Fairness of the 2026 FIFA World Cup group-stage draw

Equidad del sorteo de fase de grupos de la Copa Mundial de la FIFA 2026

Authors

Iñigo García-Atutxa [1]
Ekaitz Dudagoitia Barrio [2]
Hodei Calvo-Soraluze [3]
Leire Altadill-Legarra [4]
Francisca Villanueva-Flores [5]
Igor Garcia-Atutxa [6, 7]

[1] Universidad Alfonso X El Sabio (Spain)
[2] Universidad de Murcia (Spain)
[3] Universidad de Deusto (Spain)
[4] Consejo Superior de Investigaciones Científicas (Spain)
[5] CICATA Morelos (Mexico)
[6] Universidad Católica de Murcia (Spain)
[7] Universidad Tecmilenio (Mexico)

Corresponding author:
Francisca Villanueva-Flores
fvillanuevaf@ipn.mx

Received: 08-01-26
Accepted: 19-02-26

How to cite in APA

García-Atutxa, I., Dudagotia Barrio, E., Calvo-Soraluze, H., Altadill-Legarra, L., Villanueva-Flores, F., & Garcia-Atutxa, I. (2026). Fairness of the 2026 FIFA World Cup group-stage draw. *Retos, 78*, 357-368. https://doi.org/10.47197/retos.v78.118530

Abstract

Introduction and objective: We evaluate the ex-ante fairness of the 2026 FIFA World Cup group-stage draw under official rules and two counterfactual mechanisms using a reproducible simulation. A 48-team pool consistent with confederation quotas is paired with an ex-ante team-strength index derived solely from past World Cup final-tournament matches using Lidstone shrinkage, with robustness checks employing the FIFA Ranking (SUM) and Elo ratings.
Methodology: We compare (i) FIFA-2026: pots by ranking with confederation caps and pre-assigned hosts; (ii) Feasible-uniform: a pot-free feasible baseline under the same confederation caps; and (iii) Fair-greedy: a post-draw, within-pot swap heuristic that accepts only coefficient-of-variation (CV)–reducing moves while preserving all constraints.
Results: Across N = 500 simulated draws per mechanism, the mean CV of group-average strength is 0.0882 (Feasible-uniform), 0.0684 (FIFA-2026), and 0.0634 (Fair-greedy). Thus, relative to FIFA-2026, Feasible-uniform increases inequality by +0.0198 CV (~+28.9%), while Fair-greedy reduces it by −0.0050 (~−7.3%), with lower upper-tail risk under FIFA-2026 and Fair-greedy. Letter-level diagnostics under FIFA-2026 reveal lower difficulty for host-anchored groups and a concentration of "hardest-group" risk in a subset of non-host letters. Against a historical yardstick (1998–2022), FIFA-2026 and Fair-greedy remain within observed bounds, whereas Feasible-uniform can exceed them at the upper extreme.
Conclusions: Overall, our results suggest a simple, constraint-preserving post-draw adjustment that can improve competitive balance without changing the live-draw structure.

Keywords

Tournament design; randomized assignments; fairness; FIFA World Cup 2026; constrained matching.

Resumen

Introducción y Objetivo. Evaluamos la equidad ex ante del sorteo de la fase de grupos del Mundial de la FIFA 2026 bajo las reglas oficiales y dos mecanismos contrafactuales mediante una simulación reproducible. Un conjunto de 48 equipos, consistente con las cuotas por confederación, se empareja con un índice de fortaleza ex ante derivado exclusivamente de partidos de fases finales de Copas del Mundo anteriores, usando contracción de Lidstone, con comprobaciones de robustez empleando el Ranking FIFA (SUM) y las puntuaciones Elo.
Metodología. Comparamos: (i) FIFA-2026: bombos según ranking con topes por confederación y anfitriones preasignados; (ii) Uniforme-factible: una línea base factible sin bombos bajo los mismos topes por confederación; y (iii) Justo-voraz: una heurística posterior al sorteo de intercambios dentro del mismo bombo que acepta únicamente movimientos que reduzcan el coeficiente de variación (CV), preservando todas las restricciones.
Resultados. Con N = 500 sorteos simulados por mecanismo, el CV medio de la fortaleza promedio por grupo es 0.0882 (Uniforme-factible), 0.0684 (FIFA-2026) y 0.0634 (Justo-voraz). Así, en comparación con FIFA-2026, Uniforme-factible incrementa la desigualdad en +0.0198 de CV (~+28.9%), mientras que Justo-voraz la reduce en −0.0050 (~−7.3%), con menor riesgo en la cola superior bajo FIFA-2026 y Justo-voraz. Los diagnósticos por letra bajo FIFA-2026 revelan menor dificultad en los grupos anclados por anfitriones y una concentración del riesgo de "grupo más difícil" en un subconjunto de letras no anfitrionas. Frente a un referente histórico (1998–2022), FIFA-2026 y Justo-voraz se mantienen dentro de los límites observados, mientras que Uniforme-factible puede superarlos en el extremo superior.
Conclusiones. En conjunto, nuestros resultados sugieren un ajuste simple posterior al sorteo, que preserva las restricciones, capaz de mejorar el equilibrio competitivo sin cambiar la estructura del sorteo en vivo.

Palabras clave

Diseño de torneos; asignaciones aleatorias; equidad; Copa Mundial de la FIFA 2026; emparejamiento con restricciones.

Introduction

The 2026 FIFA Men's World Cup will be contested by 48 teams in 12 groups of four, with the top two in each group and the eight best third-placed teams advancing to a 32-team knockout round—104 matches in total. The draw is seeded by the FIFA Ranking and constrained by "geographical and sporting considerations," including pre-assignment of the three hosts to positions A1 (Mexico), B1 (Canada), and D1 (United States). FIFA's regulations also enumerate the pairing logic for third-placed teams in the Round of 32, adding operational complexity to the bracket (FIFA, 2025d).

These changes renew long-standing concerns about draw fairness and between-group balance across groups. Prior work documents that non-uniform assignment probabilities, draw order, and geographical caps can systematically bias outcomes, while operations-research approaches propose more balanced and transparent assignment mechanisms that respect real-world constraints and live-draw requirements (Cea et al., 2020; Guyon, 2015; Monks & Husch, 2009; Roberts & Rosenthal, 2024; Scarf et al., 2009; Wallace & Haigh, 2013). Recent World Cup-specific work formalizes the fairness of group draws and quantifies the effect of seeding and constraints on ex-ante outcomes (Csató, Gyimesi, et al., 2025), while related analyses consider fairness questions upstream of the draw, such as slot allocation across confederations (Csató, Kiss, et al., 2025). In parallel, recent discussions of assisted, transparent procedures for conditioned pairings (e.g., UEFA designs) indicate that operational constraints, live broadcast needs, and more uniform assignment probabilities can be reconciled if draw steps are clearly specified and dead-end risk is controlled (Boczon & Wilson, 2023; Csató, 2023). Since 2018, FIFA has used an Elo-type ranking ("SUM") for seeding (FIFA, 2025c), which updates points based on match importance and opponent strength and provides a comparatively stable input for pre-tournament strength measures; however, the mapping from these rules and constraints to realized group difficulty remains an empirical question (Boczon & Wilson, 2023; Csató, 2023; Karpov, 2016; Laliena & López, 2019, 2025).

This study contributes along three fronts. First, we build a reproducible pipeline that combines the Fjelstul World Cup Database (1930–2022) with a pre-tournament strength index derived solely from final-tournament history to construct ex-ante measures of group difficulty for 2026 (Fjelstul, 2023). Second, we compare three assignment engines: (i) FIFA-2026 (pots, constraints, and pre-assigned hosts), (ii) a feasible-uniform counterfactual that implements a pot-free feasible baseline (stepwise-uniform; not necessarily globally uniform), and (iii) a "Fair-greedy" post-optimization that preserves pots and constraints but locally swaps teams to reduce between-group inequality. Third, we connect group composition to bracket paths (Round-of-32 pairing rules), quantifying how pre-assignment and constraint geometry redistribute the probability of landing in the tournament's "hardest group" (Arlegi & Dimitrov, 2020; Chater et al., 2021; Horen & Riezman, 1985; Krumer & Moreno-Ternero, 2023; Lapré & Palazzolo, 2023; Ryvkin, 2010).

Our primary fairness outcome is the coefficient of variation (CV) of group-average strength S_g across the 12 groups, reported with bootstrap confidence intervals (CIs) and empirical cumulative distribution functions (ECDFs, stochastic dominance). We complement CV with the probability that each lettered group is the hardest in a simulation run and with sensitivity analyses on seeding metrics (FIFA, Elo, and our historical index).

We define group difficulty as the simple average of the four teams' ex-ante strengths; our headline statistic is the CV across these group averages, which captures between-group inequality in average difficulty (i.e., draw fairness). This should not be confused with within-group competitive balance, which concerns how similar teams are within a group (or, downstream, how concentrated the distribution of points is); see Triguero-Ruiz & Avila-Cano (2019) and Avila-Cano et al. (2021). For instance, in a four-team group under a {3,1,0} point system, the most concentrated final points distribution occurs when one team wins all its matches, and the remaining three teams draw among themselves (Avila-Cano et al., 2021). Because pot seeding intentionally stratifies each group (one team per pot), fully homogeneous groups—and thus high within-group competitive balance—are structurally limited under the official design (Csató & Gyimesi, 2025); our focus is therefore on whether this stratified structure yields groups of comparable difficulty.

Practical relevance. For coaches, analysts, and performance staff, draw-induced variation in group difficulty affects scouting priorities, load management, qualifying probabilities, and downstream bracket exposure. Our reporting (e.g., CV percentiles, upper-tail probabilities, and "hardest-group" risks by letter) translates directly into actionable pre-tournament risk assessments under current rules and plausible alternatives.

Method

Data

We use the Fjelstul World Cup Database (v1.2.0) to reconstruct group stages for 1998–2022 and to build a 48-team pool for the 2026 format. Pre-processing harmonizes stable team identifiers, removes historical entities no longer active for 2026 (e.g., Soviet Union), and enforces one-to-one joins on team–edition keys via referential-integrity checks (Fjelstul, 2023).

Pre-tournament team strength

To avoid circularity with contemporaneous seedings, we compute a historical, ex-ante strength index from prior final-tournament matches only. Let n_i be the number of such matches for team i, p_{im} the points earned by i in match m (3/1/0) and μ the global mean points per match across teams/editions used. We apply a Lidstone (add- α) shrinkage estimator (Fienberg & Holland, 1972):

$$s_i = \frac{\alpha\mu \ + \sum_{m=1}^{n_i} p_{im}}{\alpha + n_i}, \alpha = 8,$$

which stabilizes estimates for sparse histories. As robustness checks, we replicate all analyses using the FIFA Ranking (SUM method) and Elo-type ratings as alternative strength metrics. For FIFA, we use the FIFA/Coca-Cola Men's World Ranking (SUM) methodology as documented by FIFA and freeze the snapshot to the ranking issue date used in the official FIFA World Cup 2026 draw procedures (19 November 2025) (FIFA, 2025a). For Elo, we use the World Football Elo Ratings as a football-specific Elo adaptation, archived at the same data freeze date (FIFA, 2025b; Hvattum & Arntzen, 2010). Intuitively, Lidstone shrinkage pulls each team's historical points-per-match estimate toward the overall mean μ when match history is sparse, reducing noise from small samples.

Group strength and between-group inequality

Let $g \in \{A, \dots, L\}$ denote a group and T_g its four teams. Group strength is the simple average of member strengths (Lapré & Amato, 2025):

$$S_g = \frac{1}{4}\sum_{i \in T_g} s_i$$

Our primary fairness metric is the CV across the 12 groups (Triguero-Ruiz & Avila-Cano, 2024):

$$CV(S_g) = \frac{\sigma(S_g)}{\overline{S_g}},$$

This metric has a clear interpretation as relative imbalance between groups. We also compute Gini and Theil indices and monitor the "hardest-group" event (the group attaining $max\ S_g$ in each simulation). In words, we first summarize each group by the average strength of its four teams, then quantify how unequal these group averages are across the 12 groups; lower CV indicates groups are more similar in expected difficulty.

2026 pool construction and pots

The 48-team pool matches official quotas: UEFA 16; CONMEBOL 6; CAF 9; AFC 8; CONCACAF 6; OFC 1; plus, two intercontinental playoff berths. Hosts are pre-assigned to A1 (Mexico), B1 (Canada), and D1

(United States). When berths are undetermined, we complete the pool by confederation and simulate the playoff winners with a Bradley–Terry model (Bradley & Terry, 1952) that converts strength differentials to win probabilities; winners are placed in Pot 4. Pots comprise four sets of 12 teams each, ordered by FIFA Ranking (FIFA/Coca-Cola Men's World Ranking) in the baseline and ordered by s_i and Elo in robustness checks. Geographical constraints are enforced: UEFA ≤ 2 per group; all other confederations ≤ 1 per group.

Draw mechanisms

We compare three engines:

(i) FIFA-2026 (baseline). Pre-positions: A1/B1/D1 fixed. Order: Pot 1 (non-hosts) → Pot 2 → Pot 3 → Pot 4. Constraint handling: At each step we build a bipartite graph between remaining teams (current pot) and groups with available slots; an edge indicates that assigning that team to that group would not violate confederation caps. We then (i) select the next team using a most-constrained-first rule (smallest number of feasible groups), (ii) sample a candidate group among its feasible edges, and (iii) verify future feasibility by computing the size of a maximum bipartite matching between the remaining teams and groups via iterative augmenting paths (Kuhn algorithm (Kuhn, 1955), with Hopcroft–Karp (Hopcroft & Karp, 2006) as an optional acceleration). If the matching size equals the number of remaining teams, the placement is committed; otherwise, we backtrack and resample. This mirrors the broadcastable "pots + constraints" semantics while guaranteeing a feasible completion. Intuitively, the matching check asks whether the remaining teams can still be assigned to groups later without violating confederation caps; if not, we resample the current choice.

(ii) Feasible-uniform (counterfactual baseline; pot-free). We sample without pots from the feasible assignment space defined solely by confederation caps. Exact global uniform sampling over this constrained space is not guaranteed by a simple sequential procedure; our implementation is stepwise-uniform (uniform among currently feasible choices) with randomized backtracking and look-ahead feasibility checks. Accordingly, we interpret Feasible-uniform as a pot-free baseline rather than as an exact globally uniform draw; Section "Robustness and sensitivity" reports robustness checks to sampler design.

(iii) Fair-greedy (post-optimization). Starting from a feasible FIFA-style draw, we iteratively test within-pot 2-swaps between the groups with the largest and smallest group strength S_g. We accept only swaps that strictly reduce the between-group inequality metric CV and preserve all constraints (pots, hosts, geography). Iteration stops at a local minimum (no improving swap).

Statistical analysis

For each mechanism we run N = 500 independent draws (N is a user-set simulation budget; increasing N reduces Monte Carlo error and improves tail resolution without changing the draw logic), storing S_g and CV (S_g). We report means, standard deviations, quartiles, and ranges, and present percent effects relative to FIFA-2026. Because N = 500 is coarse for extreme-tail quantities (e.g., sample maxima or very small probabilities), we interpret tail statements qualitatively and report uncertainty via CIs rather than over-interpreting single extreme draws.

We estimate mechanism-specific means and pairwise differences for CV (S_g). CIs comes from a nonparametric bootstrap (Efron & Tibshirani, 1994) over the N simulated draws (B = 10,000 resamples); normal-approximation CIs serve as checks. These intervals quantify Monte Carlo uncertainty arising from the finite simulation size N. We compare full distributions via empirical cumulative distribution functions (ECDFs) and report one-sided Kolmogorov–Smirnov (KS) (Massey Jr., 1951) statistics when relevant (first-order stochastic dominance). For each group letter we compute the probability of being the "hardest group" with Wilson (Wilson, 1927) 95% binomial CIs (CIs omitted for brevity), marking 8.33% as the uniform baseline across 12 groups. For readers unfamiliar with these tools: bootstrap resampling forms CIs by repeatedly resampling the N simulated draws; ECDF/KS comparisons indicate whether one mechanism tends to yield lower CV values across the whole distribution, not just on average.

Robustness and sensitivity

We assess robustness only along the dimensions supported by the codebase:

(i) Strength metric and smoothing. We compare the historical, ex-ante strength index S_i (Section "Pre-tournament team strength"; baseline α = 8) against alternative strength metrics used in practice for seeding, namely the FIFA Ranking (SUM) and football-specific Elo ratings. This checks that our conclusions do not hinge on a particular strength construction.

(ii) Feasible-uniform sampler. The counterfactual pot-free is implemented via the stepwise-uniform randomized backtracking sampler with look-ahead feasibility checks described in Section "Draw mechanisms". Because stepwise uniformity does not guarantee global uniformity over all feasible assignments, we do not interpret this mechanism as an exact globally uniform sampler; instead, it serves as a transparent pot-free baseline.

(iii) Fair-greedy heuristic. Starting from a feasible FIFA-style draw, the algorithm iteratively proposes within-pot 2-swaps between the groups with the largest and smallest group strength S_g. A swap is accepted only if it strictly reduces the inequality metric CV and preserves all constraints (pots, host pre-positions A1/B1/D1, and confederation caps). The search proceeds from a single deterministic start and stops at the first local optimum (no neutral-move acceptance).

(iv) Pool composition. We perturb the 48-team pool by varying intercontinental playoff outcomes and the set of currently active teams, while keeping host pre-positions and confederation caps unchanged. For each variant, we rerun the three mechanisms and recompute the summary statistics.

Results

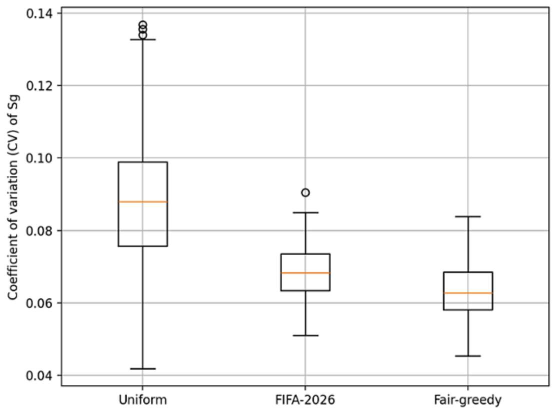
Primary analysis (2026): between-group inequality across mechanisms

Ex-ante fairness is evaluated using the coefficient of variation of group-average strength, CV (S_g), over N = 500 simulated draws per mechanism. We treat N as the Monte Carlo budget and quantify sampling uncertainty for key mean comparisons via nonparametric bootstrap CIs (Table 1). A clear and consistent hierarchy emerges, Feasible-uniform > FIFA-2026 > Fair-greedy, where higher values indicate worse between-group balance. The mean CV (S_g) is 0.0882 under Feasible-uniform, 0.0684 under FIFA-2026, and 0.0634 under Fair-greedy (Fig. 1). In absolute terms, Feasible-uniform exceeds FIFA-2026 by +0.0198 (≈+28.94% relative to FIFA-2026), whereas Fair-greedy reduces mean CV by −0.0050 (≈−7.29%). The gap Feasible-uniform vs. Fair-greedy is +0.0248 (≈+39.07% relative to Fair-greedy).

Table 1. Percent effects on mean CV vs FIFA-2026, with bootstrap (B=10,000) estimates and 95% CIs.

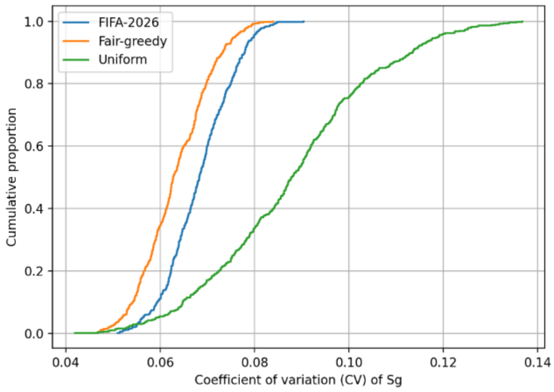
Comparison	95% CI	Δ% (mean CV)	ΔCV (%)
Uniform - FIFA-20261	[0.0180, 0.0215]	28.94	0.0198
Fair-greedy - FIFA-2026	[-0.0054, -0.0046]	-7.29	-0.0050
Uniform - Fair-greedy	[0.0230, 0.0265]	39.07	0.0248

Figure 1. CV (S_g) of group-average strength across mechanisms (N=500): Uniform > FIFA-2026 > Fair-greedy.



Distributional diagnostics corroborate these mean differences. Feasible-uniform displays the widest spread and heaviest upper tail, consistent with a materially higher risk of extreme imbalance: its standard deviation is 0.0179 and its interquartile range (IQR) is 0.0233. By contrast, FIFA-2026 concentrates outcomes in a narrower band (σ = 0.0070; IQR = 0.0102), and Fair-greedy yields a similarly narrow dispersion (σ = 0.0073; IQR = 0.0104) while shifting the distribution toward lower CV values. Relative to Feasible-uniform, the IQR contracts by ~56.3% for FIFA-2026 and ~55.4% for Fair-greedy; similarly, the σ contracts by ~60.9% (FIFA-2026) and ~59.2% (Fair-greedy). A scale-free dispersion indicator, IQR/median, confirms a large reduction in relative spread versus Feasible-uniform: 0.265 (Feasible-uniform) vs 0.149 (FIFA-2026) vs 0.165 (Fair-greedy), with FIFA-2026 exhibiting the tightest spread and Fair-greedy a slightly wider but left-shifted distribution (Fig. 2).

Figure 2. ECDFs of CV for Feasible-uniform, FIFA-2026, and Fair-greedy (2026); curves further left/upward indicate greater fairness.



Empirical cumulative distributions show first-order stochastic dominance of Fair-greedy over FIFA-2026, and of both over Feasible-uniform: for nearly any operational threshold on CV (S_g), the probability of achieving a more balanced draw is higher under Fair-greedy, followed by FIFA-2026. Extremes align with this picture: the maximum CV (S_g), under Feasible-uniform reaches 0.1368, exceeding the historical ex-ante peak observed in past editions, whereas maxima under FIFA-2026 (0.0904) and Fair-greedy (0.0839) remain within historically observed ranges. Minimum values (0.0511 for FIFA-2026; 0.0454 for Fair-greedy; 0.0419 for Feasible-uniform) confirm that all mechanisms can occasionally generate unusually balanced draws, but Feasible-uniform does so at the cost of far greater variability and upper-tail risk. Increasing N in the same pipeline would sharpen extreme-tail estimates without changing the draw logic.

From an operational standpoint, these results indicate that the pot-plus-constraints logic of FIFA-2026 already reduces systemic inequality versus a purely uniform assignment on the feasible space, and that a simple within-pot swap heuristic (Fair-greedy) yields additional, distribution-wide gains without violating live-draw constraints. In practice, this translates into fewer very high-difficulty groups (high group-average strength) scenarios and a narrower uncertainty band for coaches and analysts when planning scouting and load management.

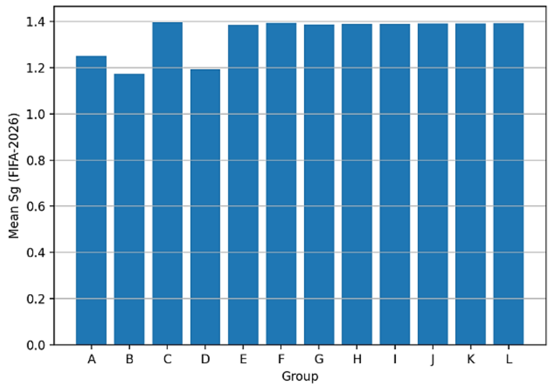
While our primary endpoint is the dispersion of group strengths, the same simulated draws can be coupled with a match-outcome model to estimate expected group-stage points and probabilities of qualification/advancement. For example, the Bradley–Terry win-probability mapping already used to simulate the intercontinental playoff provides a natural baseline for such extensions. Because these indicators depend on additional modelling choices beyond the draw mechanism itself, we treat them as complementary to the draw-fairness outcomes reported here.

Beyond dispersion in average group difficulty, each four-team group can be viewed as an independent competition whose within-group competitive balance may be measured with alternative indices (e.g., the normalized Herfindahl–Hirschman Index or the Distance to Competitive Balance measure) that operate in relative terms and can incorporate weighting schemes. We do not replicate the analysis with these indices here but note that magnitudes (and potentially rankings) may be sensitive to the chosen balance metric.

Letter-level difficulty under FIFA-2026

We analyze letter-level group composition using the group-average strength S_g across N = 500 simulated draws under the official rules with host pre-positions A1 (Mexico), B1 (Canada), and D1 (United States). The evidence indicates a systematic left shift for A, B, and D relative to the tournament-wide distribution of S_g, whereas several non-host letters, especially F, J, K, and H, exhibit higher central tendency. The difference between the lowest- and highest-mean letters is approximately 0.15–0.25 points (about 10–15% of the grand mean), which is operationally meaningful at group-stage margins (Fig. 3).

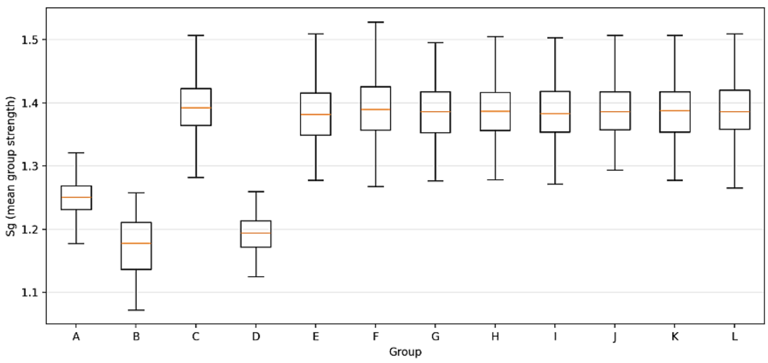
Figure 3. Group-average strength (S_g) by group letter (FIFA-2026). Bars show mean S_g for groups A–L; A/B/D (hosts) highlighted.



Distributional profiles complement these mean contrasts. Host-anchored letters (A/B/D) display compressed dispersion around the lower half of the overall S_g distribution and rarely visit the upper tail. In contrast, F/J/K/H show right-skewed shapes with heavier upper tails, consistent with a reallocation of severe team combinations toward a restricted subset of non-host letters. Robust spread measures (IQR) and tail range corroborate this asymmetry: non-host letters that concentrate difficulty present both higher medians and larger upper-quartile excursions, while A/B/D are characterized by tighter, lower-centered boxes and short upper whiskers (Fig. 4).

Two mechanisms plausibly underpin these patterns. First, because hosts are pre-assigned to Pot-1 slots, they replace a non-host Pot-1 team in groups A/B/D; to the extent that a host's strength is below the Pot-1 average, the A/B/D anchors depress in those letters. Second, confederation caps interact with the fixed host placements to constrain feasible assignments downstream, thereby concentrating difficult cross-confederation pairings in a smaller set of non-host letters. From a planning perspective, this implies (i) reduced exposure to high- difficulty profiles for A/B/D, and (ii) elevated monitoring for letters such as F/J/K/H, where heavier tails increase the likelihood of very high-difficulty groups (high group-average strength) even when the tournament-wide mean group-average strength is unchanged.

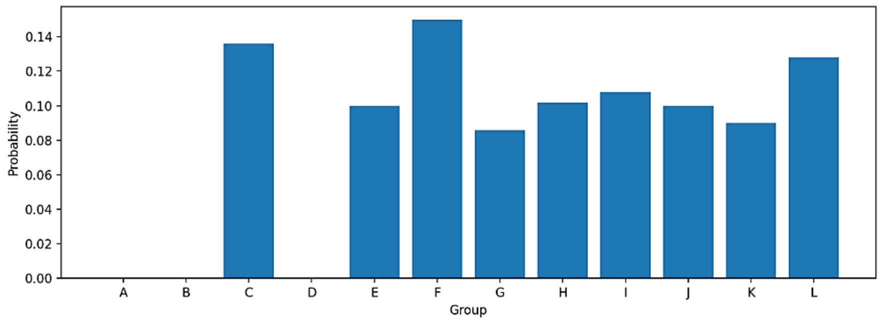
Figure 4. Boxplots S_g by group letter (FIFA-2026). Distribution of S_g for groups A–L across simulations; A/B/D annotated as host groups.



"Hardest-group" risk by letter

We quantify extreme difficulty by the probability that each letter attains the maximum group-average strength within a simulation run, $Pr\{letter = arg\ max\ S_g\}$, over N = 500 draws under the FIFA-2026 mechanism. This "hardest-group" definition is based on average strength; it differs from within-group competitive balance, and from the popular "group of death" label, which can also apply to evenly matched groups even at lower overall strength levels. With twelve groups, the symmetrical benchmark is $\frac{1}{2}$ = 8.33%. The observed probabilities are: A = 0.0%, B = 0.0%, C = 13.6%, D = 0.0%, E = 10.0%, F = 15.0%, G = 8.6%, H = 10.2%, I = 10.8%, J = 10.0%, K = 9.0%, L = 12.8% (Fig. 5).

Figure 5. Probability of being the "hardest group" for groups A–L (bars).



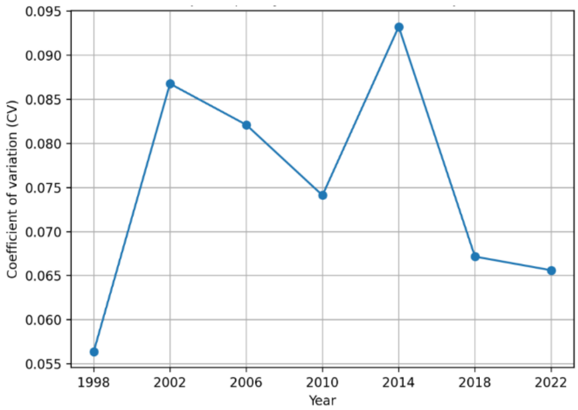
Three patterns stand out. First, the host-anchored letters (A/B/D) are effectively insulated from being the hardest group (0% across 500 runs), aligning with their lower S_g profiles in section "Letter-level difficulty under FIFA-2026". Second, a small set of non-host letters bears disproportionate risk: F = 15.0%, C = 13.6%, and L = 12.8%, which correspond to risk ratios of about 1.80×, 1.63×, and 1.54× relative to the 8.33% benchmark. Third, several letters hover near symmetry (e.g., G ≈ 8.6%, K ≈ 9.0%), indicating that the excess mass observed for F/C/L is not diffused uniformly but rather concentrated.

Mechanistically, these imbalances are consistent with the interaction between fixed host placements (A1/B1/D1) and confederation caps: by anchoring three letters with host teams, the feasible assignment space reconfigures so that severe cross-confederation combinations are more frequently realized in certain non-host letters. From a coaching and performance-planning perspective, this implies letter-specific exposure: while A/B/D carry negligible risk of becoming the tournament's hardest group, letters such as F/C/L warrant heightened pre-tournament contingency planning (scouting breadth, rotation scenarios, and load management), even when average tournament difficulty remains stable.

Historical yardstick (1998–2022)

To contextualize 2026 outcomes, we compute ex-ante CV (S_g) for past World Cups using the same historical strength index and procedure. The series spans ≈ 0.056–0.093 from 1998 to 2022, with a peak near 0.093 in 2014 and comparatively balanced editions in 2018–2022 (≈ 0.067–0.066). Relative to this yardstick, the Feasible-uniform mechanism occasionally yields CV (S_g) values above the historical maximum, indicating tail-risk profiles that have not been typical of realized tournaments. In contrast, both FIFA-2026 and Fair-Greedy remain within historical limits; moreover, the mass of the Fair-Greedy distribution concentrates below the 2018–2022 band, consistent with a systematic reduction in between-group inequality (Fig. 6).

Figure 6. Historical between-group inequality CV (S_g) at FIFA World Cups, 1998–2022, computed ex-ante from pre-tournament strength.



Historical yardstick (1998–2022)

Robust analyses are restricted to variants supported by the current codebase. Replacing the historical index s_i with FIFA (SUM) as the seeding-relevant strength metric does not alter the qualitative ordering Feasible-uniform > FIFA-2026 > Fair-greedy; point estimates shift marginally but remain within bootstrap uncertainty. The Feasible-uniform sampler based on randomized backtracking with look-ahead reproduces the same dispersion patterns across repeated runs, confirming that the high-variance, heavy-tail profile is intrinsic to this pot-free stepwise-uniform baseline under the constrained feasible space. The Fair-greedy heuristic, initialized from the FIFA-style draw produced in each simulation run; the improvement step is deterministic conditional on the draw and accepting only strict improvements in CV (S_g), consistently converges to lower and tighter CV (S_g) distributions, indicating that local within-pot swaps suffice to capture meaningful fairness gains without relaxing life-draw constraints. Finally, pool composition perturbations (intercontinental playoff outcomes and active-team lists) leave the mechanism ranking intact and preserve the letter-level asymmetry documented above (host letters A/B/D systematically benign; elevated exposure for a subset of non-host letters).

Together, these implemented checks support the stability of the main findings: the pot-and-constraints architecture already reduces structural inequality relative to Feasible-uniform assignment, and a lightweight, constraint-preserving swap routine produces additional, distribution-wide improvements in competitive balance.

Discussion

Across 500 draws per mechanism, our simulations under the assumptions described in Methods, yield a clear hierarchy in between-group inequality (CV of group-average strength): Feasible-uniform (0.0882) > FIFA-2026 (0.0684) > Fair-greedy (0.0634). Relative to FIFA-2026, the uniform counterfactual is +0.0198 CV (+28.94%), while Fair-greedy reduces CV by −0.0050 (−7.29%). Distributional diagnostics point to substantially narrower spreads and lower upper-tail risk under FIFA-2026 and Fair-greedy (e.g., max CV 0.1368 for Uniform vs 0.0904 for FIFA-2026 and 0.0839 for Fair-greedy), suggesting that a pots-and-constraints engine already mitigates extreme imbalance, with modest within-pot swaps delivering additional tournament-wide gains without violating live-draw constraints. These comparisons are ex-ante and conditional on the strength proxies and constraints described in Methods and should not be interpreted as realized match outcomes.

Letter-level patterns under FIFA-2026 reveal structural asymmetries shaped by host pre-assignments (A1/B1/D1) and confederation caps: host-anchored groups A, B, and D tend to be weaker and less dispersed, while several non-host letters (notably F, J, K, and H) skew stronger. The "hardest-group" event concentrates in F (15.0%), C (13.6%), and L (12.8%), whereas A/B/D are effectively insulated (0%), relative to the symmetric 8.33% benchmark across 12 groups. These results imply letter-specific exposure that matters for pre-tournament planning. From a sports-management standpoint, such letter-specific exposure can be used for scenario planning (logistics, scouting, and resource allocation) and supports communicating difficulty distributions alongside the draw rules.

Against a historical yardstick (1998–2022 CV ≈ 0.056–0.093), the uniform counterfactual for 2026 occasionally exceeds the historical maximum, whereas FIFA-2026 and Fair-greedy remain within historical limits; moreover, Fair-greedy's mass lies below the 2018–2022 band (≈ 0.067–0.066), consistent with systematically improved balance. We treat this historical range as descriptive context rather than a normative target, since past formats and constraints differ from 2026.

Scope conditions bound inference: the strength index is ex-ante and derived only from final-tournament history, different strength models could shift magnitudes, N = 500 limits precision in the far tail; Fair-greedy halts at a local optimum; and we do not map group strength to match outcomes. CV summarises between-group dispersion of average strength and does not capture within-group heterogeneity or downstream qualification probabilities. Feasible-uniform is a constrained baseline and does not guarantee global uniformity over feasible assignments. Nonetheless, robustness checks (alternative strength metrics, uniform sampler repetition, and pool-composition perturbations) preserve the mechanism ranking and the letter-level asymmetry.

Two low-cost prescriptions follow: (i) retain a pots-plus-constraints engine with feasibility checks, and (ii) consider a transparent, within-pot swap phase that accepts only strict CV improvements. For sports administrators, publishing the feasibility checks and swap criterion (and, where possible, a reference implementation) would improve auditability and stakeholder trust without changing the life-draw experience. Operationally, teams drawn into letters with elevated "hardest-group" risk—especially F/C/L— may wish to prepare accordingly.

Conclusions

Using fully reproducible simulations (500 runs per mechanism), we show that the FIFA-2026 pots-and-constraints architecture substantially reduces between-group inequality relative to a feasible uniform assignment, and that a minimal, rule-compatible post-draw stage of within-pot swaps ("Fair-greedy") yields additional improvements (see Table 1 and Fig. 1–2). Compared with historical World Cups (1998–2022), both FIFA-2026 and Fair-greedy fall within observed ranges of group-difficulty dispersion, while the uniform counterfactual can push into the upper tail. Beyond averages, letter-level analyses reveal systematic, non-uniform exposure: host anchoring and confederation caps shield some groups from the "hardest-group" event and concentrate risk in a subset of non-host letters, a pattern robust to alternative strength proxies. Overall, the results provide an operational benchmark and a transparent, low-cost refinement compatible with a live broadcast draw, alongside a replicable framework for evaluating and improving draw fairness in future tournaments.

Acknowledgements

During the preparation of this manuscript, the authors used ChatGPT-5 exclusively to enhance language and readability. After employing this tool, the authors reviewed and edited the content as necessary and assumed full responsibility for the content of the publication.

Financing

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by grants from the Instituto Politécnico Nacional through the "Convocatoria de Apoyos Económicos para Publicaciones y Trámite de Invenciones."

References

Arlegi, R., & Dimitrov, D. (2020). Fair elimination-type competitions. European Journal of Operational Research, 287(2), 528–535. https://doi.org/10.1016/j.ejor.2020.03.025

Avila-Cano, A., Ruiz-Sepulveda, A., & Triguero-Ruiz, F. (2021). Identifying the Maximum Concentration of Results in Bilateral Sports Competitions. Mathematics, 9(11). https://doi.org/10.3390/math9111293

Boczon, M., & Wilson, A. J. (2023). Goals, Constraints, and Transparently Fair Assignments: A Field Study of Randomization Design in the UEFA Champions League. Management Science, 69(6), 3474–3491. https://doi.org/10.1287/mnsc.2022.4528

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. Biometrika, 39(3–4), 324–345. https://doi.org/10.1093/biomet/39.3-4.324

Cea, S., Durán, G., Guajardo, M., Sauré Valenzuela, D., Siebert Sandoval, J., & Zamorano, G. (2020). An analytics approach to the FIFA ranking procedure and the World Cup final draw. Annals of Operations Research. https://doi.org/10.1007/s10479-019-03261-8

Chater, M., Arrondel, L., Gayant, J.-P., & Laslier, J.-F. (2021). Fixing match-fixing: Optimal schedules to promote competitiveness. European Journal of Operational Research, 294(2), 673–683. https://doi.org/10.1016/j.ejor.2021.02.006

Csató, L. (2023). A note on the UEFA Champions League Round of 16 draw (arXiv:2210.15555). arXiv. https://doi.org/10.48550/arXiv.2210.15555

Csató, L., & Gyimesi, A. (2025). Increasing competitiveness by imbalanced groups: The example of the 48-team FIFA World Cup. European Journal of Operational Research, S037722172500935X. https://doi.org/10.1016/j.ejor.2025.11.025

Csató, L., Gyimesi, A., Goossens, D., Devriesere, K., Lambers, R., & Spieksma, F. (2025). How to measure the uncertainty of a tournament draw: The case of European football's Champions League (arXiv:2507.15320). arXiv. https://doi.org/10.48550/arXiv.2507.15320

Csató, L., Kiss, L. M., & Szádoczki, Z. (2025). The allocation of FIFA World Cup slots based on the ranking of confederations. Annals of Operations Research, 344(1), 153–173. https://doi.org/10.1007/s10479-024-06091-5

Efron, B., & Tibshirani, R. J. (1994). An Introduction to the Bootstrap. Chapman and Hall/CRC. https://doi.org/10.1201/9780429246593

Fienberg, S. E., & Holland, P. W. (1972). On the choice of flattening constants for estimating multinomial probabilities. Journal of Multivariate Analysis, 2(1), 127–134. https://doi.org/10.1016/0047-259X(72)90014-0

FIFA. (2025a). Draw Procedures for the FIFA World Cup-2026. https://digitalhub.fifa.com/m/2d1a1ac7bab78995/original/Draw-Procedures-for-the-FIFA-World-Cup-2026.pdf

FIFA. (2025b). FIFA World Ranking Men [Internet]. https://inside.fifa.com/es/fifa-world-ranking/men

FIFA. (2025c). FIFA/Coca-Cola Men's World Ranking Procedures [Internet]. https://inside.fifa.com/fifa-world-ranking/procedure-men

FIFA. (2025d). Regulations for the FIFA World Cup 26. https://digitalhub.fifa.com/m/636f5c9c6f29771f/original/FWC2026_regulations_EN.pdf

Fjelstul, J. C. (2023). The Fjelstul World Cup Database v.1.2.0 [Dataset].

Guyon, J. (2015). Rethinking the FIFA World CupTM final draw. Journal of Quantitative Analysis in Sports, 11(3), 169–182. https://doi.org/10.1515/jqas-2014-0030

Hopcroft, J. E., & Karp, R. M. (2006). An $n^{5/2}$ Algorithm for Maximum Matchings in Bipartite Graphs. SIAM Journal on Computing. https://doi.org/10.1137/0202019

Horen, J., & Riezman, R. (1985). Comparing Draws for Single Elimination Tournaments. Operations Research, 33(2), 249–262.

Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. International Journal of Forecasting, Sports Forecasting, 26(3), 460–470. https://doi.org/10.1016/j.ijforecast.2009.10.002

Karpov, A. (2016). A new knockout tournament seeding method and its axiomatic justification. Operations Research Letters, 44(6), 706–711. https://doi.org/10.1016/j.orl.2016.09.003

Krumer, A., & Moreno-Ternero, J. D. (2023). The Allocation of Additional Slots for the FIFA World Cup. Journal of Sports Economics, 24(7), 831–850. https://doi.org/10.1177/15270025231160757

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 2(1–2), 83–97. https://doi.org/10.1002/nav.3800020109
Laliena, P., & López, F. J. (2019). Fair draws for group rounds in sport tournaments. International Transactions in Operational Research, 26(2), 439–457. https://doi.org/10.1111/itor.12565
Laliena, P., & López, F. J. (2025). Draw procedures for balanced 3-team group rounds in sports competitions. Annals of Operations Research, 346(3), 2065–2092. https://doi.org/10.1007/s10479-025-06497-9
Lapré, M. A., & Amato, J. G. (2025). The impact of imbalanced groups in UEFA Euro 1980–2024 and comparison with the FIFA World Cup. Journal of Quantitative Analysis in Sports. https://doi.org/10.1515/jqas-2024-0151
Lapré, M. A., & Palazzolo, E. M. (2023). The evolution of seeding systems and the impact of imbalanced groups in FIFA Men's World Cup tournaments 1954–2022. Journal of Quantitative Analysis in Sports, 19(4), 317–332. https://doi.org/10.1515/jqas-2022-0087
Massey Jr., F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. Journal of the American Statistical Association, 46(253), 68–78. https://doi.org/10.1080/01621459.1951.10500769
Monks, J., & Husch, J. (2009). The Impact of Seeding, Home Continent, and Hosting on FIFA World Cup Results. Journal of Sports Economics, 10(4), 391–408.
Roberts, G. O., & Rosenthal, J. S. (2024). Football group draw probabilities and corrections. Canadian Journal of Statistics, 52(3), 659–677. https://doi.org/10.1002/cjs.11798
Ryvkin, D. (2010). The selection efficiency of tournaments. European Journal of Operational Research, 206(3), 667–675.
Scarf, P., Yusof, M. M., & Bilbao, M. (2009). A numerical study of designs for sporting contests. European Journal of Operational Research, 198(1), 190–198.
Triguero-Ruiz, F., & Avila-Cano, A. (2024). The competitive balance of UEFA and CONMEBOL football leagues: On managing the number of teams in league design. Managerial and Decision Economics, 45(3), 1480–1496. https://doi.org/10.1002/mde.4083
Wallace, M., & Haigh, J. (2013). Football and Marriage – And the UEFA Draw. Significance, 10(2), 47–48. https://doi.org/10.1111/j.1740-9713.2013.00656.x
Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association, 22(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

Authors and translators' details:

Iñigo García-Atutxa	inigogaratu@gmail.com	Author
Ekaitz Dudagotia Barrio	ekaitz10ED10@gmail.com	Author
Hodei Calvo-Soraluze	hodei.calvo@deusto.es	Author
Leire Altadill-Legarra	leire.altadill@inia.csic.es	Author
Francisca Villanueva-Flores	fvillanuevaf@ipn.mx	Author
Igor Garcia-Atutxa	igarcia839@alu.ucam.edu	Author