,15	No cambio fiable en disfuncionalidad
80	Trast. Adaptativos	5	1,05	0,15	No cambio fiable en disfuncionalidad

PCF – Punto de Corte Funcional (OQ-45 Rol Social = 10.85)

**Tabla A3E2.1. Resultados individuales “OQ 45 Rol Social” con Índice de Cambio Fiable (Continuación)**

Caso	Grupo Diagnóstico	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
81	Trast. Adaptativos	0	0,00	1,15	No cambio fiable en disfuncionalidad
82	Trast. Adaptativos	6	1,26	2,15	No cambio fiable en disfuncionalidad
83	Trast. Adaptativos	-3	-0,63	-0,85	No cambio fiable en funcionalidad
84	Trast. Adaptativos	-1	-0,21	1,15	No cambio fiable en disfuncionalidad
85	Trast. Adaptativos	3	0,63	3,15	No cambio fiable en disfuncionalidad
86	Trast. Adaptativos	-7	-1,47	2,15	No cambio fiable en disfuncionalidad
87	Trast. Control Impulsos	-1	-0,21	4,15	No cambio fiable en disfuncionalidad
88	Trast. Control Impulsos	-1	-0,35	5,49	No cambio fiable en disfuncionalidad
89	Trast. Control Impulsos	-1	-0,21	-0,85	No cambio fiable en funcionalidad
90	Trast. Control Impulsos	-9	-1,89	-1,85	No cambio fiable en funcionalidad
91	Trast. Control Impulsos	-2	-0,42	5,15	No cambio fiable en disfuncionalidad
92	Trast. Control Impulsos	-10	-2,10	4,15	Mejorado en disfuncionalidad
93	Trast. Control Impulsos	-4	-0,84	3,15	No cambio fiable en disfuncionalidad
94	Trast. Control Impulsos	-4	-0,84	-0,85	No cambio fiable en funcionalidad
95	Trast. Control Impulsos	-9	-1,89	0,15	No cambio fiable en disfuncionalidad
96	Trast. Control Impulsos	-7	-1,47	5,15	No cambio fiable en disfuncionalidad
97	Trast. Control Impulsos	-10	-2,10	5,15	Mejorado en disfuncionalidad
98	Trast. Control Impulsos	-2	-0,42	0,15	No cambio fiable en funcionalidad
99	Trast. Control Impulsos	-10	-2,10	0,15	Mejorado en disfuncionalidad
100	Trast. Control Impulsos	-13	-2,73	1,15	Mejorado en disfuncionalidad

\* PCF – Punto de Corte Funcional (OQ-45 Rol Social = 10.85)

### A3E3. Tablas adicionales del Estudio 3

Tabla A3E3.1. Resultados individuales “BDI” con Índice de Cambio Fiable Grupo Clínico

Caso	Grupo Diagnóstico	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
1	T. Dep. Mayor Episodio Único	-1	-0,26	6,9	Sin cambio dentro de la disfuncionalidad
2	T. Dep. No Especificado	-9	-2,30	-10,1	Recuperado pre
3	T. Dep. Mayor Episodio Único	-14	-3,58	-3,1	Recuperado
4	T. Adapta. Ánimo Mixto	-15	-3,84	-1,1	Recuperado
5	T. Distímico	-15	-3,84	-0,1	Recuperado
6	T. Adapta. Ánimo Deprimido	-14	-3,58	-3,1	Recuperado
7	T. Adapta. Ánimo Mixto	-11	-2,82	-8,1	Recuperado
8	T. Adapta. Animo Deprimido	-3	-0,77	-3,1	Sin cambio dentro de la funcionalidad pre
9	T. Adapta. Animo Mixto	-12	-3,07	-8,1	Recuperado
10	T. Adapta. Animo Mixto	-16	-4,10	-9,1	Recuperado
11	T. Dep. Mayor Episodio Único	-19	-4,86	6,9	Mejoría dentro de la disfuncionalidad
12	T. Adapta. Animo Mixto	-14	-3,58	0,9	Mejoría dentro de la disfuncionalidad
13	T. Dep. Mayor Episodio Único	-15	-3,84	5,9	Mejoría dentro de la disfuncionalidad
14	T. Adapta. Animo Mixto	-10	-2,56	-12,1	Recuperado pre
15	T. Dep. No Especificado	-3	-0,77	0,9	Sin cambio dentro de la disfuncionalidad
16	T. Dep. Mayor Episodio Único	-22	-5,63	2,9	Mejoría dentro de la disfuncionalidad
17	T. Adapta. Animo Mixto	-11	-2,82	-8,1	Recuperado
18	T. Distímico	-19	-4,86	4,9	Mejoría dentro de la disfuncionalidad
19	T. Adapta. Animo Mixto	-4	-1,02	-0,1	Sin cambio dentro de la funcionalidad
20	T. Adapta. Animo Deprimido	-14	-3,58	-4,1	Recuperado
21	T. Dep. Mayor Episodio Único	-9	-2,30	11,9	Mejoría dentro de la disfuncionalidad
22	T. Dep. Mayor Episodio Único	-16	-4,10	6,9	Mejoría dentro de la disfuncionalidad
23	T. Dep. Mayor Episodio Único	-10	-2,56	3,9	Mejoría dentro de la disfuncionalidad
24	T. Distímico	-20	-5,12	2,9	Mejoría dentro de la disfuncionalidad
25	T. Adapta. Animo Deprimido	-14	-3,58	0,9	Mejoría dentro de la disfuncionalidad
26	T. Dep. Mayor Recidivante	-8	-2,05	14,9	Mejoría dentro de la disfuncionalidad
27	T. Dep. Mayor Episodio Único	-12	-3,07	8,9	Mejoría dentro de la disfuncionalidad
28	T. Adapta. Animo Mixto	-15	-3,84	-1,1	Recuperado
29	T. Dep. Mayor Episodio Único	-2	-0,51	12,9	Sin cambio dentro de la disfuncionalidad
30	T. Dep. Mayor Recidivante	-18	-4,61	6,9	Mejoría dentro de la disfuncionalidad
31	T. Adapta. Animo Deprimido	-19	-4,86	2,9	Mejoría dentro de la disfuncionalidad
32	T. Dep. No Especificado	0	0,00	2,9	Sin cambio dentro de la disfuncionalidad
33	T. Dep. Mayor Episodio Único	-6	-1,54	11,9	Sin cambio dentro de la disfuncionalidad
34	T. Dep. Mayor Episodio Único	-17	-4,35	6,9	Mejoría dentro de la disfuncionalidad
35	T. Adapta. Animo Deprimido	-23	-5,89	-8,1	Recuperado
36	T. Dep. Mayor Episodio Único	-17	-4,35	8,9	Mejoría dentro de la disfuncionalidad
37	T. Adapta. Animo Deprimido	-10	-2,56	-8,1	Recuperado

\* PCF – Punto de Corte Funcional (BDI= 18,10)

**Tabla A3E3.1. Resultados individuales “BDI” con Índice de Cambio Fiable Grupo Clínico (Continuación)**

Caso	Grupo Diagnóstico	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
38	T. Adapta. Animo Mixto	-7	-1,79	-8,1	Sin cambio dentro de la funcionalidad pre
39	T. Dep. No Especificado	-15	-3,84	-2,1	Recuperado
40	T. Distímico	-1	-0,26	4,9	Sin cambio dentro de la disfuncionalidad
41	T. Distímico	-11	-2,82	-5,1	Recuperado
42	T. Distímico	-14	-3,58	-4,1	Recuperado
43	T. Adapta. Animo Deprimido	-17	-4,35	-9,1	Recuperado
44	T. Adapta. Animo Mixto	-3	-0,77	-4,1	Sin cambio dentro de la funcionalidad pre
45	T. Adapta. Animo Mixto	-10	-2,56	-9,1	Recuperado
46	T. Adapta. Animo Mixto	-13	-3,33	-11,1	Recuperado
47	T. Distímico	-16	-4,10	5,9	Mejoría dentro de la disfuncionalidad
48	T. Distímico	-11	-2,82	-5,1	Recuperado
49	T. Adapta. Animo Deprimido	-11	-2,82	-9,1	Recuperado
50	T. Dep. Mayor Episodio Único	-5	-1,28	24,9	Sin cambio dentro de la disfuncionalidad
51	T. Distímico	-18	-4,61	-3,1	Recuperado
52	T. Distímico	-13	-3,33	4,9	Mejoría dentro de la disfuncionalidad
53	T. Adapta. Animo Mixto	-9	-2,30	-10,1	Recuperado pre
54	T. Distímico	-2	-0,51	1,9	Sin cambio dentro de la disfuncionalidad
55	T. Distímico	-11	-2,82	3,9	Mejoría dentro de la disfuncionalidad
56	T. Adapta. Animo Deprimido	-12	-3,07	-6,1	Recuperado
57	T. Adapta. Animo Deprimido	-21	-5,38	-9,1	Recuperado
58	T. Distímico	-14	-3,58	-2,1	Recuperado
59	T. Adapta. Animo Deprimido	-13	-3,33	-5,1	Recuperado
60	T. Dep. Mayor Recidivante	-3	-0,77	9,9	Sin cambio dentro de la disfuncionalidad
61	T. Distímico	-14	-3,58	-1,1	Recuperado
62	T. Dep. Mayor Recidivante	-17	-4,35	6,9	Mejoría dentro de la disfuncionalidad
63	T. Dep. Mayor Episodio Único	-18	-4,61	9,9	Mejoría dentro de la disfuncionalidad
64	T. Dep. Mayor Recidivante	-7	-1,79	5,9	Sin cambio dentro de la disfuncionalidad
65	T. Distímico	-15	-3,84	6,9	Mejoría dentro de la disfuncionalidad
66	T. Distímico	-14	-3,58	-1,1	Recuperado
67	T. Distímico	-16	-4,10	9,9	Mejoría dentro de la disfuncionalidad
68	T. Dep. Mayor Recidivante	-14	-3,58	7,9	Mejoría dentro de la disfuncionalidad
69	T. Distímico	-18	-4,61	5,9	Mejoría dentro de la disfuncionalidad
70	T. Dep. Mayor Episodio Único	-13	-3,33	4,9	Mejoría dentro de la disfuncionalidad
71	T. Dep. Mayor Episodio Único	-15	-3,84	-2,1	Recuperado
72	T. Distímico	-14	-3,58	-2,1	Recuperado
73	T. Distímico	-15	-3,84	-0,1	Recuperado
74	T. Dep. Mayor Recidivante	-9	-2,30	4,9	Mejoría dentro de la disfuncionalidad
75	T. Adapta. Animo Deprimido	-13	-3,33	-9,1	Recuperado
76	T. Dep. Mayor Recidivante	-16	-4,10	5,9	Mejoría dentro de la disfuncionalidad
77	T. Distímico	-11	-2,82	-7,1	Recuperado
78	T. Distímico	-14	-3,58	8,9	Mejoría dentro de la disfuncionalidad
79	T. Adapta. Animo Deprimido	-12	-3,07	-5,1	Mejoría dentro de la disfuncionalidad
80	T. Distímico	-15	-3,84	-0,1	Recuperado

\* PCF – Punto de Corte Funcional (BDI= 18,10)

**Tabla A3E3.1. Resultados individuales “BDI” con Índice de Cambio Fiable Grupo Clínico (Continuación)**

Caso	Grupo Diagnóstico	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
81	T. Distímico	-17	-4,35	4,9	Mejoría dentro de la disfuncionalidad
82	T. Dep. Mayor Recidivante	-15	-3,84	5,9	Mejoría dentro de la disfuncionalidad
83	T. Adapta. Animo Mixto	-22	-5,63	-9,1	Recuperado
84	T. Distímico	-13	-3,33	1,9	Mejoría dentro de la disfuncionalidad
85	T. Adapta. Animo Mixto	-12	-3,07	-8,1	Recuperado
86	T. Dep. Mayor Episodio Único	-18	-4,61	5,9	Mejoría dentro de la disfuncionalidad
87	T. Dep. No Especificado	-1	-0,26	14,9	Sin cambio dentro de la disfuncionalidad
88	T. Dep. Mayor Recidivante	-14	-3,58	-0,1	Recuperado
89	T. Distímico	-5	-1,28	9,9	Sin cambio dentro de la disfuncionalidad
90	T. Adapta. Animo Deprimido	-13	-3,33	-5,1	Recuperado
91	T. Adapta. Animo Mixto	-23	-5,89	-8,1	Recuperado
92	T. Adapta. Animo Deprimido	-20	-5,12	-9,1	Recuperado
93	T. Dep. Mayor Recidivante	-6	-1,54	8,9	Sin cambio dentro de la disfuncionalidad
94	T. Dep. Mayor Recidivante	-3	-0,77	7,9	Sin cambio dentro de la disfuncionalidad
95	T. Adapta. Animo Deprimido	-15	-3,84	-9,1	Recuperado
96	T. Distímico	-1	-0,26	1,9	Sin cambio dentro de la disfuncionalidad
97	T. Distímico	-5	-1,28	9,9	Sin cambio dentro de la disfuncionalidad
98	T. Distímico	-11	-2,82	-6,1	Recuperado
99	T. Dep. Mayor Recidivante	-14	-3,58	0,9	Mejoría dentro de la disfuncionalidad
100	T. Adapta. Animo Mixto	-12	-3,07	-11,1	Recuperado

\* PCF – Punto de Corte Funcional (BDI= 18,10)

**Tabla A3E3.2. Resultados individuales “BDI” con Índice de Cambio Fiable Grupo No Clínico**

Caso	Edad / Sexo	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
1	20/Mujer	5	1,28	-12,1	Sin cambio dentro de la funcionalidad
2	22/Mujer	0	0,00	-11,1	Sin cambio dentro de la funcionalidad
3	22/Mujer	-1	-0,26	-17,1	Sin cambio dentro de la funcionalidad
4	22/Mujer	3	0,77	-0,1	Sin cambio dentro de la funcionalidad
5	23/Mujer	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
6	23/Mujer	-1	-0,26	-13,1	Sin cambio dentro de la funcionalidad
7	23/Mujer	2	0,51	-7,1	Sin cambio dentro de la funcionalidad
8	24/Mujer	2	0,51	-9,1	Sin cambio dentro de la funcionalidad
9	24/Mujer	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
10	25/Mujer	1	0,26	-10,1	Sin cambio dentro de la funcionalidad
11	25/Hombre	-2	-0,51	-11,1	Sin cambio dentro de la funcionalidad
12	25/Hombre	10	2,56	-0,1	Empeoramiento dentro de funcionalidad
13	26/Mujer	0	0,00	-10,1	Sin cambio dentro de la funcionalidad
14	26/Mujer	-6	-1,54	-12,1	Sin cambio dentro de la funcionalidad
15	27/Mujer	5	1,28	-3,1	Sin cambio dentro de la funcionalidad
16	27/Mujer	-6	-1,54	-16,1	Sin cambio dentro de la funcionalidad
17	27/Mujer	-5	-1,28	-18,1	Sin cambio dentro de la funcionalidad
18	27/Hombre	-3	-0,77	-15,1	Sin cambio dentro de la funcionalidad
19	28/Mujer	-7	-1,79	-15,1	Sin cambio dentro de la funcionalidad
20	28/Hombre	-7	-1,79	-4,1	Sin cambio dentro de la funcionalidad
21	29/Hombre	-1	-0,26	-15,1	Sin cambio dentro de la funcionalidad
22	30/Mujer	2	0,51	-11,1	Sin cambio dentro de la funcionalidad
23	30/Hombre	8	2,05	-1,1	Empeoramiento dentro de funcionalidad
24	30/Hombre	8	2,05	-0,1	Empeoramiento dentro de funcionalidad
25	30/Hombre	10	2,56	-5,1	Empeoramiento dentro de funcionalidad
26	31/Mujer	4	1,02	-13,1	Sin cambio dentro de la funcionalidad
27	32/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
28	32/Mujer	-2	-0,51	-14,1	Sin cambio dentro de la funcionalidad
29	32/Hombre	1	0,26	-16,1	Sin cambio dentro de la funcionalidad
30	32/Hombre	11	2,82	-3,1	Empeoramiento dentro de funcionalidad
31	32/Hombre	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
32	33/Mujer	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
33	33/Mujer	-1	-0,26	-12,1	Sin cambio dentro de la funcionalidad
34	33/Mujer	-1	-0,26	-11,1	Sin cambio dentro de la funcionalidad
35	33/Hombre	-3	-0,77	-8,1	Sin cambio dentro de la funcionalidad
36	34/Mujer	3	0,77	-12,1	Sin cambio dentro de la funcionalidad
37	34/Mujer	-3	-0,77	-9,1	Sin cambio dentro de la funcionalidad

\* PCF – Punto de Corte Funcional (BDI= 18,10)

**Tabla A3E3.2.Resultados individuales “BDI” con Índice de Cambio Fiable Grupo No Clínico (Continuación)**

Caso	Edad/ Sexo	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
38	34/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
39	35/Mujer	-1	-0,26	-12,1	Sin cambio dentro de la funcionalidad
40	35/Mujer	-4	-1,02	-12,1	Sin cambio dentro de la funcionalidad
41	35/Mujer	0	0,00	-5,1	Sin cambio dentro de la funcionalidad
42	35/Mujer	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
43	35/Mujer	-10	-2,56	-16,1	Recuperado
44	35/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
45	35/Mujer	-2	-0,51	-12,1	Sin cambio dentro de la funcionalidad
46	35/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
47	35/Hombre	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
48	36/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
49	36/Mujer	3	0,77	-7,1	Sin cambio dentro de la funcionalidad
50	36/Hombre	-1	-0,26	-17,1	Sin cambio dentro de la funcionalidad
51	37/Mujer	0	0,00	-12,1	Sin cambio dentro de la funcionalidad
52	38/Mujer	2	0,51	-11,1	Sin cambio dentro de la funcionalidad
53	38/Mujer	0	0,00	-11,1	Sin cambio dentro de la funcionalidad
54	39/Mujer	1	0,26	-16,1	Sin cambio dentro de la funcionalidad
55	39/Mujer	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
56	39/Mujer	-4	-1,02	-11,1	Sin cambio dentro de la funcionalidad
57	39/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
58	39/Hombre	-1	-0,26	-12,1	Sin cambio dentro de la funcionalidad
59	39/Hombre	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
60	40/Mujer	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
61	40/Mujer	-2	-0,51	-17,1	Sin cambio dentro de la funcionalidad
62	40/Hombre	0	0,00	-3,1	Sin cambio dentro de la funcionalidad
63	41/Mujer	-2	-0,51	-12,1	Sin cambio dentro de la funcionalidad
64	41/Hombre	0	0,00	-10,1	Sin cambio dentro de la funcionalidad
65	41/Hombre	-1	-0,26	-12,1	Sin cambio dentro de la funcionalidad
66	43/Mujer	3	0,77	-12,1	Sin cambio dentro de la funcionalidad
67	43/Mujer	-1	-0,26	-12,1	Sin cambio dentro de la funcionalidad
68	44/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
69	44/Mujer	0	0,00	-16,1	Sin cambio dentro de la funcionalidad
70	45/Hombre	-1	-0,26	-15,1	Sin cambio dentro de la funcionalidad
71	45/Hombre	4	1,02	-10,1	Sin cambio dentro de la funcionalidad
72	45/Hombre	3	0,77	-15,1	Sin cambio dentro de la funcionalidad
73	45/Hombre	1	0,26	-13,1	Sin cambio dentro de la funcionalidad
74	46/Mujer	2	0,51	-13,1	Sin cambio dentro de la funcionalidad
75	47/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
76	49/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
77	49/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
78	49/Mujer	-2	-0,51	-18,1	Sin cambio dentro de la funcionalidad
79	49/Hombre	0	0,00	-15,1	Sin cambio dentro de la funcionalidad
80	50/Mujer	1	0,26	-9,1	Sin cambio dentro de la funcionalidad

PCF – Punto de Corte Funcional (BDI= 18,10)

**Tabla A3E3.2.- Resultados individuales “BDI” con Índice de Cambio Fiable Grupo No Clínico (Continuación)**

Caso	Edad / Sexo	Pre –Post	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
81	50/Hombre	-4	-1,02	-14,1	Sin cambio dentro de la funcionalidad
82	51/Mujer	0	0,00	-18,1	Sin cambio dentro de la funcionalidad
83	51/Hombre	0	0,00	-14,1	Sin cambio dentro de la funcionalidad
84	53/Mujer	-3	-0,77	-18,1	Sin cambio dentro de la funcionalidad
85	53/Mujer	-2	-0,51	-17,1	Sin cambio dentro de la funcionalidad
86	53/Hombre	0	0,00	-17,1	Sin cambio dentro de la funcionalidad
87	54/Mujer	0	0,00	-15,1	Sin cambio dentro de la funcionalidad
88	54/Mujer	5	1,28	-12,1	Sin cambio dentro de la funcionalidad
89	54/Mujer	3	0,77	-15,1	Sin cambio dentro de la funcionalidad
90	54/Mujer	1	0,26	-12,1	Sin cambio dentro de la funcionalidad
91	54/Hombre	-6	-1,54	-10,1	Sin cambio dentro de la funcionalidad
92	55/Mujer	-1	-0,26	-17,1	Sin cambio dentro de la funcionalidad
93	55/Hombre	-5	-1,28	-17,1	Sin cambio dentro de la funcionalidad
94	58/Hombre	0	0,00	-12,1	Sin cambio dentro de la funcionalidad
95	58/Hombre	0	0,00	-13,1	Sin cambio dentro de la funcionalidad
96	59/Mujer	0	0,00	-15,1	Sin cambio dentro de la funcionalidad
97	60/Mujer	-5	-1,28	-17,1	Sin cambio dentro de la funcionalidad
98	60/Mujer	-2	-0,51	-18,1	Sin cambio dentro de la funcionalidad
99	62/Mujer	0	0,00	-11,1	Sin cambio dentro de la funcionalidad
100	62/Mujer	0	0,00	-15,1	Sin cambio dentro de la funcionalidad

\* PCF – Punto de Corte Funcional (BDI= 18,10)

### A3E3. Tablas adicionales del Estudio 3

Tabla A3E3.1.- Índices de cambio individual alternativos a ICF<sub>JT</sub> en grupo clínico

Nº	Dif.	GLN	EN inf.	EN sup.	HA	TEI	DIT	ESG	DTW	IPP inf	IPP sup.	IPI inf.	IPI sup.	PM 50%	PM 25%
1	-1	-0,57	20,15	33,53	-2,45	-0,12	-0,18	-0,3	-0,36	20,27	22,23	14,75	27,75	0,13	0,04
2	-9	-1,98	12,68	26,06	-3,89	-1,11	-1,64	-2,68	-3,26	12,12	16,55	7,53	21,14	-8,18	0,53
3	-14	-2,77	22,64	36,02	-4,79	-1,73	-2,55	-4,16	-5,07	22,85	24,25	17,09	30,01	1,28	0,48
4	-15	-2,90	25,13	38,51	-4,97	-1,85	-2,73	-4,46	-5,43	25,20	26,51	19,4	32,31	1,08	0,47
5	-15	-2,85	25,96	39,34	-4,97	-1,85	-2,73	-4,46	-5,43	25,92	27,33	20,16	33,09	1,01	0,45
6	-14	-2,77	22,64	36,02	-4,79	-1,73	-2,55	-4,16	-5,07	22,85	24,25	17,09	30,01	1,28	0,48
7	-11	-2,33	16,00	29,38	-4,25	-1,36	-2,01	-3,27	-3,98	15,77	19,05	10,77	24,04	3,79	0,52
8	-3	-0,33	13,51	26,89	-2,81	-0,37	-0,55	-0,89	-1,09	13,04	17,17	8,35	21,86	-3,00	0,17
9	-12	-2,55	16,83	30,21	-4,43	-1,48	-2,19	-3,57	-4,34	16,68	19,68	11,58	24,78	3,08	0,55
10	-16	-3,49	19,32	32,70	-5,15	-1,97	-2,92	-4,76	-5,79	19,38	21,58	13,96	27,00	2,32	0,64
11	-19	-3,42	35,09	48,47	-5,69	-2,34	-3,46	-5,65	-6,88	32,98	37,16	28,31	41,83	0,73	0,43
12	-14	-2,59	25,96	39,34	-4,79	-1,73	-2,55	-4,16	-5,07	25,92	27,33	20,16	33,09	0,94	0,42
13	-15	-2,58	30,94	44,32	-4,97	-1,85	-2,73	-4,46	-5,43	29,85	32,62	24,66	37,80	0,72	0,38
14	-10	-2,29	11,85	25,23	-4,07	-1,23	-1,82	-2,97	-3,62	11,21	15,93	6,72	20,42	-4,76	0,63
15	-3	-0,14	16,83	30,21	-2,81	-0,37	-0,55	-0,89	-1,09	16,68	19,68	11,58	24,78	0,77	0,14
16	-22	-4,27	34,26	47,64	-6,22	-2,71	-4,01	-6,54	-7,96	32,36	36,25	27,59	41,02	0,88	0,51
17	-11	-2,33	16,00	29,38	-4,25	-1,36	-2,01	-3,27	-3,98	15,77	19,05	10,77	24,04	3,79	0,52
18	-19	-3,51	33,43	46,81	-5,69	-2,34	-3,46	-5,65	-6,88	31,73	35,34	26,86	40,21	0,79	0,45
19	-4	-0,41	16,83	30,21	-2,99	-0,49	-0,73	-1,19	-1,45	16,68	19,68	11,58	24,78	1,03	0,18
20	-14	-2,81	21,81	35,19	-4,79	-1,73	-2,55	-4,16	-5,07	22,01	23,56	16,31	29,25	1,41	0,50
21	-9	-0,98	30,94	44,32	-3,89	-1,11	-1,64	-2,68	-3,26	29,85	32,62	24,66	37,80	0,43	0,23
22	-16	-2,76	32,60	45,98	-5,15	-1,97	-2,92	-4,76	-5,79	31,11	34,43	26,13	39,41	0,70	0,39
23	-10	-1,56	25,13	38,51	-4,07	-1,23	-1,82	-2,97	-3,62	25,2	26,51	19,4	32,31	0,72	0,31
24	-20	-3,83	32,60	45,98	-5,86	-2,46	-3,65	-5,95	-7,24	31,11	34,43	26,13	39,41	0,87	0,49
25	-14	-2,59	25,96	39,34	-4,79	-1,73	-2,55	-4,16	-5,07	25,92	27,33	20,16	33,09	0,94	0,42
26	-8	-0,62	32,60	45,98	-3,71	-0,99	-1,46	-2,38	-2,90	31,11	34,43	26,13	39,41	0,35	0,20
27	-12	-1,78	30,94	44,32	-4,43	-1,48	-2,19	-3,57	-4,34	29,85	32,62	24,66	37,8	0,57	0,31
28	-15	-2,90	25,13	38,51	-4,97	-1,85	-2,73	-4,46	-5,43	25,20	26,51	19,4	32,31	1,08	0,47
29	-2	0,62	25,96	39,34	-2,63	-0,25	-0,36	-0,59	-0,72	25,92	27,33	20,16	33,09	0,13	0,06
30	-18	-3,20	34,26	47,64	-5,51	-2,22	-3,28	-5,35	-6,52	32,36	36,25	27,59	41,02	0,72	0,42
31	-19	-3,61	31,77	45,15	-5,69	-2,34	-3,46	-5,65	-6,88	30,48	33,52	25,4	38,6	0,87	0,48
32	0	0,61	16,00	29,38	-2,27	0,00	0,00	0,00	0,00	15,77	19,05	10,77	24,04	0,00	0,00
33	-6	-0,31	28,45	41,83	-3,35	-0,74	-1,09	-1,78	-2,17	27,93	29,93	22,43	35,43	0,34	0,17
34	-17	-2,98	33,43	46,81	-5,33	-2,10	-3,10	-5,05	-6,15	31,73	35,34	26,86	40,21	0,71	0,40
35	-23	-4,99	25,96	39,34	-6,40	-2,83	-4,19	-6,84	-8,33	25,92	27,33	20,16	33,09	1,54	0,70
36	-17	-2,89	35,09	48,47	-5,33	-2,10	-3,10	-5,05	-6,15	32,98	37,16	28,31	41,83	0,66	0,39
37	-10	-2,11	15,17	28,55	-4,07	-1,23	-1,82	-2,97	-3,62	14,86	18,42	9,97	23,31	5,26	0,50
38	-7	-1,44	12,68	26,06	-3,53	-0,86	-1,28	-2,08	-2,53	12,12	16,55	7,53	21,14	-6,36	0,41
39	-15	-2,94	24,30	37,68	-4,97	-1,85	-2,73	-4,46	-5,43	24,45	25,72	18,63	31,54	1,16	0,48
40	-1	0,48	18,49	31,87	-2,45	-0,12	-0,18	-0,3	-0,36	18,48	20,94	13,17	26,25	0,17	0,04

Dif.= Diferencia Post-Pre Tratamiento; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; HA= Método de Hageman-Arrindel; TEI 1,96/0,20= Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para el Pronóstico Promedio; IPI=Intervalo para el Pronóstico Individual; PM= Porcentaje de Mejoría; inf: límite inferior, sup. Límite Superior.

**Tabla A3E3.1.-Índices de cambio individual alternativos a ICF<sub>JT</sub> en grupo clínico (Continuación)**

Nº	Dif.	GLN	EN inf.	EN sup.	HA	TEI	DIT	ESG	DTW	IPP inf	IPP sup.	IPI inf.	IPI sup.	PM 50%	PM 25%
41	-11	-2,19	18,49	31,87	-4,25	-1,36	-2,01	-3,27	-3,98	18,48	20,94	13,17	26,25	1,86	0,46
42	-14	-2,81	21,81	35,19	-4,79	-1,73	-2,55	-4,16	-5,07	22,01	23,56	16,31	29,25	1,41	0,50
43	-17	-3,71	20,15	33,53	-5,33	-2,10	-3,10	-5,05	-6,15	20,27	22,23	14,75	27,75	2,15	0,65
44	-3	-0,37	12,68	26,06	-2,81	-0,37	-0,55	-0,89	-1,09	12,12	16,55	7,53	21,14	-2,73	0,18
45	-10	-2,15	14,34	27,72	-4,07	-1,23	-1,82	-2,97	-3,62	13,95	17,80	9,16	22,58	11,11	0,53
46	-13	-2,91	15,17	28,55	-4,61	-1,60	-2,37	-3,87	-4,71	14,86	18,42	9,97	23,31	6,84	0,65
47	-16	-2,80	31,77	45,15	-5,15	-1,97	-2,92	-4,76	-5,79	30,48	33,52	25,40	38,60	0,73	0,40
48	-11	-2,19	18,49	31,87	-4,25	-1,36	-2,01	-3,27	-3,98	18,48	20,94	13,17	26,25	1,86	0,46
49	-11	-2,37	15,17	28,55	-4,25	-1,36	-2,01	-3,27	-3,98	14,86	18,42	9,97	23,31	5,79	0,55
50	-5	0,50	38,41	51,79	-3,17	-0,62	-0,91	-1,49	-1,81	35,47	40,82	31,17	45,11	0,17	0,10
51	-18	-3,66	25,96	39,34	-5,51	-2,22	-3,28	-5,35	-6,52	25,92	27,33	20,16	33,09	1,21	0,55
52	-13	-2,18	28,45	41,83	-4,61	-1,60	-2,37	-3,87	-4,71	27,93	29,93	22,43	35,43	0,73	0,36
53	-9	-1,98	12,68	26,06	-3,89	-1,11	-1,64	-2,68	-3,26	12,12	16,55	7,53	21,14	-8,18	0,53
54	-2	0,12	16,83	30,21	-2,63	-0,25	-0,36	-0,59	-0,72	16,68	19,68	11,58	24,78	0,51	0,09
55	-11	-1,78	25,96	39,34	-4,25	-1,36	-2,01	-3,27	-3,98	25,92	27,33	20,16	33,09	0,74	0,33
56	-12	-2,46	18,49	31,87	-4,43	-1,48	-2,19	-3,57	-4,34	18,48	20,94	13,17	26,25	2,03	0,50
57	-21	-4,60	23,47	36,85	-6,04	-2,59	-3,83	-6,24	-7,60	23,67	24,97	17,86	30,78	1,76	0,70
58	-14	-2,72	23,47	36,85	-4,79	-1,73	-2,55	-4,16	-5,07	23,67	24,97	17,86	30,78	1,18	0,47
59	-13	-2,64	20,15	33,53	-4,61	-1,60	-2,37	-3,87	-4,71	20,27	22,23	14,75	27,75	1,65	0,50
60	-3	0,27	24,30	37,68	-2,81	-0,37	-0,55	-0,89	-1,09	24,45	25,72	18,63	31,54	0,23	0,10
61	-14	-2,68	24,30	37,68	-4,79	-1,73	-2,55	-4,16	-5,07	24,45	25,72	18,63	31,54	1,09	0,45
62	-17	-2,98	33,43	46,81	-5,33	-2,10	-3,10	-5,05	-6,15	31,73	35,34	26,86	40,21	0,71	0,40
63	-18	-3,07	36,75	50,13	-5,51	-2,22	-3,28	-5,35	-6,52	34,23	38,99	29,75	43,47	0,65	0,39
64	-7	-0,80	24,30	37,68	-3,53	-0,86	-1,28	-2,08	-2,53	24,45	25,72	18,63	31,54	0,54	0,23
65	-15	-2,54	31,77	45,15	-4,97	-1,85	-2,73	-4,46	-5,43	30,48	33,52	25,40	38,60	0,68	0,38
66	-14	-2,68	24,30	37,68	-4,79	-1,73	-2,55	-4,16	-5,07	24,45	25,72	18,63	31,54	1,09	0,45
67	-16	-2,62	35,09	48,47	-5,15	-1,97	-2,92	-4,76	-5,79	32,98	37,16	28,31	41,83	0,62	0,36
68	-14	-2,27	31,77	45,15	-4,79	-1,73	-2,55	-4,16	-5,07	30,48	33,52	25,40	38,60	0,64	0,35
69	-18	-3,25	33,43	46,81	-5,51	-2,22	-3,28	-5,35	-6,52	31,73	35,34	26,86	40,21	0,75	0,43
70	-13	-2,18	28,45	41,83	-4,61	-1,60	-2,37	-3,87	-4,71	27,93	29,93	22,43	35,43	0,73	0,36
71	-15	-2,94	24,30	37,68	-4,97	-1,85	-2,73	-4,46	-5,43	24,45	25,72	18,63	31,54	1,16	0,48
72	-14	-2,72	23,47	36,85	-4,79	-1,73	-2,55	-4,16	-5,07	23,67	24,97	17,86	30,78	1,18	0,47
73	-15	-2,85	25,96	39,34	-4,97	-1,85	-2,73	-4,46	-5,43	25,92	27,33	20,16	33,09	1,01	0,45
74	-9	-1,29	25,13	38,51	-3,89	-1,11	-1,64	-2,68	-3,26	25,2	26,51	19,40	32,31	0,65	0,28
75	-13	-2,82	16,83	30,21	-4,61	-1,60	-2,37	-3,87	-4,71	16,68	19,68	11,58	24,78	3,33	0,59
76	-16	-2,80	31,77	45,15	-5,15	-1,97	-2,92	-4,76	-5,79	30,48	33,52	25,40	38,60	0,73	0,40
77	-11	-2,28	16,83	30,21	-4,25	-1,36	-2,01	-3,27	-3,98	16,68	19,68	11,58	24,78	2,82	0,50
78	-14	-2,22	32,60	45,98	-4,79	-1,73	-2,55	-4,16	-5,07	31,11	34,43	26,13	39,41	0,61	0,34
79	-12	-2,42	19,32	32,7	-4,43	-1,48	-2,19	-3,57	-4,34	19,38	21,58	13,96	27,00	1,74	0,48
80	-15	-2,85	25,96	39,34	-4,97	-1,85	-2,73	-4,46	-5,43	25,92	27,33	20,16	33,09	1,01	0,45
81	-17	-3,07	31,77	45,15	-5,33	-2,10	-3,10	-5,05	-6,15	30,48	33,52	25,40	38,60	0,78	0,43
82	-15	-2,58	30,94	44,32	-4,97	-1,85	-2,73	-4,46	-5,43	29,85	32,62	24,66	37,80	0,72	0,38
83	-22	-4,82	24,3	37,68	-6,22	-2,71	-4,01	-6,54	-7,96	24,45	25,72	18,63	31,54	1,71	0,71
84	-13	-2,32	25,96	39,34	-4,61	-1,60	-2,37	-3,87	-4,71	25,92	27,33	20,16	33,09	0,87	0,39
85	-12	-2,55	16,83	30,21	-4,43	-1,48	-2,19	-3,57	-4,34	16,68	19,68	11,58	24,78	3,08	0,55

Dif.= Diferencia Post-Pre Tratamiento; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; HA= Método de Hageman-Arrindel; TEI 1,96/0,20= Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para el Pronóstico Promedio; IPI=Intervalo para el Pronóstico Individual; PM= Porcentaje de Mejoría; inf: límite inferior, sup. Límite Superior.

**Tabla A3E3.1. Índices de cambio individual alternativos a ICF<sub>JT</sub> en grupo clínico (Continuación)**

Nº	Dif.	GLN	EN inf.	EN sup.	HA	TEI	DIT	ESG	DTW	IPP inf	IPP sup.	IPI inf.	IPI sup.	PM 50%	PM 25%
86	-18	-3,25	33,43	46,81	-5,51	-2,22	-3,28	-5,35	-6,52	31,73	35,34	26,86	40,21	0,75	0,43
87	-1	0,94	26,79	40,17	-2,45	-0,12	-0,18	-0,30	-0,36	26,60	28,18	20,92	33,86	0,06	0,03
88	-14	-2,63	25,13	38,51	-4,79	-1,73	-2,55	-4,16	-5,07	25,20	26,51	19,40	32,31	1,01	0,44
89	-5	-0,18	25,96	39,34	-3,17	-0,62	-0,91	-1,49	-1,81	25,92	27,33	20,16	33,09	0,34	0,15
90	-13	-2,64	20,15	33,53	-4,61	-1,60	-2,37	-3,87	-4,71	20,27	22,23	14,75	27,75	1,65	0,50
91	-23	-4,99	25,96	39,34	-6,40	-2,83	-4,19	-6,84	-8,33	25,92	27,33	20,16	33,09	1,54	0,70
92	-20	-4,37	22,64	36,02	-5,86	-2,46	-3,65	-5,95	-7,24	22,85	24,25	17,09	30,01	1,83	0,69
93	-6	-0,45	25,96	39,34	-3,35	-0,74	-1,09	-1,78	-2,17	25,92	27,33	20,16	33,09	0,40	0,18
94	-3	0,18	22,64	36,02	-2,81	-0,37	-0,55	-0,89	-1,09	22,85	24,25	17,09	30,01	0,28	0,10
95	-15	-3,26	18,49	31,87	-4,97	-1,85	-2,73	-4,46	-5,43	18,48	20,94	13,17	26,25	2,54	0,63
96	-1	0,35	16,00	29,38	-2,45	-0,12	-0,18	-0,30	-0,36	15,77	19,05	10,77	24,04	0,34	0,05
97	-5	-0,18	25,96	39,34	-3,17	-0,62	-0,91	-1,49	-1,81	25,92	27,33	20,16	33,09	0,34	0,15
98	-11	-2,24	17,66	31,04	-4,25	-1,36	-2,01	-3,27	-3,98	17,58	20,31	12,38	25,51	2,24	0,48
99	-14	-2,59	25,96	39,34	-4,79	-1,73	-2,55	-4,16	-5,07	25,92	27,33	20,16	33,09	0,94	0,42
100	-12	-2,69	14,34	27,72	-4,43	-1,48	-2,19	-3,57	-4,34	13,95	17,80	9,16	22,58	13,33	0,63

Dif.= Diferencia Post-Pre Tratamiento; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; HA= Método de Hageman-Arrindel; TEI 1,96/0,20= Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para el Pronóstico Promedio; IPI=Intervalo para el Pronóstico Individual; PM= Porcentaje de Mejoría; inf: límite inferior, sup. Límite Superior.

**Tabla A3E3.2. Índices de cambio individual alternativos a ICF<sub>IT</sub> en grupo No Clínico**

Nº	Dif.	GLN	EN inf.	EN sup.	HA	TEI	DIT	ESG	DTW	IPP inf	IPP sup.	IPI inf.	IPI sup.	PM 50%	PM 25%
1	5	1,04	-1,70	5,14		1,19	1,49	1,49	1,81	1,14	2,95	-4,44	8,54	0,29	-5,00
2	0	-0,02	3,28	10,12		0,00	0,00	0,00	0,00	5,97	7,34	0,20	13,12	0,00	0,00
3	-1	-0,52	-0,87	5,97		-0,24	-0,3	-0,3	-0,36	2,01	3,62	-3,66	9,29	-0,06	0,50
4	3	1,14	9,92	16,76		0,72	0,89	0,89	1,09	11,19	14,41	6,17	19,43	0,97	-0,20
5	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
6	-1	-0,34	2,45	9,29		-0,24	-0,3	-0,3	-0,36	5,24	6,53	-0,57	12,34	-0,08	0,17
7	2	0,60	4,94	11,78		0,48	0,59	0,59	0,72	7,34	9,05	1,71	14,67	0,22	-0,22
8	2	0,51	3,28	10,12		0,48	0,59	0,59	0,72	5,97	7,34	0,20	13,12	0,18	-0,29
9	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
10	1	0,24	3,28	10,12		0,24	0,30	0,30	0,36	5,97	7,34	0,20	13,12	0,09	-0,14
11	-2	-0,47	4,94	11,78		-0,48	-0,59	-0,59	-0,72	7,34	9,05	1,71	14,67	-0,22	0,22
12	10	2,70	4,11	10,95		2,39	2,97	2,97	3,62	6,66	8,18	0,96	13,89	0,99	-1,25
13	0	0,02	4,11	10,95		0,00	0,00	0,00	0,00	6,66	8,18	0,96	13,89	0,00	0,00
14	-6	-1,40	7,43	14,27		-1,43	-1,78	-1,78	-2,17	9,29	11,7	3,96	17,03	-0,98	0,5
15	5	1,45	5,77	12,61		1,19	1,49	1,49	1,81	8,00	9,92	2,46	15,46	0,62	-0,5
16	-6	-1,58	4,11	10,95		-1,43	-1,78	-1,78	-2,17	6,66	8,18	0,96	13,89	-0,59	0,75
17	-5	-1,45	1,62	8,46		-1,19	-1,49	-1,49	-1,81	4,48	5,76	-1,33	11,57	-0,38	1,00
18	-3	-0,87	2,45	9,29		-0,72	-0,89	-0,89	-1,09	5,24	6,53	-0,57	12,34	-0,25	0,50
19	-7	-1,76	5,77	12,61		-1,67	-2,08	-2,08	-2,53	8,00	9,92	2,46	15,46	-0,86	0,70
20	-7	-1,26	14,9	21,74		-1,67	-2,08	-2,08	-2,53	14,93	19,89	10,51	24,3	2,41	0,33
21	-1	-0,43	0,79	7,63		-0,24	-0,30	-0,30	-0,36	3,69	5,01	-2,11	10,81	-0,07	0,25
22	2	0,42	1,62	8,46		0,48	0,59	0,59	0,72	4,48	5,76	-1,33	11,57	0,15	-0,40
23	8	2,21	4,94	11,78		1,91	2,38	2,38	2,90	7,34	9,05	1,71	14,67	0,88	-0,89
24	8	2,25	5,77	12,61		1,91	2,38	2,38	2,90	8,00	9,92	2,46	15,46	0,99	-0,80
25	10	2,47	-0,04	6,80		2,39	2,97	2,97	3,62	2,86	4,31	-2,88	10,05	0,66	-3,33
26	4	0,77	-1,70	5,14		0,96	1,19	1,19	1,45	1,14	2,95	-4,44	8,54	0,23	-4,00
27	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
28	-2	-0,60	2,45	9,29		-0,48	-0,59	-0,59	-0,72	5,24	6,53	-0,57	12,34	-0,17	0,33
29	1	-0,03	-1,70	5,14		0,24	0,30	0,30	0,36	1,14	2,95	-4,44	8,54	0,06	-1,00
30	11	2,78	0,79	7,63		2,63	3,27	3,27	3,98	3,69	5,01	-2,11	10,81	0,78	-2,75
31	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
32	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
33	-1	-0,29	3,28	10,12		-0,24	-0,30	-0,30	-0,36	5,97	7,34	0,2	13,12	-0,09	0,14
34	-1	-0,24	4,11	10,95		-0,24	-0,30	-0,30	-0,36	6,66	8,18	0,96	13,89	-0,10	0,13
35	-3	-0,55	8,26	15,10		-0,72	-0,89	-0,89	-1,09	9,92	12,60	4,70	17,83	-0,59	0,23
36	3	0,60	-0,04	6,80		0,72	0,89	0,89	1,09	2,86	4,31	-2,88	10,05	0,20	-1,00
37	-3	-0,60	7,43	14,27		-0,72	-0,89	-0,89	-1,09	9,29	11,7	3,96	17,03	-0,49	0,25
38	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
39	-1	-0,29	3,28	10,12		-0,24	-0,30	-0,30	-0,36	5,97	7,34	0,20	13,12	-0,09	0,14
40	-4	-0,96	5,77	12,61		-0,96	-1,19	-1,19	-1,45	8,00	9,92	2,46	15,46	-0,49	0,40

Dif.= Diferencia Post-Pre Tratamiento; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; HA= Método de Hageman-Arrindel; TEI 1,96/0,20= Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para el Pronóstico Promedio; IPI=Intervalo para el Pronóstico Individual; PM= Porcentaje de Mejoría; inf: límite inferior, sup. Límite Superior.

**Tabla A3E3.2. Índices de cambio individual alternativos a ICF<sub>JT</sub> en grupo No clínico (Continuación)**

Nº	Dif.	GLN	EN inf.	EN sup.	HA	TEI	DIT	ESG	DTW	IPP inf	IPP sup.	IPI inf.	IPI sup.	PM 50%	PM 25%
41	0	0,25	8,26	15,10		0,00	0,00	0,00	0,00	9,92	12,6	4,70	17,83	0,00	0,00
42	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
43	-10	-2,47	7,43	14,27		-2,39	-2,97	-2,97	-3,62	9,29	11,70	3,96	17,03	-1,64	0,83
44	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
45	-2	-0,51	4,11	10,95		-0,48	-0,59	-0,59	-0,72	6,66	8,18	0,96	13,89	-0,20	0,25
46	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
47	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
48	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
49	3	0,83	4,11	10,95		0,72	0,89	0,89	1,09	6,66	8,18	0,96	13,89	0,30	-0,38
50	-1	-0,52	-0,87	5,97		-0,24	-0,30	-0,30	-0,36	2,01	3,62	-3,66	9,29	-0,06	0,50
51	0	-0,07	2,45	9,29		0,00	0,00	0,00	0,00	5,24	6,53	-0,57	12,34	0,00	0,00
52	2	0,42	1,62	8,46		0,48	0,59	0,59	0,72	4,48	5,76	-1,33	11,57	0,15	-0,40
53	0	-0,02	3,28	10,12		0,00	0,00	0,00	0,00	5,97	7,34	0,2	13,12	0,00	0,00
54	1	-0,03	-1,70	5,14		0,24	0,30	0,30	0,36	1,14	2,95	-4,44	8,54	0,06	-1,00
55	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
56	-4	-0,91	6,60	13,44		-0,96	-1,19	-1,19	-1,45	8,65	10,81	3,21	16,24	-0,56	0,36
57	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
58	-1	-0,29	3,28	10,12		-0,24	-0,30	-0,30	-0,36	5,97	7,34	0,20	13,12	-0,09	0,14
59	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
60	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
61	-2	-0,74	-0,04	6,80		-0,48	-0,59	-0,59	-0,72	2,86	4,31	-2,88	10,05	-0,13	0,67
62	0	0,34	9,92	16,76		0,00	0,00	0,00	0,00	11,19	14,41	6,17	19,43	0,00	0,00
63	-2	-0,51	4,11	10,95		-0,48	-0,59	-0,59	-0,72	6,66	8,18	0,96	13,89	-0,2	0,25
64	0	0,02	4,11	10,95		0,00	0,00	0,00	0,00	6,66	8,18	0,96	13,89	0,00	0,00
65	-1	-0,29	3,28	10,12		-0,24	-0,30	-0,30	-0,36	5,97	7,34	0,20	13,12	-0,09	0,14
66	3	0,60	-0,04	6,80		0,72	0,89	0,89	1,09	2,86	4,31	-2,88	10,05	0,2	-1,00
67	-1	-0,29	3,28	10,12		-0,24	-0,30	-0,30	-0,36	5,97	7,34	0,20	13,12	-0,09	0,14
68	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
69	0	-0,25	-0,87	5,97		0,00	0,00	0,00	0,00	2,01	3,62	-3,66	9,29	0,00	0,00
70	-1	-0,43	0,79	7,63		-0,24	-0,30	-0,30	-0,36	3,69	5,01	-2,11	10,81	-0,07	0,25
71	4	0,91	0,79	7,63		0,96	1,19	1,19	1,45	3,69	5,01	-2,11	10,81	0,28	-1,00
72	3	0,46	-2,53	4,31		0,72	0,89	0,89	1,09	0,26	2,30	-5,22	7,78	0,17	0,00
73	1	0,11	0,79	7,63		0,24	0,30	0,30	0,36	3,69	5,01	-2,11	10,81	0,07	-0,25
74	2	0,33	-0,04	6,80		0,48	0,59	0,59	0,72	2,86	4,31	-2,88	10,05	0,13	-0,67
75	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
76	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
77	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
78	-2	-0,79	-0,87	5,97		-0,48	-0,59	-0,59	-0,72	2,01	3,62	-3,66	9,29	-0,12	1,00
79	0	-0,20	-0,04	6,80		0,00	0,00	0,00	0,00	2,86	4,31	-2,88	10,05	0,00	0,00
80	1	0,29	4,11	10,95		0,24	0,30	0,30	0,36	6,66	8,18	0,96	13,89	0,10	-0,13
81	-4	-1,05	4,11	10,95		-0,96	-1,19	-1,19	-1,45	6,66	8,18	0,96	13,89	-0,40	0,50
82	0	-0,34	-2,53	4,31		0,00	0,00	0,00	0,00	0,26	2,30	-5,22	7,78	0,00	0,00
83	0	-0,16	0,79	7,63		0,00	0,00	0,00	0,00	3,69	5,01	-2,11	10,81	0,00	0,00
84	-3	-1,01	-0,04	6,80		-0,72	-0,89	-0,89	-1,09	2,86	4,31	-2,88	10,05	-0,20	1,00
85	-2	-0,74	-0,04	6,80		-0,48	-0,59	-0,59	-0,72	2,86	4,31	-2,88	10,05	-0,13	0,67

Dif.= Diferencia Post-Pre Tratamiento; GLN= Método Gullisken Lord-Novik; EN= Método Edward-Nunnally; HA= Método de Hageman-Arrindel; TEI 1,96/0,20= Tamaño del efecto individual; DIT 1,96/0,20= Diferencia individual Tipificada; ESG 1,96/0,20= Estadístico de sensibilidad Guyyatt; DTW= Diferencia Tipificada de Wyrwich; IPP= Intervalo para el Pronóstico Promedio; IPI=Intervalo para el Pronóstico Individual; PM= Porcentaje de Mejoría; inf. Límite inferior, sup. Límite Superior.

**Tabla A3E3.2. Índices de cambio individual alternativos a ICF<sub>JT</sub> en grupo No clínico (Continuación)**

Nº	Dif.	GLN	EN inf.	EN sup.	HA	TEI	DIT	ESG	DTW	IPP inf	IPP sup.	IPI inf.	IPI sup.	PM 50%	PM 25%
86	0	-0,30	-1,70	5,14		0,00	0,00	0,00	0,00	1,14	2,95	-4,44	8,54	0,00	0,00
87	0	-0,20	-0,04	6,80		0,00	0,00	0,00	0,00	2,86	4,31	-2,88	10,05	0,00	0,00
88	5	1,04	-1,70	5,14		1,19	1,49	1,49	1,81	1,14	2,95	-4,44	8,54	0,29	-5,00
89	3	0,46	-2,53	4,31		0,72	0,89	0,89	1,09	0,26	2,30	-5,22	7,78	0,17	0,00
90	1	0,15	1,62	8,46		0,24	0,30	0,30	0,36	4,48	5,76	-1,33	11,57	0,08	-0,20
91	-6	-1,31	9,09	15,93		-1,43	-1,78	-1,78	-2,17	10,56	13,51	5,44	18,63	-1,46	0,43
92	-1	-0,52	-0,87	5,97		-0,24	-0,30	-0,30	-0,36	2,01	3,62	-3,66	9,29	-0,06	0,50
93	-5	-1,41	2,45	9,29		-1,19	-1,49	-1,49	-1,81	5,24	6,53	-0,57	12,34	-0,41	0,83
94	0	-0,07	2,45	9,29		0,00	0,00	0,00	0,00	5,24	6,53	-0,57	12,34	0,00	0,00
95	0	-0,11	1,62	8,46		0,00	0,00	0,00	0,00	4,48	5,76	-1,33	11,57	0,00	0,00
96	0	-0,20	-0,04	6,80		0,00	0,00	0,00	0,00	2,86	4,31	-2,88	10,05	0,00	0,00
97	-5	-1,41	2,45	9,29		-1,19	-1,49	-1,49	-1,81	5,24	6,53	-0,57	12,34	-0,41	0,83
98	-2	-0,79	-0,87	5,97		-0,48	-0,59	-0,59	-0,72	2,01	3,62	-3,66	9,29	-0,12	1,00
99	0	-0,02	3,28	10,12		0,00	0,00	0,00	0,00	5,97	7,34	0,20	13,12	0,00	0,00
100	0	-0,20	-0,04	6,80		0,00	0,00	0,00	0,00	2,86	4,31	-2,88	10,05	0,00	0,00

**Dif.**= Diferencia Post-Pre Tratamiento; **GLN**= Método Gullisken Lord-Novik; **EN**= Método Edward-Nunnally; **HA**= Método de Hageman-Arrindel; **TEI 1,96/0,20**= Tamaño del efecto individual; **DIT 1,96/0,20**= Diferencia individual Tipificada; **ESG 1,96/0,20**= Estadístico de sensibilidad Guyyatt; **DTW**= Diferencia Tipificada de Wyrwich; **IPP**= Intervalo para el Pronóstico Promedio; **IPI**=Intervalo para el Pronóstico Individual; **PM**= Porcentaje de Mejoría; **inf**: límite inferior, **sup**. Límite Superior.

### A3E4. Tablas adicionales del Estudio 4

**Tabla A3E4.1. Resultados individuales “Peak Flow” con Índice de Cambio Fiable en Grupo Experimental**

Caso	Edad / Sexo	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
1	48/ Hombre	2,40	6,54	-3,23	Mejoría dentro de la disfuncionalidad
2	58/ Hombre	1,00	2,73	-4,43	Mejoría dentro de la disfuncionalidad
3	28/ Hombre	-1,30	-3,54	3,87	Empeoramiento dentro de funcionalidad
4	50/ Hombre	1,50	4,09	-4,53	Mejoría dentro de la disfuncionalidad
5	37/ Hombre	0,60	1,64	-4,23	Sin cambio dentro de la disfuncionalidad
6	58/ Hombre	0,10	0,27	-4,83	Sin cambio dentro de la disfuncionalidad
7	16/ Hombre	-0,10	-0,27	-0,93	Sin cambio dentro de la disfuncionalidad
8	39/ Hombre	-0,82	-2,24	-0,15	Empeoramiento dentro de disfuncionalidad
9	24/ Hombre	0,00	0,00	-0,03	Sin cambio dentro de la disfuncionalidad
10	49/ Hombre	0,24	0,65	-5,35	Sin cambio dentro de la disfuncionalidad
11	33/ Hombre	1,15	3,14	3,60	Recuperado
12	48/ Hombre	1,30	3,54	-2,03	Mejoría dentro de la disfuncionalidad
13	25/ Hombre	-1,20	-3,27	-2,53	Empeoramiento dentro de disfuncionalidad
14	37/ Mujer	-0,30	-1,23	-1,89	Sin cambio dentro de la disfuncionalidad
15	62/ Mujer	-0,40	-1,64	-3,19	Sin cambio dentro de la disfuncionalidad
16	19/ Mujer	0,30	1,23	0,91	Sin cambio dentro de la funcionalidad
17	29/ Mujer	0,00	0,00	2,11	Sin cambio dentro de la funcionalidad
18	53/ Mujer	0,30	1,23	0,21	Sin cambio dentro de la funcionalidad
19	26/ Mujer	0,90	3,69	4,31	Recuperado
20	39/ Mujer	0,90	3,69	1,61	Recuperado
21	44/ Mujer	0,70	2,87	-2,09	Mejoría dentro de la disfuncionalidad
22	50/ Mujer	1,00	4,10	-1,79	Mejoría dentro de la disfuncionalidad
23	48/ Mujer	-0,14	-0,57	1,45	Sin cambio dentro de la funcionalidad
24	44/ Mujer	0,65	2,67	-0,69	Mejoría dentro de la disfuncionalidad
25	48/ Mujer	0,25	1,03	-1,31	Sin cambio dentro de la disfuncionalidad
26	58/ Mujer	0,11	0,45	-3,43	Sin cambio dentro de la disfuncionalidad
27	52/ Mujer	3,30	13,54	0,21	Recuperado
28	19/ Mujer	1,20	4,92	0,91	Recuperado
29	45/ Mujer	-0,20	-0,82	-2,19	Sin cambio dentro de la disfuncionalidad
30	26/ Mujer	1,30	5,33	2,11	Recuperado
31	53/ Mujer	0,59	2,30	-1,20	Mejoria dentro de la disfuncionalidad
32	56/ Mujer	0,27	1,11	-2,50	Sin cambio dentro de la disfuncionalidad
33	47/ Mujer	0,44	1,81	-1,60	Sin cambio dentro de la disfuncionalidad
34	23/ Mujer	-1,55	-6,36	1,38	Empeoramiento dentro de funcionalidad
35	23/ Mujer	-0,70	-2,87	-2,93	Empeoramiento dentro de disfuncionalidad
36	56/ Mujer	-0,46	-1,89	-1,89	Sin cambio dentro de la disfuncionalidad
37	35/ Mujer	1,02	4,18	0,16	Recuperado
38	51/ Mujer	-1,03	-4,23	-1,90	Empeoramiento dentro de disfuncionalidad

\* PCF – Punto de Corte Funcional ( Hombres Peak Flow= 8,63; Mujeres Peak Flow= 5,59)

**Tabla A3E4.1.Resultados individuales “Peak Flow” con Índice de Cambio Fiable en Grupo Control**

Caso	Edad / Sexo	Post- Pre	ICF <sub>JT</sub>	Post – PCF *	Clasificación de situación final ( ICF <sub>JT</sub> +Punto de Corte Funcional)
1	54/ Hombre	-,80	-2,18	-3,33	Empeoramiento dentro de disfuncionalidad
2	16/ Hombre	-,40	-1,09	-0,93	Sin cambio dentro de la disfuncionalidad
3	18/ Hombre	0,00	0,00	-0,83	Sin cambio dentro de la disfuncionalidad
4	54/ Hombre	0,00	0,00	-3,63	Sin cambio dentro de la disfuncionalidad
5	54/ Hombre	0,40	1,09	-2,23	Sin cambio dentro de la disfuncionalidad
6	55/ Hombre	0,00	0,00	-2,63	Sin cambio dentro de la disfuncionalidad
7	47/ Hombre	0,00	0,00	-6,13	Sin cambio dentro de la disfuncionalidad
8	59/ Hombre	-0,33	-0,90	-0,67	Sin cambio dentro de la disfuncionalidad
9	38/ Hombre	-1,40	-3,82	-2,83	Empeoramiento dentro de disfuncionalidad
10	57/ Hombre	-0,83	-2,26	-5,24	Empeoramiento dentro de disfuncionalidad
11	22/ Hombre	0,53	1,44	1,37	Sin cambio dentro de la funcionalidad
12	19/ Hombre	-3,60	-9,81	-0,83	Empeoramiento dentro de disfuncionalidad
13	58/ Hombre	0,00	0,00	-4,13	Sin cambio dentro de la disfuncionalidad
14	61/ Hombre	0,00	0,00	-2,73	Sin cambio dentro de la disfuncionalidad
15	16/ Hombre	-0,90	-2,45	2,57	Empeoramiento dentro de la funcionalidad
16	52/ Hombre	0,00	0,00	-3,53	Sin cambio dentro de la disfuncionalidad
17	42/ Hombre	-0,77	-2,10	-0,77	Empeoramiento dentro de disfuncionalidad
18	16/ Hombre	0,00	0,00	-4,02	Sin cambio dentro de la disfuncionalidad
19	18/ Hombre	-0,04	-0,11	1,33	Sin cambio dentro de la funcionalidad
20	51/ Hombre	0,25	0,68	-5,04	Sin cambio dentro de la disfuncionalidad
21	29/ Hombre	-1,65	-4,50	3,43	Empeoramiento dentro de la funcionalidad
22	44/ Hombre	0,85	2,32	1,37	Recuperado
23	27/ Hombre	-1,35	-3,68	-1,18	Empeoramiento dentro de la disfuncionalidad
24	21/ Hombre	0,00	0,00	-0,23	Sin cambio dentro de la disfuncionalidad
25	51/ Mujer	0,40	1,64	-3,09	Sin cambio dentro de la disfuncionalidad
26	59/ Mujer	0,00	0,00	-3,39	Sin cambio dentro de la disfuncionalidad
27	15/ Mujer	0,50	2,05	0,21	Recuperado
28	59 / Mujer	-0,50	-2,05	-1,69	Empeoramiento dentro de la disfuncionalidad
29	45/ Mujer	0,00	0,00	-1,29	Sin cambio dentro de la disfuncionalidad
30	45/ Mujer	0,33	1,35	0,30	Sin cambio dentro de la funcionalidad
31	41/ Mujer	0,00	0,00	2,97	Sin cambio dentro de la funcionalidad
32	42/ Mujer	0,00	0,00	2,41	Sin cambio dentro de la funcionalidad
33	24/ Mujer	0,00	0,00	-1,39	Sin cambio dentro de la disfuncionalidad
34	63/ Mujer	-1,10	-4,51	-2,39	Empeoramiento dentro de la disfuncionalidad
35	60/ Mujer	0,20	0,82	-2,59	Sin cambio dentro de la disfuncionalidad
36	65/ Mujer	0,00	0,00	-4,09	Sin cambio dentro de la disfuncionalidad
37	62/ Mujer	-0,70	-2,87	-4,05	Empeoramiento dentro de la disfuncionalidad
38	18/ Mujer	0,00	0,00	2,61	Sin cambio dentro de la funcionalidad

\* PCF – Punto de Corte Funcional ( Hombres Peak Flow= 8,63; Mujeres Peak Flow= 5,59)

