Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Three Weapons, Three Regimes of Order: Per-Touch Skill Transmission and the Structural Predictability of Fencing Tournaments

Naci Kalkan[1]
Manisa Celal Bayar University
Emre Gürbüz[2]
emregurbuz@topkapi.edu.tr

Istanbul Topkapi University
Metehan Özsoy[3]
Manisa Celal Bayar University
Said Enes Yılmaz[1]
Manisa Celal Bayar University

Research Article

Keywords: intransitivity, paired-comparison models, Bayesian hierarchical models, tournament seeding, right-of-way scoring, rank concordance

Posted Date: August 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10558072/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Competitive sport sits between skill and chance, and where a discipline falls depends both on the rules that decide a single contest and on the structure that turns many contests into a ranking. Olympic fencing lets us separate the two: three weapons (épée, foil, sabre) run under one tournament format but score touches differently, so weapon varies the rules of play while competition design stays fixed.

Methods

We assembled 158,652 pool bouts across 474 FIE Junior World Cup competitions (2015–2026) and measured predictability through three lenses: the rate of intransitive (cyclic) triads within pools; a custom race-to-five per-touch likelihood, fitted by maximum likelihood and as a Bayesian hierarchical model; and a tournament-level concordance built on an artifact-free earned-tier outcome, with the pool-fitted per-touch slope carried into the elimination tableau as a structural out-of-sample test.

Results

Disorder ranked the weapons épée > sabre > foil (14.5%, 12.4%, and 10.3% cyclic triads, respectively); per-touch skill transmission ran the reverse (marginal slope épée 0.48, sabre 0.80, foil 0.83), with weapon-specific between-tournament heterogeneity largest in foil. Earned-tier concordance (foil 0.684, sabre 0.650, épée 0.638) fell well below artifact-inflated raw-rank values (0.81–0.85).

Conclusions

We expose and correct a tie-break artifact that inflates conventional fencing predictability and show that the épée prediction gap reflects where the independent-touch reference stops rather than a weapon-specific mechanism: the same departure appears in the priority weapons, where simultaneous touches cannot occur.

1 Background

Every sport answers one question implicitly: how often does the better competitor win? The answer never belongs to the athletes alone. It falls out of the interaction between the rules that govern a single contest and the structure that aggregates many contests into a final standing. The resulting balance between skill and chance matters both to the science of performance and to the practical work of designing competition that is fair and worth watching [1, 2]. A format that is kind to underdogs, a scoring rule that rewards one decisive action, a seeding procedure that throws the strongest entrants together early: each bends the path from latent ability to final placement.

The practical stakes are concrete. If foil's between-tournament heterogeneity reflects refereeing variance in right-of-way calls, that is an actionable finding for referee training and certification, not just a statistical curiosity. If épée's lower per-touch transmission means individual bouts carry less information about relative skill, that has direct implications for how much weight junior selection committees should place on a single result versus a season's aggregate. And if official rankings are inflated by a tie-break artifact, federations relying on those rankings for seeding future events are seeding on partly circular information. Each of the paper's three pillars therefore speaks to a decision someone in the sport makes.

Fencing lets us study that path at scale and under close to experimental control. International fencing runs in three weapons, épée, foil, and sabre, that share one competitive scaffold. Athletes pass through a round-robin pool phase, which seeds a single-elimination tableau, under one governing body, one calendar, and one set of refereeing and ranking conventions. What separates the weapons is how a touch is resolved. Épée counts simultaneous (double) touches, awarding a point to both fencers at once. Foil and sabre resolve simultaneous actions by priority, or right-of-way, giving the touch to a single fencer [3]. Because the architecture stays fixed while the within-bout rule set changes, weapon works as a natural manipulation of the rules of play. It isolates their effect on the skill-chance balance from the effect of competition design.

The literature on how tournament structure shapes predictability is large. Knockout formats are only imperfectly monotone: a stronger entrant need not advance further than a weaker one, and seeding choices move the probability that the best competitor prevails [4–10]. A separate strand on intransitivity shows that head-to-head results routinely contain rock-paper-scissors cycles that no single ranking can reconcile, and that the expected level of such cyclicity depends on how competitor abilities translate into win probabilities [11]. Both strands speak to the same skill-chance question, yet they have rarely met on one large, rule-varying dataset.

These design questions have a data-driven counterpart in how noisy a sport's per-touch outcomes are. A format choice, whether that is pool size, the number of preliminary rounds, or how deep seeding protection extends into the bracket, trades fairness against length, and the right trade-off depends on how much signal a single bout carries about the underlying skill gap [4, 10]. A weapon whose bouts transmit skill weakly needs more rounds, or larger pools, to reach the same seed-to-outcome reliability as a weapon whose bouts transmit skill strongly. FIE currently applies the same format uniformly across all three weapons, so it is worth asking whether that format is equally informative in each. Our weapon-specific estimates give this question a number instead of a guess.

How much a sport rewards skill over chance has lately become a quantitative question, and the tools for answering it have advanced quickly. [12] generalize the paired-comparison framework to estimate both the depth of a competitive hierarchy and the residual probability of an upset and find that most sports and games are "shallow," with little room for upsets, a continuum on which any discipline can be located. Agent-based calibrations put numbers to the same idea inside fencing itself: in junior épée the share of

outcome attributable to talent sits at or just below one-half, an unusually high chance component for an individual sport [13], and in elite tennis the talent weight is lower still [14]. Variance-decomposition studies reach compatible conclusions in team sport, attributing roughly forty percent of seasonal over-performance to skill and the remainder to luck [15], while format comparisons show that shorter, more sequential contests mechanically enlarge the luck share [16]. The intransitivity strand has matured in parallel: cyclic, rock-paper-scissors dominance is now documented as a stable and even bet-beating feature of effortful professional sport [17, 18] and embedded in Bayesian generative models that separate transitive strength from genuine cyclicity [19]. Paired-comparison and Elo-type models have likewise been extended to strength-dependent ties, margins of victory, and hierarchical random effects [20–23]. Fencing has begun to receive this treatment directly. The three weapons differ sharply in tempo, with work-to-rest ratios running from roughly 1:1 in épée to nearly 1:9 in sabre [24], and network analyses of fencing results confirm that single-elimination structure raises the incidence of cyclic outcomes relative to round-robin play [25]. What the field still lacks is a single corpus large enough to estimate these effects per weapon, model them at the level of the individual touch, and confront the predictability they imply with an outcome measure free of seeding artifacts.

Fencing's large-scale structural study is, in fact, still thin, and the existing work carries three constraints. Samples have usually been too small to estimate weapon-specific effects with precision, or to characterize the heterogeneity across tournaments. Analyses have treated the bout as the atomic unit, which leaves the per-touch mechanism, the level at which the scoring rules operate, unmodelled, even though within-match fencing performance is serially structured rather than a string of independent trials [26]. And tournament-level predictability has been judged against published final rankings, which we show are contaminated by a tie-breaking artifact that mechanically inflates the apparent agreement between seeding and outcome.

This paper takes on all three using a corpus of 158,652 pool bouts from 474 Junior World Cup competitions contested between 2015 and 2026. We analyze predictability through three complementary pillars. We first quantify model-free disorder as the rate of intransitive triads within pools. We then specify a per-touch likelihood, a symmetric race-to-five process in the tradition of paired-comparison models [27, 28], and estimate weapon-specific skill transmission by maximum likelihood and within a Bayesian hierarchical framework that carries tournament-level uncertainty [29, 30]. We finally introduce an earned-tier outcome that removes the ranking artifact, then test whether the pool-fitted per-touch mechanism predicts elimination-phase orderliness out of sample.

Three objectives organize the work. The first is to describe the skill-chance balance of each weapon at three nested levels of analysis. The second, and the study's principal methodological contribution, is to expose and correct a tie-break artifact that silently inflates conventional fencing predictability measures, turning a correlation of seeding with itself into the appearance of a meritocratic ranking. The third is to test how far a deliberately simple per-touch model carries from the pool phase to the elimination tableau, with explicit attention to where it stops and why.

2 Methods

2.1 Data source and corpus construction

A complete replication package, including the preprocessed data, all analysis scripts (Python and R), and supplementary figures, has been deposited at https://osf.io/jtb8h.

We compiled competition results from the International Fencing Federation (FIE) for all junior-category World Cup events. Pool sheets, seedings, and final classifications were retrieved from the FIE results portal and parsed into structured tables. Events before 2015 were dropped because their pool and tableau files are empty placeholders in the source archive. The analytic corpus then comprises 158,652 pool bouts across 474 competitions held between 2015 and 2026 in the six junior individual disciplines (men's and women's épée, foil, and sabre). Each FIE competition record corresponds to a single weapon-and-gender event, so the 474 competitions partition exactly into épée (173), foil (155), and sabre (146). Every pool bout records the two competitors, their entry seeds, the final score (touches for and against), and the winner; every competition records the final classification and the elimination round each fencer reached. We retained the final score of each bout but not its touch-by-touch sequence, a constraint that bears directly on the per-touch diagnostics below. Table 1 summarizes the corpus. Throughout, seed denotes a fencer's entry ranking into the event, and a bout's seed difference is the signed gap between competitors' seeds, standardized within the modelling sample (z-score, written x_z) so that per-touch slopes compare cleanly across weapons. Every analysis was implemented independently in Python and R and reconciled to numerical precision, an internal guard against coding error.

Table 1
Corpus by weapon.

Weapon	Competitions	Pool bouts	Cyclic-triad rate	Seasons
Foil	155	52,201	10.3%	2015–2026
Sabre	146	45,634	12.4%	2015–2026
Épée	173	60,817	14.5%	2015–2026
All weapons	474	158,652	n/a	2015–2026

Note: Counts are taken from the parsed pool-bout records; the cyclic-triad rate is the share of intransitive pool triads, defined in § 2.2.

2.2 Pillar I: Observed intransitivity in the pool phase

The pool phase is a round-robin, so within any pool of n fencers there are C(n,3) triads of head-to-head results. A triad is transitive when its three pairwise outcomes admit a consistent ordering (A beats B, B beats C, A beats C) and intransitive, or cyclic, when they do not (A beats B, B beats C, C beats A). The cyclic-triad rate, the share of evaluable triads that form a cycle, is a model-free index of disorder. It assumes nothing about latent ability and reflects only the internal consistency of results [11]. We

computed this rate for every pool and aggregated by weapon. For complete pools the count has a closed form: the number of cyclic triads equals C(n,3) minus the sum over fencers of $C(w_i, 2)$, where w_i is a fencer's win total within the pool. As complementary model-free indices we computed, for each weapon, the higher-seed win rate (the probability that the better-seeded fencer wins a pool bout) and the Kendall rank correlation (τ) between entry seed and final classification [31].

As a robustness check on the cyclic-triad rate, we also normalized circular-triad counts against the [32] theoretical maximum, a combinatorial bound that is a function of pool size n [$\frac{n^3-n}{24}$ for odd n, $\frac{n^3-4n}{24}$ for even n], rather than against the total number of triads, C(n,3). This distinction matters when pool size varies substantially across the units being compared, since the achievable ceiling on cyclic triads is itself n-dependent. In our corpus, pool size is nearly invariant across weapons (mean 6.80–6.84, with 79.8–83.6% of pools at n = 7; Table 2), so the two normalizations should agree closely if the weapon ordering is genuine rather than an artifact of differing pool sizes. Table 2 confirms this: both the raw rate and the Kendall–Smith-normalized rate preserve the same weapon ordering and, more importantly, an almost identical relative gap between weapons (foil-to-épée ratio of 0.71 under both normalizations), indicating that the reported disorder ordering is not an artifact of pool-size variation.

Table 2
Cyclic-triad rate under two normalizations, by weapon.

Weapon	Mean pool size (n)	% of pools at n = 7	Raw cyclic-triad rate (% of C(n,3))	Kendall–Smith normalized rate (% of theoretical max)	Ratio to Foil
Foil	6.81	81.3%	10.3%	25.8%	1.00
Sabre	6.80	79.8%	12.4%	31.1%	1.20
Épée	6.84	83.6%	14.5%	36.3%	1.41

2.3 Pillar II: A per-touch skill-transmission model

Bout-level summaries throw away the structure that the scoring rules act on. To recover it, we modelled the bout as a sequence of touches. We assume a symmetric race-to-five process: each touch goes to the higher-seeded fencer with probability p and to the opponent with probability 1 − p, independently across touches, and the bout ends when one fencer reaches five. The per-touch log-odds are linear in the standardized seed difference, $\operatorname{logit}(p) = \beta \cdot x_z$, so β is the per-touch skill-transmission slope, the rate at which a seed advantage becomes a per-touch probability advantage. The formulation is a per-touch analogue of the Bradley-Terry paired-comparison model [27, 28, 33]. Because a race-to-five outcome is not a single Bernoulli draw, we used the exact race likelihood, the probability of the observed final score under repeated independent touches, rather than approximating the bout as one logistic trial.

We fitted the model two ways. The first was maximum likelihood, pooling within each weapon for a point estimate of β with a cluster-bootstrap confidence interval that resamples whole competitions. This pooled estimate is a population-averaged, or marginal, quantity, and we treat it and its bootstrap interval as the primary result. The second was a Bayesian hierarchical model built on a custom brms/Stan

likelihood family that encodes the race-to-five process [29, 30]. The linear predictor was 0 + weapon:xz, giving a separate per-touch slope per weapon, with a competition-level random slope that we allowed to differ in dispersion by weapon, gr(comp, by = weapon), so that both the seed effect and its between-tournament variance can vary across the three disciplines and bout-level estimates inherit cluster-aware uncertainty. Priors were weakly informative, Normal(0, 2) on the weapon slopes and Exponential(1) on the random-slope standard deviations, and sampling used Stan's no-U-turn sampler, with convergence checked through the potential scale reduction factor and effective sample sizes; the model was refitted independently in NumPyro/JAX as a second engine and reconciled. The hierarchical slopes are conditional, or tournament-specific, quantities. Because the link is nonlinear, they are not directly comparable in magnitude to the marginal maximum-likelihood estimate, a distinction we keep in view when reporting both, and the conditional-versus-marginal divergence is governed by the weapon-specific between-tournament variance the hierarchical model estimates.

As an internal diagnostic, we compared two routes to β within each weapon. The first uses the full score distribution, which draws on information from every touch. The second uses the bout win rate alone, which draws only on the binary outcome. Under independent touches the two should agree; a systematic gap signals that touches are not independent and identically distributed, and its size measures the departure. We make the comparison explicit. From the bout-win logistic slope γ and the race-to-five amplification factor k, the win-rate route implies a per-touch slope $\frac{\gamma}{k}$, which we set against the score-based estimate. Here k is the derivative of the bout-win probability with respect to the per-touch probability at p = 0.5: because a race to five is equivalent to winning at least five of nine independent touches, the bout-win probability is $G(p) = P(\mathrm{Bin}(9, p) \geq 5)$, and the tail-derivative identity $\frac{dG}{dp} = n \cdot C(n-1, m-1) \cdot p^{m-1}(1-p)^{n-m}$ with n = 9 and m = 5 gives $k = 9 \cdot C(8,4) \cdot \left(\frac{1}{2}\right)^8 = \frac{630}{256} = 2.461$. We call the ratio of the two a seriality ratio. On this reading the independent-touch model is not a claim about how fencing is actually played but a reference process, a null against which serial structure can be measured: a seriality ratio of 1 marks independence, and the amount by which it exceeds 1 quantifies how much more decisive a bout is than independent touches would make it. We attach a cluster-bootstrap confidence interval to the ratio so that cross-weapon differences can be tested rather than merely described.

2.4 Pillar III: The earned-tier outcome and structural out-of-sample prediction

2.4.1 The tie-break artifact and its correction

Correlating entry seed with final classification looks like a natural measure of tournament predictability. It is misleading. The FIE final classification breaks ties among fencers eliminated in the same direct-elimination (DE) round by reverting to their pre-DE seeding. Every fencer who exits in each round is therefore ordered within that round by the very seeding the analysis is trying to validate. That injects a deterministic, seed-driven ranking into the outcome variable and inflates seed-outcome concordance

regardless of what happened on the strip. This is exactly the kind of structure-induced distortion the tournament-design literature warns can separate placement from merit [4, 6, 7].

To remove the artifact, we recoded each fencer's outcome as the earned tier, the DE round reached, computed from final rank as $\mathrm{tier} = \lceil \log_2(\mathrm{rank}) \rceil$. The champion occupies tier 0, the losing finalist tier 1, the two losing semifinalists tier 2, the four losing quarterfinalists tier 3, and so on. The earned tier records only the round a fencer reached on the strip and is invariant to the within-round, seed-based ordering that contaminates raw rank. We then measured predictability as the Kendall rank correlation (τ_b, which accommodates the heavy ties inherent in a tiered outcome) between entry seed and earned tier [31], with bootstrap confidence intervals over competitions.

2.4.2 Structural prediction from pool to tableau

Earned-tier concordance is an observed property of the elimination phase. We asked whether it can be predicted from the per-touch mechanism estimated entirely in the pool phase. We took each weapon's pool-fitted per-touch slope and carried it forward to a simulated DE tableau in which bouts run to fifteen touches, the junior DE format, while preserving the realised seeding structure of each event. Simulating the tableau under the pool-derived per-touch process yields a predicted earned-tier concordance, which we compared to the observed value. This is a structural out-of-sample test, not a causal or experimental claim. It asks whether a mechanism calibrated on one phase generalizes to another phase under a different bout length, with tournament structure held fixed.

3 Results

3.1 Pool-phase disorder runs épée > sabre > foil

The model-free indices place the three weapons on a clear, consistent gradient. Cyclic triads were most common in épée, where 14.5% of evaluable pool triads were intransitive, intermediate in sabre at 12.4%, and least common in foil at 10.3% (Fig. 1). The complementary indices line up the same way. The better-seeded fencer won pool bouts most often in foil and least often in épée (higher-seed win rate: foil 86.7%, sabre 84.6%, épée 82.5%), and the correlation between entry seed and final classification was strongest in foil and weakest in épée (Kendall τ: foil 0.847, sabre 0.823, épée 0.807). The same picture appears in the seed-gap discriminability curves, which trace the probability that the higher-seeded fencer wins as the standardized seed gap grows. Foil and sabre rise more steeply than épée, so a given seed advantage buys more in the priority weapons (Fig. 2). Across these measures foil is the most orderly weapon and épée the least, with sabre between them (Table 3).

Because a fully random field is not the appropriate reference for a cyclic-triad rate, we also benchmarked the observed rate against a skill-adjusted null. Under a seed-based logistic reference fitted separately to each weapon's bout outcomes, the expected cyclic-triad rate is 6–10%, far below the 25% of a fully random field; observed cyclicity exceeds this skill-based null in every weapon. Once heterogeneous skill is accounted for, the excess intransitivity (the "transitivity gap") is significantly larger in épée and sabre

than in foil, while épée and sabre are statistically indistinguishable from one another; we therefore do not order épée and sabre, and report only that both carry more true cyclicity than foil, a result stable across four independent null specifications (Supplementary Table S1, Figure S1).

*[*Figure 1 *about here]*

Table 3
Pool-phase disorder by weapon. All three model-free indices order the weapons foil > sabre > épée.

Weapon	Cyclic-triad rate	Higher-seed win rate	Seed-final τ
Foil	10.3%	86.7%	0.847
Sabre	12.4%	84.6%	0.823
Épée	14.5%	82.5%	0.807
*[*Figure 2 *about here]*			

3.2 Per-touch skill transmission is weakest in épée and strongest in foil

The per-touch model recovers the same ordering at the mechanistic level and inverts it relative to disorder: the more orderly a weapon, the more strongly each touch carries a seeding advantage. The marginal maximum-likelihood slope was lowest for épée (β = 0.48, 95% bootstrap CI [0.46, 0.51]), higher for sabre (0.80, [0.77, 0.86]), and highest for foil (0.83, [0.79, 0.89]) (Fig. 3). The intervals separate épée cleanly from the two priority weapons, which sit close together.

The Bayesian hierarchical fit tells a more layered story, and it does not simply reproduce the marginal numbers. Its conditional, tournament-specific posterior means are substantially larger than the marginal estimates (épée 1.20, foil 2.29, sabre 1.37), and that is expected rather than contradictory. In a nonlinear mixed model, the conditional (subject-specific) slope and the marginal (population-averaged) slope are distinct estimands; once the conditional slope is averaged over the random-effect distribution it shrinks toward the marginal value, so the two diverge by construction, and the larger the between-cluster variance, the larger the gap. Allowing that variance to differ by weapon makes the divergence interpretable rather than puzzling: the between-competition random-slope standard deviation is smallest in épée (1.21), larger in sabre (1.61), and largest by a clear margin in foil (2.01), with the three posterior distributions clearly separated and épée's well clear of foil's (Fig. 4). Foil's conditional slope is the most inflated relative to its marginal value precisely because foil carries the most between-tournament dispersion; the apparent anomaly of a foil conditional slope near 2.3 is a direct consequence of that dispersion, not a sign that the estimates disagree. What matters for our argument is that both estimands rank the weapons identically, foil above sabre above épée, so the ordering is a property of the data and not of the estimator. The dispersion also has a substantive consequence: with a conditional mean near 1.2 in épée and a standard deviation above that value, the seed-to-touch slope is effectively flat, and

occasionally reversed, in a non-trivial minority of tournaments, so any single summary slope averages over events that differ sharply in how strongly seeding predicts touches (Table 4).

To test the between-weapon differences directly rather than by interval overlap, we computed cluster-bootstrap confidence intervals for the pairwise slope differences. The foil–épée and sabre–épée differences were clearly separated from zero ($\Delta = 0.35$, 95% CI [0.28, 0.42]; and $\Delta = 0.32$, 95% CI [0.26, 0.39]), whereas the foil–sabre difference was not ($\Delta = 0.03$, 95% CI [– 0.06, 0.12]), confirming that épée separates from both priority weapons while foil and sabre do not separate from each other.

*[*Figure 3 *about here]*

*[*Figure 4 *about here]*

Table 4
Per-touch skill-transmission slopes by weapon.

Weapon	**β marginal (MLE)**	**95% CI**	**β conditional (hier.)**	**Between-comp. SD (σ)**	**Seriality ratio [95% CI]**
Foil	0.83	[0.79, 0.89]	2.29	2.01	1.64 [1.58, 1.70]
Sabre	0.80	[0.77, 0.86]	1.37	1.61	1.64 [1.59, 1.69]
Épée	0.48	[0.46, 0.51]	1.20	1.21	1.85 [1.80, 1.91]

Note: The marginal MLE with its cluster-bootstrap interval is the primary estimate. The conditional hierarchical means are larger because the conditional and marginal slopes are different estimands under a nonlinear link, with the gap scaled by each weapon's between-competition standard deviation σ; both estimands rank the weapons identically. The seriality ratio is the per-touch slope implied by the bout win rate alone relative to that implied by the full score distribution; it equals 1 under independent touches, and the excess above 1 indexes within-bout serial structure. Épée's ratio is significantly higher than that in both priority weapons: its bootstrap lower bound (1.80) exceeds their upper bounds (1.70 and 1.69).

The diagnostic puts a number on the departure from the independent-touch reference, and it does so in every weapon. The per-touch slope implied by the full score distribution falls below the slope implied by the bout win rate alone, so the seriality ratio is 1.85 in épée, 1.64 in sabre, and 1.64 in foil. Because the two routes coincide only under independent touches, a ratio above 1 in all three weapons measures real within-bout serial structure, consistent with the serial dependence documented for fencing at the point level [26]. The ratio is, crucially, well above 1 even in the priority weapons, where simultaneous touches are impossible, but épée's ratio is significantly higher than the priority pair rather than merely numerically larger: its bootstrap confidence interval clears theirs entirely. This pattern is the single most important constraint on how the épée result can be read, and we develop it in the Discussion: a departure that appears where simultaneous touches cannot occur cannot be a product of simultaneous touches, while the significant épée excess marks the bounded additional contribution that double touches can plausibly make.

3.3 Earned-tier concordance corrects an inflated artifact and keeps the weapon ordering

Recoding outcomes as earned tiers sharply lowered apparent predictability relative to raw final ranks. Raw-rank concordance sat between 0.81 and 0.85 across weapons (foil 0.847, sabre 0.823, épée 0.807), values inflated by the seed-based tie-break ordering described in § 2.4.1. Earned-tier concordance was markedly lower: foil τ_b = 0.684, sabre 0.650, épée 0.638. The drop is what the artifact predicts. Once the deterministic, seed-driven within-round ordering is removed, the concordance that remains reflects only what fencers achieved on the strip. The weapon ordering survives the correction, with foil the most concordant and its bootstrap interval clear of the other two, echoing the pool-phase and per-touch results (Table 5).

Table 5
Seed-to-outcome concordance under raw final rank versus the earned-tier outcome. The corrected metric removes the seed-based tie-break inflation. The reduction of roughly 0.16 to 0.19 is near-uniform across weapons, consistent with an artifact that operates through the shared seeding regime rather than through any weapon-specific feature.

Weapon	Raw-rank τ (inflated)	Earned-tier τ_b (corrected)
Foil	0.847 [0.830, 0.860]	0.684 [0.669, 0.696]
Sabre	0.823 [0.815, 0.831]	0.650 [0.640, 0.661]
Épée	0.807 [0.789, 0.821]	0.638 [0.623, 0.652]

3.4 The per-touch mechanism predicts the tableau for foil and sabre and leaves a residual gap for épée

Carrying each weapon's pool-fitted per-touch slope into the simulated fifteen-touch DE tableau produced predictions that tracked observation closely for two of the three weapons (Fig. 5). For foil, predicted earned-tier concordance (0.672) sat almost on top of the observed value (0.684). For sabre, the model slightly over-predicted (0.664 against 0.650). For épée, it under-predicted (0.573 against 0.638): the elimination phase was more orderly than the pool-calibrated reference expected. The direction and size of the épée gap line up with the diagnostic of § 3.2, where épée carried the significantly largest seriality ratio. The gap is therefore the downstream consequence of épée's stronger within-bout serial structure, read through a reference model that omits it, not a separate finding about épée.

*[*Figure 5 *about here]*

4 Discussion

Across three nested levels of analysis, model-free pool disorder, a mechanistic per-touch model, and tournament-level concordance, the three weapons hold a stable order. Foil is the most orderly: fewest intransitive triads, strongest per-touch transmission of seeding advantage, highest earned-tier

concordance. Épée is the least orderly on every measure, and sabre keeps a consistent middle position. That the ordering recurs across observational and mechanistic measures, and across independent Python and R implementations, points to a real structural property of the weapons rather than an artifact of one method.

The foil-épée contrast reads naturally against the scoring rules [3]. In foil and sabre, priority resolves simultaneous actions in favor of one fencer, so almost every exchange is a discriminating event that can express a skill difference. In épée, the simultaneous-touch rule lets both fencers score at once, and such double touches move the score without settling who was the better fencer in that exchange. A weapon that regularly produces non-discriminating touches will, all else equal, transmit a seeding advantage more weakly per touch and generate more disorder at the level of bouts and pools. The per-touch slopes and cyclic-triad rates fit that account, and the link between how abilities map to win probabilities and the expected level of cyclicity is what trait-performance models of intransitivity formalize [11]; that single-elimination structure should amplify such cycles relative to round-robin play, as observed directly in fencing result networks [25], is consistent with the gap we find between the artifact-inflated raw ranking and the earned-tier outcome.

4.1 Between-tournament heterogeneity and the referee's hand

The between-tournament heterogeneity is not uniform across weapons, and its ordering is itself informative. The competition-level random slope is most dispersed in foil ($\sigma = 2.01$), intermediate in sabre (1.61), and tightest in épée (1.21), the same ranking, inverted, that governs every other measure: the weapon whose touches most strongly encode a seeding advantage is also the weapon whose skill transmission varies most from one event to the next. The scoring rules again offer a natural reading. Épée is adjudicated by an objective electrical apparatus that registers a touch with negligible human interpretation; foil and sabre add right-of-way, a convention whose application rests on a referee's reading of who attacked first and whether an attack was properly parried. Right-of-way is precisely the part of fencing that resists codification. A prototype automated foil referee, trained on thousands of exchanges, still adjudicates priority well below the accuracy of objective hit detection and is, by its authors' own assessment, not ready to replace human officials [34]. What is hard for an algorithm is variable across human officials, and across the refereeing pools that staff different World Cup events, so a weapon that depends on right-of-way should inherit more between-competition variance in how cleanly a seeding advantage converts into touches than a weapon that does not. The pattern mirrors findings from other adjudicated combat sports, where Bayesian hierarchical models reveal that individual judges weight the same actions differently and that judging tendencies shift with venue and context [35]. Foil's elevated dispersion, and the inflated conditional slope that accompanies it, is therefore consistent with refereeing subjectivity rather than a modelling artifact: a large between-competition standard deviation is exactly what a referee-dependent scoring rule predicts, and the orderly σ gradient from épée to foil tracks the orderly gradient from objective to interpretive adjudication.

4.2 The tie-break artifact

The methodological core of the paper is the tie-break artifact. Because the FIE final classification orders same-round losers by their pre-DE seeding, any concordance computed against raw final rank is partly a correlation of seeding with itself. The inflation is large and silent: raw concordance near 0.81 to 0.85 collapses to 0.64 to 0.68 once the artifact is removed, and nothing in the published rankings flags it. The earned-tier outcome, which records only the round reached on the strip, is an artifact-free alternative we recommend for any seed-to-outcome analysis in fencing, and for any single-elimination sport whose tie-breaks revert to seeding [4, 6, 36]. That the correction is near uniform across weapons supports the diagnosis: the artifact works through the shared seeding regime, so it shifts all three weapons by a similar amount and leaves their ordering intact. The general lesson for tournament modelling is that the outcome variable deserves the same scrutiny as the model. A contaminated outcome can manufacture the appearance of order where less exists.

4.3 Reach and boundary of the per-touch reference

The structural out-of-sample test marks both the reach and the boundary of a deliberately simple per-touch reference. A reference calibrated only on pool bouts predicts foil's elimination concordance almost exactly and sabre's to within a small margin, which is a real validation: for the priority weapons, one skill-transmission process appears to govern both phases despite their different bout lengths. Épée departs from the pattern, and the departure needs care. The tempting move is to pin it on the one feature unique to épée, the double touch, and to conclude that épée obeys a different law. The data do not support that reading. The seriality ratio sits well above 1 in all three weapons, 1.64 in foil and 1.64 in sabre, where simultaneous touches cannot occur, against 1.85 in épée. If the departure were a product of double touches, it would be absent or much weaker in the priority weapons; instead, it is nearly as large there. A weapon-specific mechanism cannot generate an effect in weapons that lack the mechanism. The parsimonious reading is serial dependence common to all three weapons, the momentum, tactical adjustment, and score-dependent risk-taking that an independent-touch reference does not encode [26]. Épée's simultaneous-touch rule plausibly adds to this, and the addition is now statistically resolved rather than merely suggested: épée's seriality ratio is significantly higher than the priority-weapon baseline, with non-overlapping bootstrap intervals, so doubles can account for the residual excess of roughly 0.21 between épée's ratio and the priority pair, but not for the shared departure that all three weapons display.

This reframing also answers an obvious modelling question: why not fit an explicit double-touch model to épée, for instance an extended paired-comparison model with a tie-like third outcome [37]? Two reasons. First, the analysis above shows that doubles are not the source of the common departure, so a doubles-specific model would address the smaller part of the problem while leaving the shared serial-dependence component untouched in every weapon. Second, and decisive for this corpus, the data cannot support such a model. Our records carry the final score of each bout but not its touch-by-touch sequence, and the number of double touches cannot be recovered from a final score, because a double increments both fencers' tallies and many double counts are consistent with the same final score. A

faithful épée model would need exchange-level data that the FIE results portal does not expose. We therefore treat the épée gap as a transparent feature of where the per-touch reference stops, sharpened rather than hidden by the cross-weapon comparison, not as evidence of a weapon-specific law.

4.4 Sensitivity analysis: the role of seed endogeneity

Two further cautions shape the causal reading, and the second deserves its own analysis because it bears directly on the cross-weapon comparison. The first concerns the bout-length extrapolation. A five-touch pool bout and a fifteen-touch DE bout are not the same interaction stretched to different lengths; their momentum and pacing differ, and épée in particular layers passivity rules onto the later periods of a long bout, which press on the score in ways a length-invariant per-touch slope cannot capture. Part of the épée under-prediction may reflect these tactical features of the long format rather than the pool mechanism alone, which only reinforces the reading of the gap as a model limitation.

The second concerns the predictor itself. Entry seed is not an exogenous measure of ability; it is an output of the same FIE ranking system whose tie-break behavior we criticize, which exposes the seed-based slopes to two distinct problems. The first is attenuation: as a noisy proxy for latent skill, classical measurement error biases every seed-based slope toward zero, so the per-touch β and the seed-to-outcome concordances we report are conservative lower bounds on the true skill signal. The second is a subtler endogeneity: part of the cross-weapon gap we attribute to scoring rules could instead reflect weapon-specific differences in how well the ranking system tracks ability.

To probe this, we re-estimated the per-touch slopes using a strength measure built independently of the FIE ranking. From the head-to-head pool results alone, we computed a prospective Elo rating for every fencer, updated bout by bout and frozen at the start of each competition, so that the rating entering an event reflects only prior on-strip results and shares the pre-tournament timing of the seed without inheriting its construction. The cross-weapon ordering is robust to this substitution: the Elo-based per-touch slope is again largest in foil, intermediate in sabre, and smallest in épée, and that ordering holds across every Elo specification we tried (learning rates from 16 to 48, history thresholds from 5 to 30 prior bouts). The weapon effect is therefore not an artifact of seeding validating itself; it survives an ability signal that the FIE ranking did not generate. The magnitude, however, attenuates. Roughly half of the foil-épée gap persists under the independent rating (46 percent in the main specification, and 37 to 48 percent across the learning-rate and history-threshold grid), with the remainder attributable to weapon-specific differences in how sharply the seed tracks ability: the rank correlation between the seed gap and the independent rating gap is itself higher in foil (0.67) than in épée (0.55), with sabre intermediate (0.59), so a unit of seed advantage encodes more genuine skill in foil than in épée. In a joint model the independent rating adds little beyond seed, and seed remains the dominant predictor, consistent with the official seed being the richer ability summary, drawn from a longer competitive history than our corpus alone can reconstruct, while still carrying some weapon-specific endogeneity. The cross-weapon ordering is thus robust to seed endogeneity; only the precise magnitude of the weapon effect is partly contingent on how ability is operationalized. This sits within the broader evidence that seeding is

endogenous to outcome: when seeding is isolated experimentally, as in the regression-discontinuity design of [38], a marginal seeding advantage need not translate into deeper tournament progress at all.

Whether the same weapon ordering holds among senior elites, whose game is widely seen as more mechanical and tactically settled than the junior game, is a natural next test. Every method used here transfers directly to that population, and a senior replication would show whether junior status inflates or dampens the disorder we observe; we frame this as a hypothesis for future work rather than a claim the present data can settle. For sports science and tournament design, the results show that predictability is authored jointly by the rules of play and the structure of competition, and that the two can be separated when one is held fixed. They also restate a caution from the skill-versus-luck literature [1, 2, 12]: a headline concordance figure is only as trustworthy as the outcome coding beneath it.

4.5 Practical implications for coaching, selection, and officiating

The three pillars translate into concrete guidance for practitioners, not only for tournament modellers. First, foil's elevated between-competition variance ($\sigma = 2.01$, against 1.61 in sabre and 1.21 in épée) is consistent with referee-dependent scoring rather than pure noise, and it is measurable at the level of individual competitions. A federation that tracked this quantity across its certified referee pool would have a direct, data-driven target for right-of-way training and certification, in the same way Bayesian judge-effect models have been used to standardize adjudication in other combat sports [35]. Second, the weak per-touch transmission we find in épée (marginal slope 0.48, against 0.80–0.83 in the priority weapons) means a single épée bout carries comparatively little information about the underlying skill gap between two fencers. For talent identification and national squad selection, this argues for weighting seasonal aggregates over single-event results more heavily in épée than in foil or sabre, where a comparable slope difference does not hold. Third, the tie-break artifact we expose is not a modelling nicety: any federation, ranking committee, or media outlet using official final classification as a proxy for merit is, in part, reading back the seeding it started with. The earned-tier correction offers a low-cost alternative — round reached on the strip, rather than tie-broken final rank — that any federation could adopt without new data collection.

4.6 Implications for tournament design

The cross-weapon differences we document also speak to format design, separately from the coaching and officiating implications above. FIE runs all three weapons through the same pool-to-tableau architecture, but our per-touch estimates show that a single épée bout carries markedly less information about the underlying skill gap than a single foil or sabre bout does (marginal slope 0.48 against 0.80 to 0.83). Under a fixed format, this asymmetry means épée's single-elimination phase is structurally noisier. The same seeding protocol and bracket depth that reliably protect a top seed in foil offer weaker protection in épée, not because the seeding is flawed but because each individual result is simply a less reliable signal to seed on. This fits network-level evidence that single-elimination structure amplifies cyclic, upset-prone outcomes relative to round-robin play [25], and it suggests a weapon-uniform format

is not actually weapon-neutral in its consequences. Épée would likely need either a longer pool phase or more seeding protection in the early bracket rounds to match the outcome reliability that foil and sabre already achieve under the current format. The earned-tier correction we introduce gives federations a way to measure this reliability directly, rather than inferring it from a tie-break-inflated ranking, and it offers a concrete, weapon-specific quantity, the gap between earned-tier concordance and its raw-rank counterpart, that any proposed format change could be evaluated against.

Limitations

Several considerations bound the conclusions. The per-touch model treats touches as independent by design: it is a reference process, and the seriality ratio (1.6 to 1.9 across weapons) measures how far real bouts depart from it. That departure is informative rather than a defect, because it isolates and puts a single number on the within-bout serial dependence, the momentum, tactical adjustment, and score-dependent risk-taking, that the reference deliberately omits [26]. What the present data cannot do is decompose that departure into its sources. With only final scores and no touch-by-touch sequence, we cannot separate momentum from score-dependent stopping or from heterogeneity across exchanges, so the per-touch slopes are best read as effective summaries of a richer process rather than literal exchange-level probabilities.

Related to this, the model has no representation of the simultaneous-touch rule unique to épée. As the Discussion sets out, this is not the source of the common departure, since a large seriality ratio also appears in the priority weapons, but épée's significant excess over that baseline is consistent with a genuine, if bounded, double-touch contribution that the model does not represent. Modelling it properly would require exchange-level data that record each double touch, which the FIE results portal does not provide, so the question is left open rather than resolved here.

The hierarchical results carry their own caveat. The random-slope standard deviation is large relative to the conditional slopes in every weapon, and largest in foil ($\sigma = 2.01$), which means the strength with which seeding predicts touches varies widely across tournaments and is effectively absent in some. We report the marginal maximum-likelihood slope with a cluster-bootstrap interval as the primary estimate for this reason and treat the conditional hierarchical means as a description of tournament-specific behavior rather than as a competing population estimate. The two should not be conflated, and the gap between them is a feature of the nonlinear model scaled by the weapon-specific between-tournament variance, not a sign of instability.

The evidence is observational throughout. The earned-tier concordance and the pool-to-tableau propagation are a structural out-of-sample prediction, not a randomized experiment or a causal identification. We can say whether a pool-calibrated mechanism generalizes to the tableau; we cannot attribute that generalization to a single causal factor. The seed-endogeneity sensitivity analysis sharpens but does not eliminate this limit: it shows that the cross-weapon ordering survives an independent, FIE-ranking-free Elo rating built from head-to-head results, not that the difference is causally identified, and the magnitude of the weapon effect remains partly contingent on how ability is measured.

As a sensitivity check, we re-estimated all three headline measures separately within men's and women's events (Supplementary Table S5). The weapon ordering, épée least orderly, foil most, is preserved in both genders across all three pillars, so the reported ordering is not an artifact of pooling across gender. Per-touch slopes are uniformly higher in women's events (e.g. foil $\beta = 0.98$ vs 0.75), but the cross-weapon ordering is identical.

Finally, the corpus is limited to FIE junior individual World Cup events from 2015 to 2026. Junior fencers may differ from seniors in consistency and depth of field, and the World Cup population is a selected, elite sample, so whether the weapon ordering and its magnitudes carry over to senior, veteran, team, or domestic competition is an open empirical question.

The structural out-of-sample prediction in § 2.4 assumes the FIE's standard single-elimination seeding bracket (top seeds in opposite halves, byes to the highest seeds). This matches the sport's documented procedure, but it is load bearing: re-running the prediction under a random draw or adjacent (1–2, 3–4) seeding collapses predicted concordance in every weapon, and the épée under-prediction of 0.064 is specific to the standard, seeded bracket (Supplementary Table S2). We therefore report the épée gap as conditional on the documented FIE seeding procedure, not as a bracket-independent property of the sport.

A further concern is that the reported weapon differences could reflect differences in competitive depth or ability spread rather than the rules of play. Mean field size and mean pool size are closely matched across weapons (mean pool size 6.80–6.84), and the spread of a pool-indicator proxy for within-field ability is smallest, not largest, in épée (Supplementary Table S3), the opposite of what a wider-field explanation of its high cyclicity would require.

Bouts that did not reach five touches within the time limit were excluded from the per-touch model; this exclusion is asymmetric across weapons (épée 6.7%, foil 1.9%, sabre 0.01%). Refitting the marginal per-touch slope with these bouts included, using their realized winning score, changes the épée estimate by only − 0.008 and leaves the épée–foil ordering intact (Supplementary Table S4), so the result reported in § 3.2 is not an artifact of this exclusion.

Conclusions

Using the largest structural corpus of fencing results assembled to date, 158,652 pool bouts across 474 Junior World Cup competitions, we show that the three weapons hold stable, distinct positions on the skill-chance continuum. Foil is the most orderly, épée the least, and sabre between them, consistently across model-free, mechanistic, and tournament-level measures. The study's central methodological result is the tie-break artifact: because the FIE final classification orders same-round losers by their pre-DE seeding, conventional seed-to-rank concordance is inflated from roughly 0.64 to 0.68 up to 0.81 to 0.85, a correlation of seeding with itself that the published rankings nowhere flag. The earned-tier outcome removes it with a single recoding and transfers to any single-elimination sport whose tie-breaks revert to seeding. A deliberately simple per-touch reference, calibrated on pool play, predicts elimination-

phase orderliness for the priority weapons and leaves a residual gap for épée. We attribute that gap to where the independent-touch reference stops rather than to a weapon-specific mechanism, a reading the cross-weapon evidence makes hard to avoid: the same departure from independence appears in foil and sabre, where simultaneous touches cannot occur, and épée's significant excess over that shared baseline marks only the bounded part a double-touch rule can plausibly add.

Two practical implications follow, offered as directions rather than finished tools. First, the earned-tier outcome gives organizers a fairness diagnostic they currently lack comparing seed to earned tier, rather than to published rank, shows how much of a competition's apparent seed-to-result agreement is real and how much is the tie-break echo, and the same recoding can audit whether a seeding procedure is tracking on-strip results or merely reproducing its own prior. Second, because the weapons differ in how strongly seeding predicts outcome, a single seeding-and-format template need not serve all three equally; a discipline with weaker per-touch transmission and more pool disorder, such as épée, has more room for an underdog and may warrant different bracket or pool-sizing choices than foil if equal competitive fairness is the goal. We raise these as questions for coaches and tournament designers to weigh, not as prescriptions the present analysis can validate on its own.

The broader contribution is methodological as much as empirical. In fencing, as in any tournament sport, the measure of predictability is only as sound as the outcome it is built on.

Abbreviations

CI: confidence interval

DE: direct elimination

FIE: International Fencing Federation (Fédération Internationale d'Escrime)

MLE: maximum likelihood estimation

OSF: Open Science Framework

SD: standard deviation

τ_b: Kendall's tau-b concordance coefficient

Declarations

Availability of data and materials. The analytic corpus and all analysis scripts (Python and R) are openly available in an Open Science Framework (OSF) repository (https://osf.io/jtb8h). The data derive from publicly available FIE Junior World Cup results (2015–2026).

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests. The authors (NK, EG, MÖ, SEY) declare that they have no competing interests.

Authors' contributions. NK: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing, Project administration. EG: Formal analysis, Software, Data curation, Validation, Writing – review & editing. MÖ: Investigation, Data curation, Resources, Writing – review & editing. SEY: Investigation, Software, Formal analysis, Visualization, Writing – review & editing. All authors read and approved the final manuscript.

Ethics approval and consent to participate. This study did not involve human participants, human tissue, or animals, and no interventions were carried out. It is based entirely on the secondary analysis of publicly available competition results published by the International Fencing Federation (FIE) for its Junior World Cup events (2015–2026). The analysis uses bout- and tournament-level records (seeds, scores, and final classifications) and does not collect or process any private or identifiable personal data beyond what the FIE has already made public. In accordance with institutional and national guidelines for research based solely on publicly available records, ethics committee approval and informed consent were not required.

Consent for publication. Not applicable.

Acknowledgements. Not applicable.

Use of generative AI and AI-assisted technologies. During the preparation of this work, the authors used generative AI tools to assist with the English-language translation and editing of text originally drafted in Turkish, and to assist in debugging portions of the analysis code (Python and R). After using these tools, the authors reviewed, verified, and edited all outputs as needed and take full responsibility for the content of the manuscript, including the correctness of the analyses and the integrity of the results. No AI tool is listed as an author or fulfils the criteria for authorship.

References

1. Aoki RYS, Assuncao RM, Vaz de Melo POS. Luck is Hard to Beat. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2017. pp. 1367–76. https://doi.org/10.1145/3097983.3098045
2. Mauboussin MJ. The success equation: Untangling skill and luck in business, sports, and investing. Harvard Business Review; 2012.
3. Roi GS, Bianchedi D. The Science of Fencing. Sports Med. 2008;38:465–81. https://doi.org/10.2165/00007256-200838060-00003.
4. Csató L. Tournament Design. Cham: Springer International Publishing; 2021. https://doi.org/10.1007/978-3-030-59844-0.
5. Horen J, Riezman R. Comparing Draws for Single Elimination Tournaments. Oper Res. 1985;33:249–62. https://doi.org/10.1287/opre.33.2.249.

6. Hwang FK. New Concepts in Seeding Knockout Tournaments. Am Math Monthly. 1982;89:235–9. https://doi.org/10.1080/00029890.1982.11995420.
7. Israel RB. Stronger Players Need Not Win More Knockout Tournaments. J Am Stat Assoc. 1981;76:950–1. https://doi.org/10.1080/01621459.1981.10477747.
8. Karpov A. A new knockout tournament seeding method and its axiomatic justification. Oper Res Lett. 2016;44:706–11. https://doi.org/10.1016/j.orl.2016.09.003.
9. Mcgarry T, Schutz RW. Efficacy of traditional sport tournament structures. J Oper Res Soc. 1997;48:65–74. https://doi.org/10.1057/palgrave.jors.2600330.
10. Scarf P, Yusof MM, Bilbao M. A numerical study of designs for sporting contests. Eur J Oper Res. 2009;198:190–8. https://doi.org/10.1016/j.ejor.2008.07.029.
11. Strang A, Abbott KC, Thomas PJ, The Network HHD. Quantifying Cyclic Competition in Trait-Performance Models of Tournaments. SIAM Rev. 2022;64:360–91. https://doi.org/10.1137/20M1321012.
12. Jerdee M, Newman MEJ. Luck, skill, and depth of competition in games and social hierarchies. Sci Adv. 2024;10. https://doi.org/10.1126/sciadv.adn2654.
13. Zappalà C, Pluchino A, Rapisarda A, Biondo AE, Sobkowicz P. On the role of chance in fencing tournaments: An agent-based approach. PLoS ONE. 2022;17:e0267541. https://doi.org/10.1371/journal.pone.0267541.
14. Zappalà C, Biondo AE, Pluchino A, Rapisarda A. The paradox of talent: How chance affects success in tennis tournaments. Chaos Solitons Fractals. 2023;176:114088. https://doi.org/10.1016/j.chaos.2023.114088.
15. Holzmeister F, Johannesson M. Skills vs. Luck: Decomposing Deviations from Expected Performance in European Football Leagues. J Sports Econom. 2025;26:953–75. https://doi.org/10.1177/15270025251374620.
16. Csurilla G, Boros Z, Fűrész DI, Gyimesi A, Raab M, Sterbenz T. How Much Is Winning a Matter of Luck? A Comparison of 3 × 3 and 5v5 Basketball. Int J Environ Res Public Health. 2023;20:2911. https://doi.org/10.3390/ijerph20042911.
17. van Ours JC. Nontransitive Patterns in Long-Term Football Rivalries. J Sports Econom. 2024;25:802–26. https://doi.org/10.1177/15270025241264235.
18. van Ours JC. Non-transitive patterns in sports match outcomes: a profitable anomaly. Empir Econ. 2025;69:4057–87. https://doi.org/10.1007/s00181-025-02838-6.
19. Spearing J, Tawn J, Irons D, Paulden T. Modeling Intransitivity in Pairwise Comparisons with Application to Baseball Data. J Comput Graphical Stat. 2023;32:1383–92. https://doi.org/10.1080/10618600.2023.2177299.
20. Glickman ME. Paired comparison models with strength-dependent ties and order effects. Stat Modelling. 2025. https://doi.org/10.1177/1471082X251400474.

21. Ingram M. How to extend Elo: a Bayesian perspective. J Quant Anal Sports. 2021;17:203–19. https://doi.org/10.1515/jqas-2020-0066.
22. Kovalchik S. Extension of the Elo rating system to margin of victory. Int J Forecast. 2020;36:1329–41. https://doi.org/10.1016/j.ijforecast.2020.01.006.
23. Newman MEJ. Efficient computation of rankings from pairwise comparisons. J Mach Learn Res. 2023;24:1–25.
24. Tarragó R, Bottoms L, Iglesias X. Temporal demands of elite fencing. PLoS ONE. 2023;18:e0285033. https://doi.org/10.1371/journal.pone.0285033.
25. Asztalos B, Balázs B, Palla G, Vicsek T. Hierarchy and ranking in fencing and tennis. Sci Rep. 2026. https://doi.org/10.1038/s41598-026-53459-7.
26. Doron J, Gaudreau P. A Point-by-Point Analysis of Performance in a Fencing Match: Psychological Processes Associated with Winning and Losing Streaks. J Sport Exerc Psychol. 2014;36:3–13. https://doi.org/10.1123/jsep.2013-0043.
27. Bradley RA, Terry ME. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. Biometrika. 1952;39(3/4):324. https://doi.org/10.2307/2334029.
28. Cattelan M. Stat Sci. 2012;27. https://doi.org/10.1214/12-STS396. Models for Paired Comparison Data: A Review with Emphasis on Dependent Data.
29. Bürkner P-C. brms: An *R* Package for Bayesian Multilevel Models Using *Stan*. J Stat Softw. 2017;80. https://doi.org/10.18637/jss.v080.i01.
30. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. *Stan*: A Probabilistic Programming Language. J Stat Softw. 2017;76. https://doi.org/10.18637/jss.v076.i01.
31. Kendall MG. A New Measure of Rank Correlation. Biometrika. 1938;30. https://doi.org/10.2307/2332226. 1/2:81.
32. Kendall MG, Smith BB. On the Method of Paired Comparisons. Biometrika. 1940;31. https://doi.org/10.2307/2332613. 3/4:324.
33. Turner H, Firth D. Bradley-Terry Models in *R*: The BradleyTerry2 Package. J Stat Softw. 2012;48. https://doi.org/10.18637/jss.v048.i09.
34. Chen Z, Wang Z. FERA: A Pose-Based Framework for Rule-Grounded Multimedia Decision Support with a Foil Fencing Case Study. 2026.
35. Holmes B, McHale IG, Żychaluk K. Detecting individual preferences and erroneous verdicts in mixed martial arts judging using Bayesian hierarchical models. Eur J Oper Res. 2024;312:733–45. https://doi.org/10.1016/j.ejor.2023.07.004.
36. Csató L. How to avoid uncompetitive games? The importance of tie-breaking rules. Eur J Oper Res. 2023;307:1260–9. https://doi.org/10.1016/j.ejor.2022.11.015.
37. Davidson RR. On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments. J Am Stat Assoc. 1970;65:317–28. https://doi.org/10.1080/01621459.1970.10481082.

38. Engist O, Merkus E, Schafmeister F. The Effect of Seeding on Tournament Outcomes: Evidence From a Regression-Discontinuity Design. J Sports Econom. 2021;22:115–36. https://doi.org/10.1177/1527002520955212.

Figures

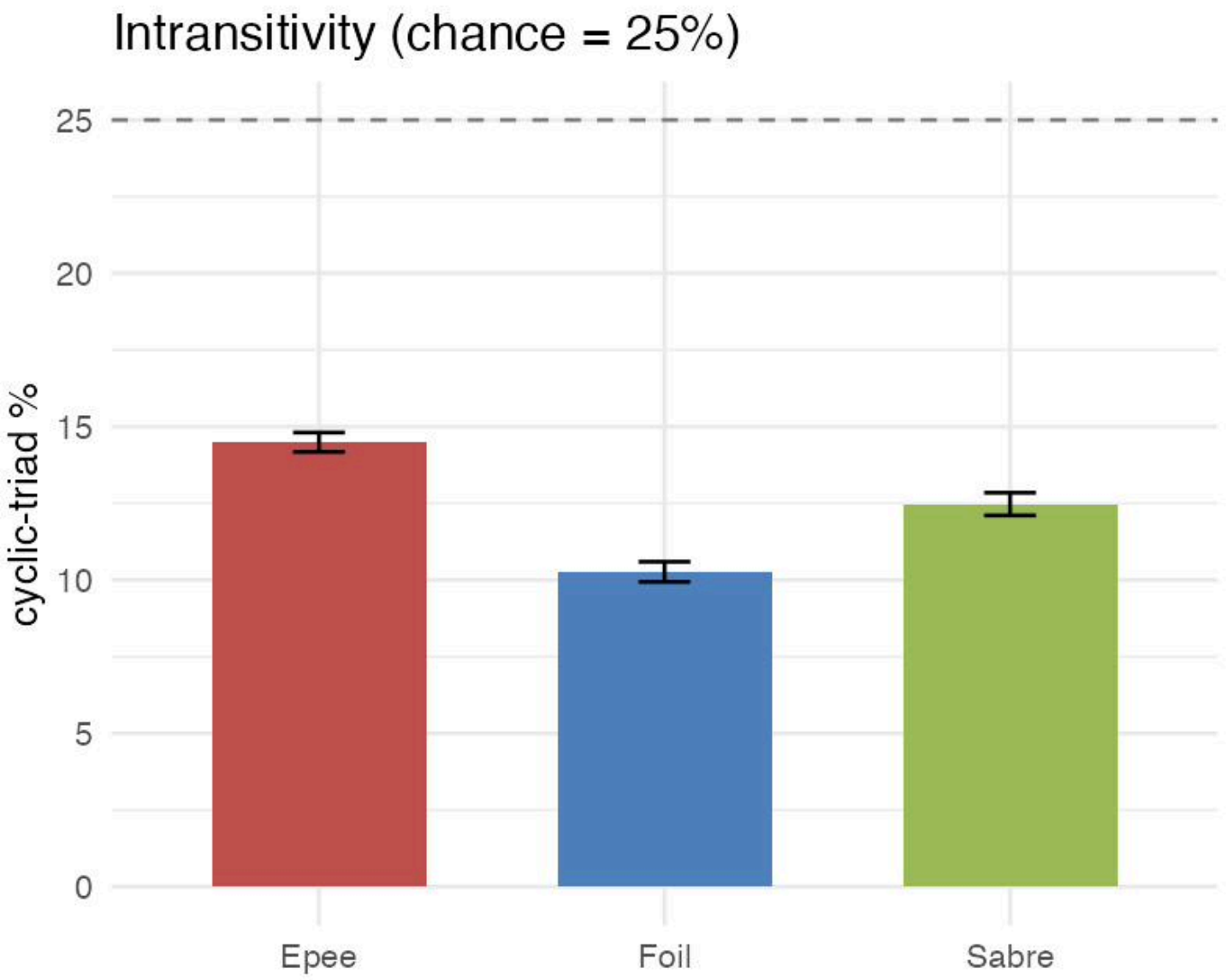


Figure 1

Cyclic-triad rate by weapon. The share of intransitive head-to-head triads within pools, a model-free index of disorder, rises from foil to sabre to épée. The dashed line marks the 25% rate expected under fully random outcomes.

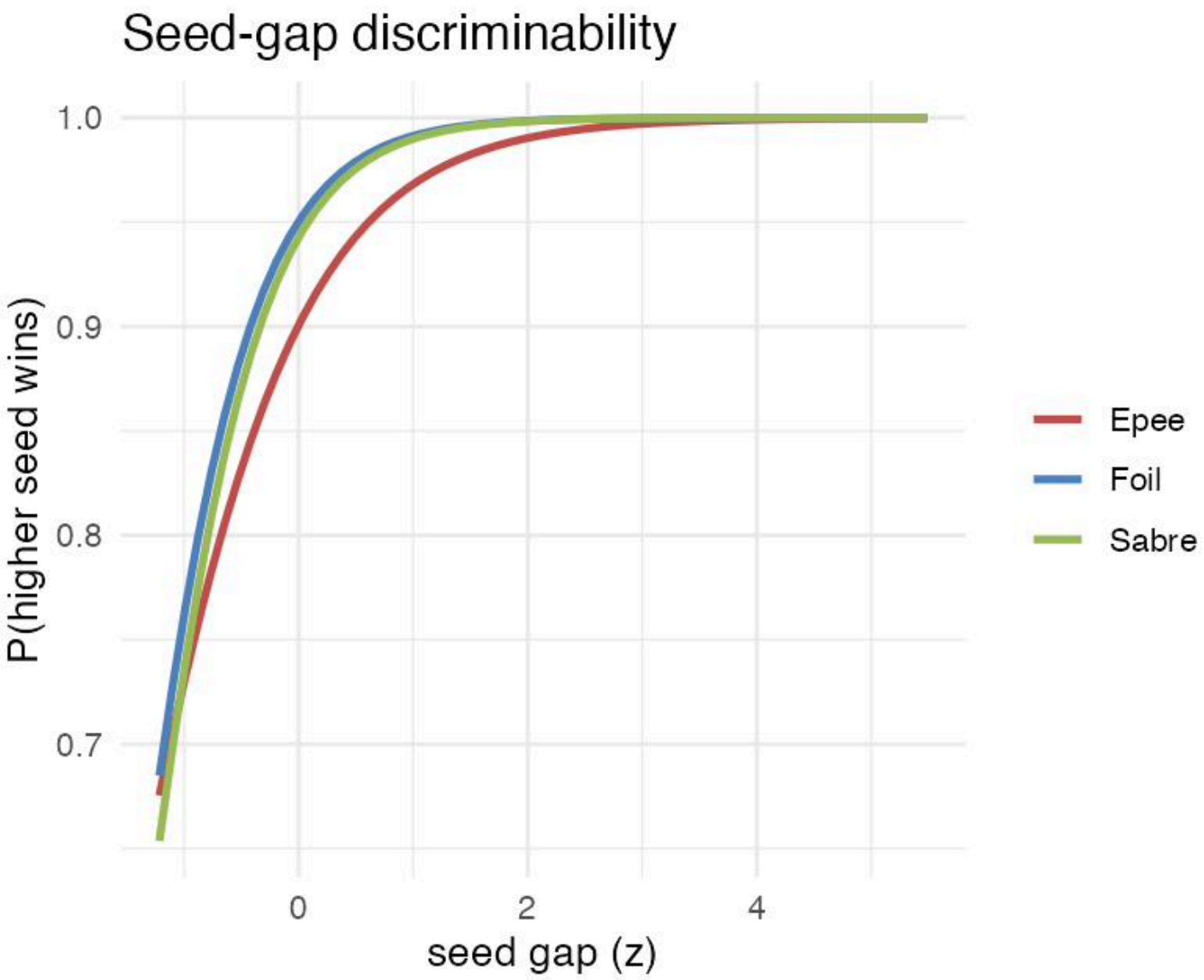


Figure 2

Seed-gap discriminability by weapon. Each curve gives the probability that the higher-seeded fencer wins as a function of the standardized seed gap. The foil and sabre curves rise more steeply than épée, so a fixed seed advantage translates into a larger win-probability advantage in the priority weapons.



Figure 3

Per-touch skill-transmission curves by weapon. Each curve is the fitted probability that the better-seeded fencer wins a touch as a function of the standardized seed advantage, using the marginal maximum-likelihood slope. A steeper curve means a given seed advantage is expressed more strongly at the level of the individual touch. Foil and sabre nearly coincide; épée is flatter throughout.

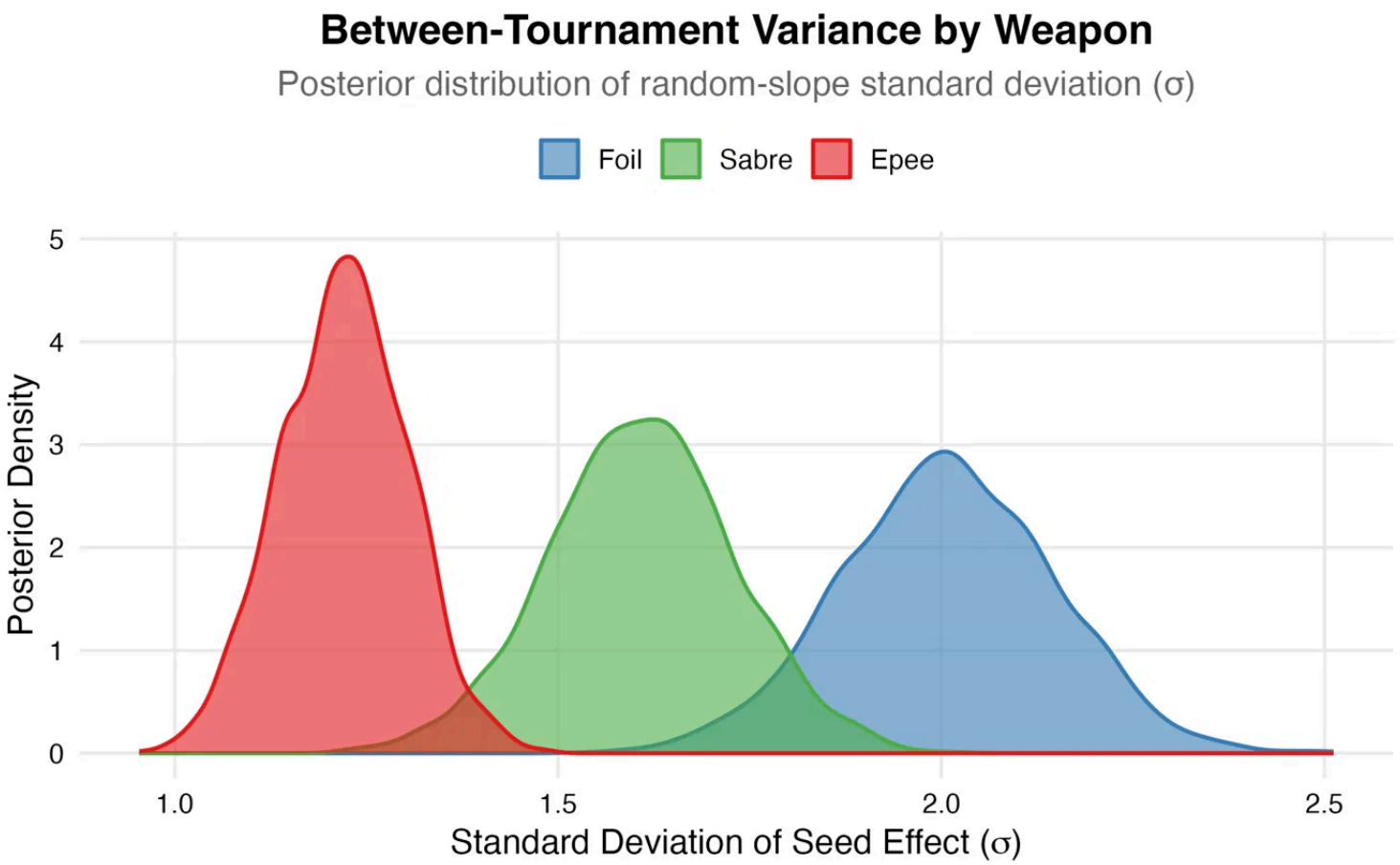


Figure 4

Between-tournament variance by weapon: posterior distributions of the competition-level random-slope standard deviation (σ). The three posteriors are clearly separated — épée most concentrated and smallest (σ ≈ 1.21), sabre intermediate (≈ 1.61), and foil largest and most dispersed (≈ 2.01) — so the conditional slope is inflated most in the weapon that carries the most between-tournament heterogeneity.

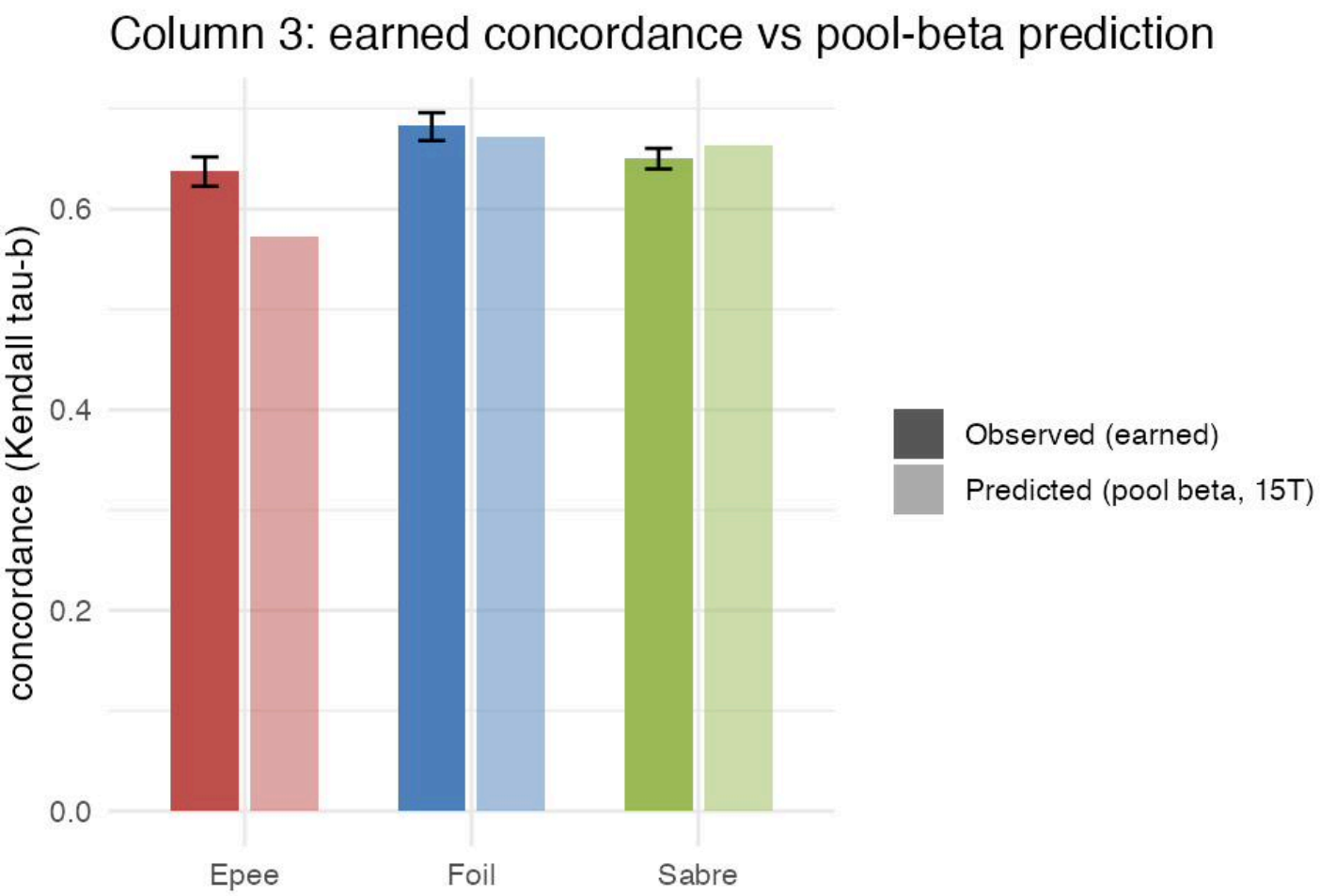


Figure 5

Observed versus structurally predicted earned-tier concordance. Dark bars are the observed earned-tier τ_b with bootstrap intervals; light bars are the value predicted by carrying the pool-fitted per-touch slope into a simulated fifteen-touch tableau. Foil and sabre nearly match; the predicted épée value falls short of what the elimination phase delivered.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx