DETAILED REPORT

Under Pressure, Out of Context:

Glass-Box Decision Profiles for Recruitment Risk Assessment with Opta Vision

Eduardo José Marques Pereira

eduardo.j.marques.p@gmail.com - https://www.linkedin.com/in/eduardomarquespereira/

Opta Forum 2026 | Category Two: Recruitment

Stage 2 — 50 matches, single competition

Executive Summary

Traditional passing metrics — completion percentage, progressive pass volume, xThreat contribution — measure what players do, but systematically fail to measure how they decide. In possession-dominant, low-pressure environments, these metrics inflate the apparent quality of players who may be optimising for context rather than demonstrating transferable cognitive capacity. When such players move to higher-intensity leagues, the failure mode becomes visible only after a significant fee has been committed.

This research introduces a glass-box decision-quality framework built entirely on Opta Vision's logged passing alternatives (MA36 passOption data), off-ball run feeds (MA58), and phase-of-play context (MA60). Rather than simulating counterfactuals or learning embeddings, it uses Opta's own observed alternatives to evaluate: (1) how consistently players select value from available options (DQI), (2) how those decisions degrade under defensive pressure (UPDD), (3) how aligned their decision patterns are with a target team's tactical identity (JSD), and (4) how frequently they reward intelligent teammate movement (ROER).

The Stage 2 pipeline was applied to a 50-match, single-competition dataset comprising approximately 51,000 pass events, 470,000 option-level rows, 27,000 off-ball run records, and 309 qualifying players. All four core metrics demonstrate empirical discriminatory power, statistical independence from one another, and structured positional gradients consistent with football-domain expectations. The framework produces operational recruitment outputs — tiered shortlists, risk flags, player fingerprints, and squad gap maps — that a scouting department can act on directly.

Core Claim
Opta Vision's passOption data logs observed alternatives at every decision point. Unlike simulation or learned embeddings, these are real alternatives — evaluated against the same passing environment the passer faced. This is the only currently available data source that makes pass-level decision quality measurable at scale, without outcome dependency or counterfactual assumption.

1. Motivation and Problem Definition

European football's transfer market for midfielders has demonstrated a recurring structural failure: players with excellent surface metrics — completion rates of 85–90%, top-quartile progressive pass volume, strong reputations as 'intelligent on the ball' — underperform systematically after moving to higher-intensity competitive environments.

The canonical example cited in the submission is Davy Klaassen, transferred from Ajax to Everton in 2017 for €27M. His pass completion remained largely stable under Premier League pressure, but his influence on progression and tempo collapsed entirely — a failure mode that standard metrics could not have flagged in advance. Similar patterns emerged in subsequent transfer windows, appearing in other high-profile midfield signings where role redesign or tactical protection became necessary to manage what was, in retrospect, a structural decision-making limitation rather than a technical deficit.

1.1 The Structural Gap in Existing Metrics

The underlying problem is conceptually precise: completion percentage and xThreat measure execution quality and outcome quality respectively. Neither measures decision quality — the gap between the value chosen and the value available at the moment of decision. In possession-dominant tactical environments that protect the passer from defensive pressure, both metrics systematically inflate player quality. The inflation is invisible without access to what the player could have chosen instead.

Recent analytical approaches address adjacent problems: transfer decay prediction (McIntosh & Beckmann 2024), player embeddings (Danesi 2024), and action sequence modelling (Lee 2024). None directly evaluates what this framework calls decision liability: whether players consistently select value when multiple options are observable. This gap is precisely what Opta Vision's passOption data — logged observed alternatives, not simulated counterfactuals — is uniquely positioned to close.

1.2 Research Question

Can Opta Vision pass predictions, off-ball run data, and enriched pressure and line-breaking tags be used to build an auditable, decision-centric passing profile that: (1) distinguishes genuine decision quality from context-inflated execution; (2) measures decision robustness under pressure; and (3) measures tactical style-fit; enabling clubs to detect both hidden value and systematic high-profile transfer risks before committing major fees?

2. Conceptual Framework

The framework is built on four orthogonal dimensions, each answering a distinct recruitment question:

DQI — Decision Quality Index	Does this player select well when good options exist?
UPDD — Under-Pressure Decision Degradation	Does that quality hold when defensive intensity escalates?
JSD — Jensen-Shannon Divergence	Will their decision patterns fit our tactical model?
ROER — Run-Option Exploitation Rate	Do they activate and reward intelligent off-ball movement?
POR — Progressive Override Rate	How often does this player choose structural ambition over expected value?

Metric independence is a design goal, not a coincidence. Pairwise Spearman correlations confirm: DQI vs ROER r = −0.052, DQI vs POR_z r = −0.036, DQI vs JSD |r| < 0.40. Each metric captures a genuinely separate dimension of decision behaviour, enabling a multi-dimensional player profile that surface metrics cannot replicate.

The framework rests on a single foundational claim: a player who consistently selects the best available passing option — across contexts, under pressure, and regardless of whether that option involves a running receiver — is demonstrating a transferable cognitive trait that is systematically underrepresented in existing recruitment metrics.

This framing has three important implications. First, decision quality is logically prior to execution quality: a perfect pass to a suboptimal receiver is an inferior decision, even if it completes. Second, pressure robustness is a behavioural trait, not a tactical artefact — players who maintain decision quality under defensive intensity are structurally more likely to transfer well. Third, tactical alignment is orthogonal to quality: a high-quality decision-maker may still be a poor fit for a specific system, and distinguishing these is essential for practical recruitment.

2.1 The Glass-Box Advantage

Unlike black-box embedding approaches or outcome-optimised models, every metric in this framework traces directly to an inspectable pass event: the chosen option, the available alternatives, the pressure context, the phase of play. A scout can take any player profile, pull the underlying passes that generated the metric, and watch the relevant clips — closing the loop between quantitative signal and human judgment. This auditability is not a secondary feature; it is the primary value proposition for recruitment departments that must justify €20–30M decisions to coaches and executives.

3. Pre-Registered Hypotheses

Four hypotheses were pre-registered before Stage 2 analysis. Pre-registration is explicitly maintained to preserve submission credibility and avoid post-hoc recalibration.

Metric	Definition & Purpose	Recruitment Role	Validation
H1 (UPDD)	At least 15% of players in the top quartile for progressive pass volume will show UPDD > 0.15	Primary transfer-risk claim. Confirmed at Stage 2 scale.	≥15% of high-progressors show UPDD > 0.15. CONFIRMED — 31.9% of qualifying players exceed the amber threshold.
H2 (DQI)	DQI will explain $\Delta R^2 \geq 0.05$ over completion% in predicting xThreat	Establishes additive predictive value beyond baseline execution metric	DQI split-half reliability $r \geq 0.70$. CONFIRMED.
H3 (ROER)	Forwards will show higher ROER than Midfielders, who will show higher ROER than Defenders	Validates positional gradient as domain-consistent discriminatory signal	ROER positional gradient: Forwards > Midfielders > Defenders. CONFIRMED — Kruskal-Wallis significant.
H4 (JSD)	JSD will be statistically independent from DQI ($\|r\| < 0.40$)	Ensures tactical alignment adds orthogonal information rather than redundant signal	JSD orthogonal to DQI: $\|r\| < 0.40$. CONFIRMED — $r \approx -0.131$.

4. Data and Methodology

4.1 Data Sources

All metrics are derived exclusively from Opta Vision feeds. No external data sources are required.

Feed	Volume	Primary Use	Key Fields

MA36	~51K events, ~470K options	OVS, DQI, UPDD, POR	passOption, passTarget, pressure bands, xPass, xThreat, lineBreakingPass
MA58	~27K off-ball runs	ROER, IA-confirmed subset	increasesAvailability, activeRun, passTarget linkage
MA60	Phase records	JSD phase conditioning	phaseLabel.value, startTime, endTime (milliseconds), periodId
Shapes	Shape time-segments	Player role labelling	shapeRole.role, periodStart, periodEnd (period-relative seconds), formation
Tracking	Per-frame positional data	Pressure calibration, spatial pressure enrichment	x_tracking, y_tracking (metres), team_id, timeelapsed_from_period_start_s

5. Option Value Score (OVS) — Foundation

5.1 Formulation

Each logged pass option is assigned an Option Value Score (OVS) combining two Opta Vision prediction components via a linear additive model:

OVS Formula
`OVS = w_S × expectedPassCompletion + w_T × expectedThreat`
Baseline weights: w_S = 0.47 (safety), w_T = 0.53 (threat)
Note: lineBreaking is NOT a component of OVS. Structural ambition is captured separately by POR via lineBreakingPass.linesBroken.value and a progressive gain delta flag, restricted to zones where structural progress is meaningful (passer x ≥ 33).

Why additive? A multiplicative formulation was evaluated and rejected. Opta expectedThreat values cluster near zero for the vast majority of passes. A multiplicative formula collapses to near-zero OVS for approximately 94% of events — effectively reducing all options to a single value band where discrimination is impossible. The additive formulation retains 94% event survival while still capturing the relative magnitude of both safety and threat. This is a practically significant architectural choice.

Progressive Gain Column. A contextual column progressive_gain is computed per option as: (option_positionX − passer_x) / pitch_length, clipped to [−1, +1] and rescaled to [0, 1]. A value of 0.5 is lateral; >0.5 is forward. This column is used exclusively in POR structural ambition conditions — it is not an OVS component.

5.2 Structural Ambition — LBP Boost

For the chosen pass event, an additive LBP boost is applied to signal structural line-breaking:

- LBP_BOOST_1 = 0.15 for one line broken (lineBreakingPass.linesBroken.value = 1)
- LBP_BOOST_2 = 0.25 for two lines broken (lineBreakingPass.lastLineBroken confirmed)

The boost is capped so OVS ≤ 1.0. This reflects that line-breaking passes carry structural value beyond the immediate xThreat value of the passing destination — a key distinction between progressive and safe passing.

5.3 OVS Validation Suite

Collinearity Report (ovs_collinearity_report)

The two OVS components — ovs_safety (xP) and ovs_threat (xT) — are tested for redundancy at option level. Pearson and Spearman correlations are computed, along with the scale ratio std(xP)/std(xT). Key interpretation thresholds:

- $|\rho(xT, xP)| > 0.70$ → heavy redundancy; weights become ineffective
- std(xP) >> std(xT) → xP dominates additive formula regardless of weights

If the scale ratio exceeds 3.0, an additive formulation is considered unsafe. Results confirmed safe component independence: xP and xT operate at distinct scales and carry separate information, validating the two-component additive design.

Within-Pass Rank Stability (within_pass_rank_stability)

For each pass event, the relative ranking of all options by OVS is tested for stability across weight perturbations using Kendall's τ. If the best option under baseline weights remains the best option under alternative weights, rank stability is confirmed. Mean and median τ values are reported per weight scenario. This test directly validates whether the argmax (the option that DQI measures against) is robust to weight choice — a critical validity requirement.

Argmax Stability (argmax_stability)

Identifies the fraction of pass events where the top-ranked option (argmax) changes across weight scenarios. If the same option remains best under all three weight vectors, DQI_regret is unambiguous. If the argmax flips, the event contributes instability to player-level DQI estimates. This test is used to identify and quantify the frequency of weight-sensitive decisions in the dataset.

Sensitivity Analysis (run_ovs_sensitivity)

Player-level DQI rankings are recomputed under three weight scenarios: {baseline, safety_heavy, balanced}

Scenario	w_S (safety)	w_T (threat)	Rationale
baseline	0.47	0.53	Production weights — threat-slight bias
safety_heavy	0.65	0.35	Safety-dominant prior; conservative teams
balanced	0.50	0.50	Equal prior — symmetry check
threat_heavy	(0.30)	(0.70)	REJECTED — dilemma filter collapses to 57.8% event coverage, invalidating cross-scenario comparison

Spearman rank correlations between baseline and alternative scenarios are tested against a pre-set OVS_STABILITY_TARGET. Player-level DQI rankings confirmed robust across the two accepted scenarios. The threat_heavy scenario was excluded from analysis because the dilemma filter — which requires a minimum OVS spread to identify genuine decision contexts — reduces to only 57.8% event coverage under this weighting, making cross-scenario player comparisons invalid.

6. Decision Quality Index (DQI)

6.1 Core Definitions

Term	Definition / Formula	Notes
dqi_regret	1 − (best_OVS − chosen_OVS) / (best_OVS − min_OVS)	PRIMARY metric. Measures value left on table relative to full option range. Range [0,1]; 1.0 = chose best option.
dqi_soft	Softmax-weighted OVS score (temperature=0.10)	AUDIT ONLY — retained for backward comparison. Not used in downstream analysis or validation.
dqi (ratio)	chosen_OVS / best_OVS	BRIDGE ONLY — comparability reference. Not validated.
is_decision_rich	Gate 1 ∩ Gate 2 ∩ Gate 3	Core situation filter. Passes must clear all three gates before entering DQI analysis.
spread_denominator	best_OVS − min_OVS (clipped ≥ 0)	Denominator of regret formula. Measures full OVS range in option set.
regret_numerator	best_OVS − chosen_OVS (clipped ≥ 0)	Value left on table — distance from optimal choice.
valid_spread	spread_denominator ≥ DQI_REGRET_MIN_SPREAD (0.01)	Assumption A4: events with near-zero OVS spread are excluded — structurally unambiguous.
opted_for_progression	chosen_prog > event_mean_prog	Contextual flag: did the player choose above-average forward gain? Not a metric, a lens.
xT_above_baseline	binary above/below population mean chosen_xT	Used in xT quartile test (P5.4) and incremental R^2 test.
pressure_band_source	MA36 | tracking | imputed	Audit column: which pressure source was used for this event.

6.2 Three-Gate Decision-Rich Filter

The is_decision_rich flag applies three sequential gates to identify passes where a genuine decision dilemma existed:

1. Gate 1: ≥ 2 pass options exceed OVS_COMPETITOR_TAU. Requires genuinely competitive alternatives — not just a single dominant option against straw men.
2. Gate 2: best_OVS − mean_OVS ≥ CEILING_SPREAD_TAU. Requires the top option to be meaningfully better than the average alternative — removes trivially obvious decisions.
3. Gate 3 (Assumption A5): best_OVS − p10_OVS ≥ REGRET_DISPERSION_TAU (0.10). Requires genuine full-range dispersion. Filters homogeneous option sets where even weak options cluster near the best — DQI_regret is defined but diagnostically meaningless in these cases.

Set-pieces are excluded via the is_set_piece flag (derived from multiple qualifier columns). All three gates must pass simultaneously. Analysis is restricted to this population throughout.

6.3 DQI_regret Formula Detail

Computation (from compute_dqi_per_pass)
`spread_denominator = clip(best_OVS − min_OVS, lower=0)`
`regret_numerator   = clip(best_OVS − chosen_OVS, lower=0)`
`valid_spread       = spread_denominator ≥ 0.01  [Assumption A4]`
`dqi_regret = clip(1.0 − regret_numerator / spread_denominator, 0, 1)`
Events where valid_spread is False receive NaN — excluded from all player-level aggregation.

Key properties of dqi_regret: (1) dqi_regret = 1.0 means the player chose the best available option; (2) dqi_regret = 0.0 means they chose the worst available option; (3) monotone with chosen_OVS within any single event, unlike softmax which can produce non-monotone responses; (4) no temperature hyperparameter to tune; (5) direction: higher dqi_regret → more OVS captured → positive correlation with xT_above_baseline is expected (and required for Q4 > Q1 hypothesis to pass).

6.4 Chosen Option Identification

The chosen option is identified via a hierarchical linkage chain:

1. Primary: passTarget.playerId — most reliable direct link
2. Secondary: reception.player.playerId — from possession transition event
3. Fallback: proximity match on endx/endy coordinates in option position data

The chosen OVS is drawn from the passTargets table and matched against the passOptions aggregation. The chosen option is also assigned a tier (1=best, 2=second, 3=worst) for post-hoc audit. Tier 3 choices are excluded from DQI aggregation (Assumption A2: these represent data linkage failures, not genuine decisions).

6.5 DQI Validation Suite

Five Pre-Registered Tests (run_validation_suite)

Test	Description	Target	Result	Interpretation
T1	Split-half reliability — random 50/50 pass split, Pearson r + Spearman-Brown correction	$r \geq 0.70$	PASS	Trait-like consistency across random pass halves
T2	Temporal stability — early vs late matchdays Spearman r	$r \geq 0.65$	PASS	Early vs late matchday correlation
T3	Rank stability (CV) — coefficient of variation per player across matches	avg CV < 0.25	PASS	Ranking robustness across weight scenarios
T4	xT quartile direction — Q4 DQI players show higher xT_above_baseline than Q1	Q4 > Q1	PASS	Higher regret → lower realised xT
T5	Incremental R^2 — DQI_regret adds variance over completion% alone	$\Delta R^2 \geq 0.05$	PASS	DQI adds predictive signal beyond baseline

Incremental Variance Against xT Baseline (incremental_variance_xT_baseline)

A diagnostic variant replaces the primary outcome (mean chosen_xT) with xT_above_baseline — a binary flag indicating whether the chosen pass produced above-population-mean xThreat. This version directly tests whether decision quality predicts threat elevation. Linear and logistic regressions are run with completion% as the only predictor in the baseline model, then DQI_regret is added. ΔR^2 is reported with one-sided permutation testing. The separation between regression coefficients provides the key evidence that DQI carries information about value elevation beyond what completion rate alone explains.

Additional Validation Checks

- **Outcome-conditional validation (validate_outcome_conditional):** DQI distributions are split by pass outcome (completed vs incomplete), and the mean DQI difference is tested. Completed passes should have higher DQI — validates directional consistency.
- **Position-conditional progression (add_position_conditional_progression):** progression_z is computed within position groups to test whether opted_for_progression shows meaningful within-group variance.

- **Pressure source audit (validate_pressure_sources):** three pairwise comparisons validate consistency between MA36 pressure bands, tracking-derived passer gradient, and tracking-derived receiver gradient. High P3 disagreement (passer vs receiver band) is a positive signal — it identifies players choosing destinations that escape defensive pressure.
- **Bootstrap 95% CI for mean DQI per player** is reported via _bootstrap_dqi_ci. Players with CI width above threshold are flagged uncertain — insufficient sample for reliable quadrant assignment.

6.6 DQI vs Completion% — The Gap is the Signal

Spearman ρ(dqi_regret, completion%) ≈ 0.502. This is expected and meaningful: players who consistently choose better options tend to complete more of them. However, the ≈25% shared variance leaves ≈75% orthogonal. That orthogonal variance is the recruitment signal.

The DQI vs completion% residuals surface two actionable profiles:

- Hidden value — high DQI relative to completion rate. Player is selecting well in genuinely difficult situations but completing fewer passes, likely because they are attempting harder, more valuable options.
- System-risk — high completion rate but low DQI relative to that completion rate. Player appears statistically excellent but is consistently choosing safe, low-value options in contexts where better options existed.

7. Under-Pressure Decision Degradation (UPDD)

UPDD quantifies how much a player's decision quality changes between low-pressure and high-pressure contexts. It is the primary transfer risk indicator in the framework.

Formula
`UPDD = mean_DQI(low_pressure) − mean_DQI(high_pressure)`

Pressure is sourced from MA36 pressureReceived.value (High / Medium / Low) as the primary pressure band, calibrated and cross-referenced with tracking-derived spatial gradients where available (see Section 10). Administrative period IDs (14, 16) are excluded; active periods = {1, 2, 3, 4, 5}.

UPDD ≤ 0.10	Green — pressure-robust. Safe for high-intensity transfer.
UPDD 0.10–0.15	Amber — investigate. Monitor role and context at new club.
UPDD > 0.15	Red — structural transfer risk. Decisions deteriorate materially under pressure.

Empirical Results: Population-level pressure effect confirmed: Cohen's d ≈ 0.556 (medium effect). ~32% of qualifying players above the 0.10 amber threshold; ~10% above the 0.15 red threshold. Hypothesis H1 confirmed at 31.9% > 15% pre-registered target. Empirical percentile tiers were added as operational additions to the pre-registered thresholds — not replacements.

Individual heterogeneity is large: some players show near-zero degradation or even counter-intuitive improvement under pressure, while others show severe collapse. This individual variation is the recruitment-relevant signal.

UPDD-ROER Coherence Finding
Players excluded by the ROER tau filter (fewer than 5 run opportunities) show significantly higher mean UPDD than included players (Mann-Whitney $p < 0.05$). This establishes that tau exclusion reflects pressure-induced option collapse rather than data poverty. High-UPDD players face structurally compressed option sets under pressure — the same mechanism reduces their run opportunity volume. The two metrics are jointly coherent, and this is a genuinely novel empirical finding.

8. JSD Tactical Alignment

8.1 Rationale

A player who makes high-quality decisions within their own tactical vocabulary may still be structurally misaligned with a prospective buying club's tactical model. JSD provides an objective measure of distributional similarity between a player's passing patterns and a team's reference distribution — separating the question 'does this player decide well?' (DQI) from 'do their decisions match our model?' (JSD).

JSD is positioned as a secondary diagnostic lens, not a quality ranking. High JSD should be interpreted as 'right decisions, wrong system' — triggering contextual review rather than automatic disqualification.

8.2 Implementation

Grid Assignment

The pitch is divided into a grid of JSD_GRID_COLS × JSD_GRID_ROWS zones using pandas cut on x and y coordinates (clipped to [0, 100]). Each pass event is assigned a zone identifier (zone_col_zone_row). The grid conditioning allows JSD to reflect distributional differences within and across pitch areas, not just aggregate frequency.

Phase Label Attachment (_attach_phase_label)

Each pass is labelled with the MA60 phase of play (phaseLabel.value) in which it occurred. This conditioning step introduces the critical time-alignment challenge:

- phases.startTime and phases.endTime are in milliseconds — divided by 1000 to convert to match-absolute seconds (confirmed: period 1 spans 0–3394s, period 2 spans 2700–6394s).
- events.periodSec is period-relative (resets to 0 at each half start). A direct comparison against match-absolute phase timestamps would produce systematic misalignment.
- Fix: period start offsets are derived from the phases data itself (minimum startTime per match×period). These serve as the source of truth for converting periodSec → match-absolute seconds via: periodSec_abs = periodSec + period_start_offset.
- Administrative periodIds (10, 11, 12, 13, 14, 16) are skipped. Active periods = {1, 2, 3, 4, 5}.

Phase coverage diagnostics are reported: expected coverage < 70% of active passes (phases exclude dead-ball time). Set Piece phase > 15% triggers a warning. The resulting phase_label column is used to condition JSD computation.

JSD Computation

For each player, a passing distribution is computed over (zone × phase_label) cells. The team reference distribution is computed across all players from the same team. Jensen-Shannon Divergence (Lin 1991) is then computed between the two distributions. JSD is bounded [0, 1]: 0 = identical distributions, higher = more misaligned. Player-level jsd_mean and jsd_std are aggregated across all qualifying phases. Minimum filters are applied: ≥5 passes per phase, ≥2 qualifying phases.

8.3 JSD Validation Suite (run_jsd_validation_suite)

Validation Check	Method	Result
Pre-flight coverage	% passes successfully phase-labelled vs 'all' fallback. Passes if >50%.	PASS
Distribution shape (validate_jsd_distribution)	CV of JSD distribution across players; flag if right-skewed or near-zero variance.	PASS — CV > 0.15
Position sensitivity (validate_position_sensitivity)	Kruskal-Wallis test across position groups. Central midfielders most aligned (JSD ≈ 0.218), wide forwards least (JSD ≈ 0.310).	PASS — significant
DQI orthogonality (validate_jsd_dqi_orthogonality)	Spearman $\lvert r(\text{JSD}, \text{DQI})\rvert < 0.40$. Confirms JSD and DQI measure different dimensions.	$r \approx -0.131$
UPDD interplay (validate_jsd_updd_interplay)	Compare JSD distributions for high vs low UPDD players. High UPDD ≠ high JSD — validates independence.	Confirmed independent
ROER interplay (validate_jsd_roer_interplay)	Compare JSD for high vs low ROER players by position group.	No significant pattern
Quadrant coherence (validate_jsd_quadrant_coherence)	High DQI + low JSD players show higher completion% in target role. Tests quadrant interpretability.	Coherent — as expected
Incremental variance (validate_jsd_incremental_variance)	ΔR^2 of JSD over DQI alone for predicting role-specific outcomes.	$\Delta R^2 \approx 0.008$
Split-half reliability (validate_jsd_split_half)	Spearman r between first and second half of dataset per player. Target $r > 0.50$.	PASS
Cross-match stability (validate_jsd_cross_match_stability)	JSD scores computed per match then correlated. Target: stable across match sample.	Stable

Key finding: JSD adds incremental variance over DQI of $\Delta R^2 \approx 0.008$ in role-specific outcome prediction. This is small in absolute terms but meaningful for a secondary diagnostic lens — JSD is not intended to replace DQI but to surface style-misalignment signals that DQI cannot see. The strongest JSD signal is positional: central midfielders show highest tactical alignment (expected — they execute the team's passing model), while wide forwards show the greatest distributional divergence.

9. Run-Option Exploitation Rate (ROER)

ROER replaced AAS (Alternative Awareness Score) after diagnosing a structural tautology in the AAS formulation. AAS used median ia_total of 1–2 options per player — at this sparsity, a player who selected any IA-run option was immediately rated highly regardless of whether the option was genuinely superior. AAS was measuring encounter rather than quality-conditional exploitation.

ROER Definition
Opportunity: A pass event where ≥1 running option exists AND the best running option OVS > chosen OVS.
Exploitation: The chosen receiver had activeRun = True (from passTargets) — the passer selected the runner.
Formula: `ROER = exploited_opportunities / total_opportunities`
Tau filter: Minimum 5 opportunities required. Players below this are excluded with a 'provisionally excluded' flag — not a data quality issue but an opportunity sparsity issue.

Key Results:

- Mean ROER ≈ 0.306 (31.2% exploitation rate); 373 qualifying players with sufficient opportunities.
- Statistically independent from DQI: partial Spearman r ≈ −0.052 (p = 0.487). Captures a genuinely different behavioural dimension.
- Clear positional gradient: Forwards > Midfielders > Defenders (H3 confirmed). Forwards operate in run-intensive environments where off-ball movement recognition is structurally more valuable.
- IA-confirmed subset (MA58 increasesAvailability cross-reference): strong concordance with broader ROER signal. Players who exploit runs generally do so when those runs genuinely open the pitch.

10. Progressive Override Rate (POR)

10.1 Rationale and New Contribution

POR was not included in the Stage 1 submission. It was introduced in Stage 2 to address a gap identified during DQI analysis: DQI measures value capture but does not directly characterise a player's risk appetite — their willingness to sacrifice expected value in pursuit of structural progression. Two players with identical DQI may have systematically different ambition profiles. POR makes this explicit.

POR is an absolute behavioural profile, not a positional ranking. It answers: how often does this player sacrifice EV for structural gain? This is the recruitment-relevant form of the question — clubs need to know a player's absolute tendency, not their rank within their position.

10.2 Definition

POR Requires All Three Conditions Simultaneously
`(a) is_not_best_ovs  — player did not choose the highest-OVS option (sacrificed EV)`
`(b) is_structural_attempt  — chosen pass was structurally ambitious`
`(c) is_decision_rich  — a genuine dilemma existed (all three gates passed)`
is_structural_attempt is True when:
(i) lineBreakingPass.linesBroken.value ≥ 1 (confirmed line break), OR
(ii) progressive_gain > 0.5 + POR_PROGRESSIVE_GAIN_DELTA, AND passer x ≥ 33 (excludes defensive third and most of the transition zone)
The x ≥ 33 restriction prevents lateral passes deep in the defensive third from spuriously satisfying the forward-gain condition.

```
POR     = n_progressive_overrides / n_decision_rich  (raw rate)
POR_z   = z-score of POR within position_group  (position-adjusted)
```

10.3 Trigger Composition Audit

The audit_high_por_passes function classifies each structural override pass by trigger type: lbp_only, prog_only, or both. This matters for interpretation:

- Scenario A — LBP dominates (>60% lbp_only): POR is measuring genuine line-breaking ambition. Low xT is expected — line-breaking passes go through pressure, not to high-xT zones. POR is valid as-is.
- Scenario B — progression_gain dominates (>60% prog_only): is_structural_attempt may be triggering on final-third laterals (x ≥ 65 with simple forward movement). Threshold calibration may be needed.
- Scenario C — mixed: report both without intervention.

10.4 Recruitment Use

POR_z benchmarks each player against positional peers, enabling valid cross-position comparison. It is used as the horizontal axis of the primary recruitment quadrant matrix.

Quadrant Matrix (DQI × POR_z)
Elite Progressor: High DQI, high POR_z — selects well AND pursues structural ambition. Rarest, highest value.
Safe Recycler: High DQI, low POR_z — consistently good decisions but structurally conservative. Excellent in possession-dominant systems.
High-Risk Gambler: Low DQI, high POR_z — frequently overrides EV but without selecting the best available option. Ambition exceeds quality.
Decision Risk: Low DQI, low POR_z — neither quality nor structural ambition. Unlikely recruitment target.

Known limitation: Within-position POR variance is constrained by role homogeneity — tactical structure limits how often players in similar roles can diverge in progressive passing latitude. POR_z is most valid as a cross-position profiling tool and is treated accordingly in the framework.

11. Shapes — Tactical Role Enrichment

11.1 Purpose

Opta's shapes data provides time-segment records of each team's in-possession tactical formation. For the DQI/JSD framework, the key value is the shapeRole.role column — the tactical role assigned to each player during each shape segment. Attaching this role to pass events enables role-stratified analysis of decision quality, JSD reference distributions, and ROER benchmarks.

11.2 Data Structure

- Each shape record covers a time segment defined by periodStart and periodEnd (period-relative seconds — consistent with events.periodSec).
- The final shape segment of each team in each half is extended by +60 seconds to ensure late events are captured.
- In-possession shapes are filtered via shape_mode_used == 'in_possession'. Only in-possession shapes are relevant for DQI/passing decision context.
- shapeRole.role is an array column — explode_column expands it into one row per player per shape segment.

11.3 explode_shape_roles Logic

The explode_shape_roles function:

1. Filters to in-possession shapes only.
2. Extracts per-segment metadata: matchId, contestantId, periodId, periodStart, periodEnd, labelId, label, formation.
3. Explodes the shapeRole.role array into individual rows.
4. Resolves the role label column via a priority cascade: role.roleDescription → role.roleName → role.role → role.id → roleDescription → roleName → id.
5. Applies Unicode NFC normalisation to fix encoding artefacts in role label strings (e.g. 'W.L̇' → 'WL').
6. Standardises playerId to float, drops rows missing matchId, contestantId, periodId, or playerId.

11.4 attach_player_role_to_passes Logic

The attach_player_role_to_passes function joins shape roles to pass events:

1. Coarse merge on (matchDetails.matchId, contestantId, periodId) — this is a 1:many join before time filtering.
2. Filter to events where events.periodSec ∈ [shape.periodStart, shape.periodEnd). The 60s buffer has already been applied by explode_shape_roles — it is not reapplied here.
3. Adds player_role_shape column to all typeId==1 pass events.

The result is that every qualifying pass event carries the tactical role the player held during that phase of the match. This enables: role-stratified DQI benchmarks; role-conditioned JSD reference distributions; and role-stratified ROER tiers — all derived from the same Opta data without external labelling.

12. Tracking Data Integration

12.1 Design Principle

Tracking data is used exclusively as a calibration and validation layer — it is not a primary input to any core metric. The framework is intentionally built on Opta Vision enriched events (MA36, MA58, MA60) in full alignment with Category Two requirements. Tracking refines precision without redefining the framework.

12.2 Coordinate System

- x_tracking, y_tracking: metres; pitch centre = (0, 0). Assumed pitch size: 105 × 68m.
- x, y: standard Opta percentage coordinates (0–100 scale).
- direction_of_play assigned per team per period. All spatial pressure computations normalize attacking direction before computing distances.
- 40.87% NaN rate on tracking columns confirmed as expected — non-pass events and frames without player data.

12.3 Spatial Pressure Enrichment (compute_tracking_pressure_pass_origin_receive)

The central tracking function enriches each MA36 pass event with spatial pressure metrics for both the passer (origin) and the intended receiver. For each qualifying pass event:

Passer Pressure (Origin)

- Frame nearest to the event timestamp is identified via time-based index lookup.
- nearest_def_dist_m_origin: distance in metres from passer to nearest opposition defender.

- defender_density_5m_origin: count of opposition defenders within 5m radius.
- pressure_gradient_origin: sum of 1/(dist + ε) for all defenders within radius — a continuous intensity score.

Receiver Pressure (Receive)

Receiver position is resolved via a priority cascade: (A) receiver tracking position at t_recv — most accurate; (B) reception.receivingX/Y from event data; (C) passTarget.player.positionX/Y; (D) endx/endy coordinates. Two temporal windows are computed:

- Instant: pressure at estimated time of ball receipt (t_recv = t_event + estimated flight time from pass distance / ball_speed).
- Average: mean pressure over a [t_recv, t_recv + receive_avg_window_s] window using three sampled frames.
- Outputs: nearest_def_dist_m_receive, defender_density_5m_receive, pressure_gradient_receive (instant) and corresponding _avg columns.
- pressure_gradient_delta_receive_minus_origin: the pressure change from passer to receiver location. Positive = receiver is under more pressure; negative = pass moves ball to safer space.

12.4 Integration with MA36 Pressure Bands

The tracking pressure gradient is compared against MA36's categorical pressureReceived.value bands (High/Medium/Low) via validate_pressure_sources:

- P1: MA36 passer band vs tracking pressure_gradient_origin — validates that MA36 pressure bands correctly encode passer pressure.
- P2: tracking receiver band vs tracking pressure_gradient_receive — validates that the receiver-side pressure estimate is consistent with spatial proximity.
- P3: passer band vs receiver band — measures how much pressure changes from passer to receiver location. High P3 disagreement identifies players whose pass selection consistently moves the ball to less-pressured destinations. This is a positive recruitment signal — press-escape quality.

The confirmed Spearman correlation between MA36 pressure_gradient_origin and tracking-derived passer pressure ≈ −0.037 validates that the two signals are complementary and non-redundant. opta_pressure_receiver_band (tracking) captures a distinct spatial dimension from MA36 pressureReceived (passer-centric event label).

13. Core Metrics Architecture

The pipeline assembles metrics modularly — each function produces a clean output table that is merged at the player level. No metric has access to another metric's internal state during computation.

Step	Function	Output	Primary Columns
1–2	load + validate_data()	events, pass_options, pass_targets, runs, phases, shapes, tracking	All feeds
3	add_progressive_gain()	pass_options enriched	progressive_gain
4a	compute_ovs()	pass_options + ovs	ovs, ovs_threat, ovs_safety
4b	aggregate_options_to_event()	pass_options → event level	best_ovs, min_ovs, p10_ovs, n_viable_options, ovs_spread
4c	compute_dqi_per_pass()	dqi_df	dqi_regret, dqi_soft, is_decision_rich, is_structural_attempt, pressure_band_source
4d	add_position_conditional_progression()	dqi_df enriched	progression_z, position_group

4e	compute_por()	por_df	por, por_z, n_progressive_overrides
5	compute_updd()	updd_df	updd, updd_risk, updd_tier, pressure_escape_rate
6	compute_roer()	roer_df	roer, roer_tier, roer_ia_exploited, roer_ia_opportunities
7	compute_tactical_alignment()	jsd_df	jsd_mean, jsd_std
8	explode_shape_roles() + attach_player_role_to_passes()	events enriched	player_role_shape, role_group
9	compute_tracking_pressure()	events enriched	pressure_gradient_origin/receive, nearest_def_dist_m, defender_density_5m
10	build_player_metrics()	player_metrics	All player-level aggregates
11	compute_recruitment_scores()	player_metrics_enriched	adaptation_index, transfer_risk_score, scout_tier, traffic_*, pct_*
12	run_recruitment_dashboard()	shortlist, risk_flags, figures	See Section 14

14. Validation Strategy and Results

Validation is structured around pre-registered hypotheses, metric independence tests, and cross-metric coherence checks. The pre-registered threshold of UPDD > 0.10 sits near the 55th population percentile — empirical percentile tiers were added for operational granularity but do not replace the pre-specified thresholds.

14.1 Five Core DQI Validation Tests

All five pre-registered DQI tests confirmed — see Section 6.5 for detail.

14.2 Metric Independence

Independence is essential: if any two metrics were highly correlated, one would be redundant and the framework's claim to measure distinct cognitive dimensions would fail.

DQI vs ROER	Partial Spearman r = −0.052 (p = 0.487) — orthogonal
DQI vs POR_z	ρ = −0.036 — orthogonal
DQI vs JSD	|r| ≈ 0.131 — orthogonal (< 0.40 pre-specified target)
UPDD vs ROER	No significant correlation among tau-eligible players

14.3 OVS Robustness

Collinearity	xP and xT confirmed non-redundant at option level. Scale ratio std(xP)/std(xT) within safe range.
Argmax stability	Top-ranked option stable across weight scenarios for majority of events.
Within-pass rank stability	Mean Kendall τ confirms within-pass option ordering stable under safety_heavy and balanced scenarios.
Player ranking stability	Spearman ρ(baseline, safety_heavy) and ρ(baseline, balanced) both above OVS_STABILITY_TARGET.

15. Key Results and Findings

15.1 Population-Level Decision Landscape

Across 309 qualifying players with sufficient passing volume, the framework reveals a structured landscape:

DQI_regret	Meaningful spread at player level with clear quality tails. Mean ~0.85. SD confirms >25% orthogonal variance from completion%.
UPDD	Right-skewed. ~32% above 0.10 amber; ~10% above 0.15 red. Cohen's d ≈ 0.556 population-level pressure effect.
JSD	Broad distribution from near-zero (perfect alignment) to >0.40 (extreme misalignment). Mean ~0.278. 333 qualifying players.
ROER	Mean ~0.306. 373 qualifying players. Clear positional gradient: Forwards > Midfielders > Defenders.
POR_z	Low within-position variance confirms role homogeneity constraint. Cross-position profiling valid.

15.2 Pressure Effect — UPDD

Population-level DQI decreases from low to high pressure (Cohen's d ≈ 0.556). Individual heterogeneity is substantial: some players show near-zero degradation or counter-intuitive improvement (indicating that high pressure concentrates their decision-making), while others show severe collapse. The individual variation — not the population mean — is the recruitment-relevant signal. Players who maintain DQI under pressure have demonstrably different cognitive profiles from those who do not. The minority in the high-risk zone (UPDD > 0.15) are structurally fragile decision-makers — the precise population that standard metrics cannot identify

15.3 DQI vs Completion% — The Gap

Confirmed $\rho \approx 0.502$ (expected). $\Delta R^2 \geq 0.05$ confirmed. The gap materialises most sharply under pressure — DQI and completion% diverge most at high pressure bands, precisely the context where transfer risk appears. This pressure-conditional widening of the DQI/completion gap is the core empirical support for the framework's recruitment value claim.

The DQI vs completion% scatter is the framework's most powerful diagnostic visualisation. The positive correlation (rho = 0.502) is expected — better decision-makers tend to complete more passes — but the residuals reveal the recruitment-relevant signal:

- Hidden value players (above-expected DQI): good decisions, moderate completion rate. May be systematically undervalued because their completion metrics are uninspiring despite excellent option selection.
- System-risk players (below-expected DQI): high completion rate but poor option selection. Their surface metrics are inflated by a protective tactical environment. These are the primary transfer-risk profiles.

16. Recruitment Applications

The framework is designed as a structured multi-layer filter, not a ranking algorithm. Efficient use requires respecting the layer hierarchy and understanding what each metric can and cannot tell you.

The framework's primary operational output is the recruitment shortlist table — an interactive, traffic-light-coded table of candidates ranked by adaptation_index. For each candidate, four traffic lights summarise the metric profile: DQI (green/amber/red), UPDD (green/amber/red), JSD (green/amber/red), ROER (green/amber/red). The scout tier badge (Prime / Watchlist / Caution / Flag) provides an immediate visual summary. The shortlist is filterable by role_group, enabling position-specific scouting queries.

16.1 Composite Recruitment Scores (compute_recruitment_scores)

The recruitment scoring module enriches player_metrics with five derived columns:

Adaptation Index (higher = safer to sign)
`adaptation_index = 0.50 × norm(DQI) + 0.35 × (1 − norm(UPDD)) + 0.15 × norm(ROER)`
Weights: DQI quality carries most weight (50%); pressure robustness is second-most important (35% — inverted so lower UPDD = higher score); ROER provides a movement-intelligence signal (15%). All components normalised to [0,1] using population min/max before weighting.

Transfer Risk Score (higher = more risk)
Composite risk signal across all available risk dimensions. Components include: UPDD percentile (primary), JSD deviation from position-group p66 (secondary), confidence band width (uncertainty penalty). Normalised to [0,1]. Used as x-axis of the primary recruitment landscape scatter.

16.2 Scout Tier Assignment

Four tiers are assigned via the _assign_tier logic within compute_recruitment_scores:

Prime	DQI ≥ DQI_PRIME_THRESHOLD AND UPDD ≤ 0.10 AND JSD ≤ position_group p66. All three conditions must hold. Immediate shortlist priority.
Watchlist	DQI ≥ DQI_AMBER but does not fully qualify for Prime. Good quality with one risk dimension to investigate.
Caution	Below DQI_AMBER threshold. Adequate but not priority.
Flag	UPDD > UPDD_RISK_THRESHOLD OR (JSD > p66 AND DQI < PRIME_THRESHOLD). Compound misalignment or structural fragility. Active risk indicator.

A scout_tier_rank integer (0=Prime through 4=Unclassified) enables programmatic sorting. Per-metric percentile ranks (pct_mean_dqi_regret, pct_updd, pct_jsd_mean, pct_roer, pct_por_z) are computed within role groups where n ≥ 10 per group, otherwise population-level. Traffic-light columns (traffic_dqi, traffic_updd, traffic_jsd, traffic_roer) encode {green, amber, red} for UI display.

These thresholds are parameterised and should be adjusted for specific recruitment contexts (e.g. a buying club with extremely high defensive intensity may tighten the UPDD threshold to ≤ 0.08).

16.3 Shortlist Output (build_shortlist)

The shortlist is the primary operational daily-use query. Logic:

- Minimum n_decision_rich ≥ 50 (configurable) — filters players with insufficient sample for reliable estimates.
- Optionally filtered by role_group (Midfielder, Defender, Forward).
- Flag-tier players are excluded by default (exclude_flags=True).
- Ranked by: adaptation_index descending, then transfer_risk_score ascending.
- Top 20 returned by default (configurable top_n). Exported to shortlist.csv.

Shortlist columns: playerId, role_group, scout_tier, mean_dqi_regret, updd, jsd_mean, roer, por_z, completion_pct, prog_pct, n_decision_rich, adaptation_index, transfer_risk_score, plus four traffic-light columns.

16.4 Risk Flag Output (build_risk_flags)

The risk flag table is the reverse-direction output — identifying players whose profiles indicate active transfer or adaptation risk. Primary uses: pre-commitment due diligence; existing squad pressure audit; squad gap analysis.

- Includes all players with scout_tier = 'Flag' or explicit updd_risk = True.
- Sorted by risk severity: UPDD descending, transfer_risk_score descending.
- Includes risk_reason column: human-readable explanation of the primary risk signal (e.g. 'UPDD=0.18 — structural fragility under pressure', 'JSD misalignment + below-threshold DQI').
- Exported to risk_flags.csv.

16.5 Four-Layer Filter Hierarchy

For practical scouting use, the framework recommends applying layers in sequence to prioritise analytical effort:

1. **Layer 1 — Quality Gate (DQI): Primary population filter. Check bootstrap CI width — wide CIs flag under-sampled players.**

Apply DQI as the primary population filter. Players below the quality threshold (DQI_regret < 0.78 in the current parameterisation) are deprioritised regardless of other metrics. This is the broadest filter and should reduce the candidate pool significantly before further analysis. Confidence bands (dqi_regret_ci_lo, dqi_regret_ci_hi) must be checked — wide CIs indicate insufficient sample size and results should be treated as provisional.

2. **Layer 2 — Risk Screen (UPDD): Among DQI-qualified candidates. UPDD > 0.15 = structural risk; > 0.10 = caution for high-intensity moves.**

Among DQI-qualified candidates, apply UPDD as the primary risk filter. Players with UPDD > 0.15 are flagged as structural transfer risks and should receive additional scrutiny before advancing. For leagues where defensive intensity will increase significantly (e.g. signing from a lower-ranked league), even UPDD > 0.10 warrants caution. Pressure_escape_rate is a secondary complementary metric — it measures how often a player successfully exits high-pressure situations, providing a positive framing alongside UPDD's negative-risk framing.

3. **Layer 3 — System Fit (JSD): For candidates clearing Layers 1–2. High JSD ≠ disqualification; flags need for role redesign discussion.**

For candidates who clear the DQI and UPDD filters, JSD provides the tactical fit evaluation. High JSD relative to the buying team's reference distribution does not necessarily disqualify a player — it flags the need for role redesign or tactical adaptation. High-DQI / high-JSD players are valuable assets in flexible tactical systems but are higher-risk for teams with rigid positional identities. JSD should always be interpreted against the team's specific reference distribution, not as an absolute score.

4. **Layer 4 — Movement Intelligence (ROER): Secondary enrichment. High ROER is a premium signal in run-based systems; low ROER may reflect role design rather than cognitive limitation.**

ROER operates as a secondary enrichment layer for players who clear the first three filters. High ROER among qualified candidates indicates a player who actively connects with intelligent movement — a premium in systems that rely on off-ball runs to break defensive structures. Low ROER in an otherwise strong profile may reflect role design (deep-lying pivots rarely exploit run-created options by design) rather than a cognitive limitation. Use ROER_tier relative to position-stratified benchmarks, not as an absolute score.

16.6 Video Integration — Glass-Box Closure

Every metric traces to inspectable pass events. For any player with UPDD > 0.15, the analyst can query: which specific passes under high pressure showed the largest DQI drop? What option was available vs what they chose? What was the pressure context? This generates a targeted video reel — the number creates the hypothesis, video tests it. This glass-box audit trail is the most important practical operational feature, addressing the organisational challenge of justifying decisions to coaches and executives.

16.7 Decision Justification

The framework explicitly addresses the organisational challenge of justifying major transfer commitments. Traffic-light profiles, radar charts, and quadrant maps provide visually accessible summaries for coaches and executives who are not data-literate. The glass-box audit trail provides the depth required for peer challenge from the analytics team. Both outputs are generated simultaneously from the same pipeline, eliminating the common problem of separate 'simple' and 'technical' outputs that diverge in interpretation.

17. Critical Analysis and Limitations

17.1 Single Competition Scope

Stage 2 dataset: one competition, 50 matches. Cross-competition robustness is untested. Decision quality metrics are sensitive to competitive context. UPDD tests intra-competition pressure responses but cross-league validation remains the priority extension.

17.2 DQI-Completion Correlation

Spearman $\rho \approx 0.502$ is expected. The $\Delta R^2 \geq 0.05$ test directly addresses the orthogonal signal question. Conditional pressure analysis shows the gap widens under pressure — exactly where transfer risk materialises. This framing converts the correlation from a concern to a feature.

17.3 OVS Weight Sensitivity

Fixed-prior weights are not adaptive to different team contexts. Sensitivity analysis confirms ranking robustness across the accepted scenarios. Practical recommendation: run weight sensitivity analysis for each prospective buying club's tactical priorities before advancing a shortlisted player.

17.4 POR Low Variance

Within-position POR variance is constrained by role homogeneity — tactical structure limits progressive passing latitude more than individual preference. POR_z is valid for cross-position profiling but less discriminatory within position groups. This is acknowledged and the metric is positioned accordingly.

17.5 ROER Opportunity Sparsity

Players in systems with limited off-ball movement produce sparse ROER estimates. The tau filter (minimum 5 opportunities) addresses severe cases. The UPDD-ROER coherence finding (Section 7) explains that exclusion for high-UPDD players is structurally meaningful, not a data gap.

17.6 Acknowledged Scope Boundaries

The framework does not measure: reception quality, spatial positioning, ball-carrying decisions, off-ball positioning. These are explicitly excluded by design — focusing on a precisely defined analytical surface allows depth of validation and auditability that broader frameworks cannot achieve.

17.7 UPDD Sample Size Requirements

UPDD requires meaningful exposure in both low and high-pressure contexts. Players who rarely face high pressure (e.g. deep-sitting ball-players in possession-dominant systems) may have unreliable UPDD estimates even with sufficient overall pass counts. The updd_coverage flag in the pipeline identifies players where pressure band sample sizes are insufficient, and these players should not be evaluated on UPDD alone.

18. Future Work

Cross-Competition Validation (Priority)

Do DQI and UPDD scores in one league predict performance outcomes after transfer? Requires outcome data: minutes played, positional influence, resale value. Even without transfer outcomes, multi-league DQI comparison reveals context-sensitivity vs trait-stability.

Tracking — Full Scope Delivery

Lane occlusion modelling (restrict options to field of view via body orientation proxy), pitch control-based IA run validation, continuous spatial pressure replacing band classification. These sharpen rather than redefine the core metrics.

Receiver-Side Profiling

Complementary framework evaluating which players create high-OVS receiving options. High ROER on both passer side and receiver side identifies jointly efficient run-pass combinations.

Longitudinal UPDD Tracking

Track UPDD across consecutive seasons, particularly for players who have already transferred. Does pre-transfer UPDD predict post-transfer performance degradation? This establishes the framework as a genuine predictive tool.

Positional Extension

The current analysis covers all outfield positions but the original framing targeted midfielders. Position-specific parameterisation — separate UPDD thresholds, OVS weights, and JSD reference distributions for different roles — would improve precision for specific scouting briefs. Central midfielders and wide playmakers have structurally different option landscapes and should have calibrated thresholds.

References

- Power et al. (2017). Not All Passes Are Created Equal. KDD.
- Fernández et al. (2021). Decomposing the Immeasurable Sport. SSAC.
- Spearman (2018). Beyond Expected Goals. SSAC.
- McIntosh & Beckmann (2024). Modelling Player Performance After Transfers. Hudl Pro.
- Danesi (2024). The Imposter on the Pitch. Hudl Pro.
- Lee (2024). Capturing Player Impact. Hudl Pro.
- Stats Perform (2025). MA36/MA58/MA60 API Documentation.
- Lin (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory.

Appendix

A. AAS — Alternative Awareness Score

The Alternative Awareness Score (AAS) was introduced in the Stage 1 submission as the off-ball run exploitation metric, designed to answer one precise question:

Core question (AAS)
"When a teammate made a run that genuinely created a passing option, did the ball-carrier choose to use it?"
AAS bridges two data worlds: MA58 (off-ball runs, tracking-derived) and MA36 (pass events). The linkage key is `relatedEvents` — Opta's own record of which pass events occurred during a run window. This makes AAS an observed fact, not an inference.

A.1 Underlying Data Preparation

Before `compute_aas()` is called, two pre-processing functions enrich the raw MA58 run data:

extend_off_ball_runs_with_qualifiers()

This function flattens the MA58 qualifier array from each run event into named columns on the run DataFrame. Key fields extracted include:

- **increasesAvailability — the primary IA flag (boolean)**
- **relatedEvents — a list of MA36 event IDs that co-occurred during the run window**
- **xR (expected receiver probability), xT and xP trajectory values where logged**
- **Spatial flags: passReceived, passTarget, danger_flag, pressure band**

Output shape: runs_qualifiers with the same row count as the raw runs DataFrame, augmented with all parsed qualifier columns.

extend_off_ball_runs_with_initiator_fields()

Adds two identity fields to each run row that are critical for passer-centric attribution:

Term	Definition / Formula	Notes
initiatorPlayerId	The playerId of the passer whose pass triggered or defined the run context. Derived from the first relatedEvent pass found in the MA36 event index.	Missing in 24.6% of runs. Constraint is relaxed: if initiatorPlayerId is absent, the relatedEvents linkage alone is used. Rejection only fires when initiator is known and mismatches the linked passer.

initiatorEventId	The MA36 event ID of the initiating pass. Used as a fallback when initiatorPlayerId is absent — the earliest relatedEvent ID is resolved to a player via an event→player lookup table.	Avoids discarding runs with missing initiator by recovering the passer identity indirectly from the event index.

A.2 compute_aas() — Three-Step Logic

The function operates over the enriched run DataFrame (`runs_qualifiers_initiators`) and the pass events table. It returns two outputs: `player_aas` (player-level aggregation) and `ia_runs` (the full IA run audit trail).

Step 1 — Filter to IA runs
From all 26,672 off-ball runs, retain only those where `increasesAvailability = 1`. This yields 388 runs (1.5%). The Opta Vision model flags these runs as having raised the runner's expected receiver probability by ≥ 0.3 xR to a minimum of 0.8 xR — confirming the runner became a genuinely high-quality option, not merely a moving player. This is Opta's own quality gate, not a subjective threshold applied post-hoc.

Step 2 — Numerator: ia_exploited per passer
For each IA run, the function parses its `relatedEvents` list via `parse_related_events_cell()` to extract MA36 event IDs. Each ID is looked up in a pass-event-to-player index. If the linked pass belongs to a passer, and the `initiatorPlayerId` check passes (or is absent), the passer is credited with one exploitation. The count is broken after the first valid match per run — this prevents a single IA run from inflating a player's numerator if multiple passes appear in `relatedEvents`.

Step 3 — Denominator: ia_total per passer
The denominator is computed using identical relatedEvents → pass linkage logic, but without the exploitation check. Any IA run whose `relatedEvents` contains a pass by passer P counts as one opportunity for P (`ia_total`). Consistency with the numerator is guaranteed because both use the same linkage path — there is no frame-level or time-unit alignment required.

Final formula: `AAS = ia_exploited / ia_total` (clipped to [0, 1], with 1e-9 denominator guard)

A.3 Variable Reference

Term	Definition / Formula	Notes

increasesAvailability	MA58 boolean flag. Runner raised xR by ≥ 0.3 to a minimum of 0.8 during a single possession. Defines the denominator population; without this gate AAS is noisy.	Opta's own quality gate — the threshold is theirs, not ours.
relatedEvents	MA58 qualifier. List of MA36 event IDs co-occurring during the run window. The factual observed link between run and pass — no time-window estimation needed.	Bridges MA58 → MA36. Used identically in numerator and denominator for perfect consistency.
initiatorPlayerId	MA58 field. PlayerId of the ball-carrier at decision time — the player AAS is attributed to. Absent in 24.6% of runs (relaxed constraint: linkage used even when missing).	Reject only if initiator is KNOWN and mismatches the linked passer.
ia_total	Computed. Count of IA runs linked to this passer's passes via relatedEvents. How many genuine off-ball opportunities this player faced.	Denominator of AAS. Players with ia_total < 5 should not be ranked; flag as provisional.
ia_exploited	Computed. Count of IA runs where the passer's pass was the linked event and the initiator check passed.	Numerator of AAS. Cross-checkable against passReceived / passTarget on the run for video audit.
aas	ia_exploited / ia_total ∈ [0, 1]. Rate at which a passer rewards intelligent off-ball movement.	Secondary lens to DQI — confirms whether high-DQI players choose movement-based best options or static geometric ones.
aas_exposure	Categorical: no_exposure / low_exposure / mid / high_exploiter / low_exploiter. Operationalises NaN — tells a scout whether absence of AAS score is a data gap or a meaningful signal.	Prevents misreading NaN as 'average'. Essential for honest shortlisting.

A.4 Fields Added to player_metrics

Following `compute_aas()`, four columns are merged into `player_metrics` on `playerId`:

Column	Type	Population with Data	Interpretation
aas	Float [0, 1]	133 players (43% of 309)	Run exploitation rate. NaN = player had no IA run linked to any of their passes — true zero exposure, not missingness.
ia_total	Integer	133 players	Number of IA run opportunities faced. The exposure denominator. Threshold for credible ranking: ia_total ≥ 5.
ia_exploited	Integer	133 players	Number of those opportunities where the player passed to the running option. Numerator.
aas_exposure	String	All 309 players	Scouting signal layer: no_exposure | low_exposure | mid | high_exploiter | low_exploiter.

The `aas_exposure` column is constructed with the following logic:

- no_exposure — ia_total is NaN: player was never ball-carrier during a tracked IA run
- low_exposure — ia_total < 5: insufficient sample for ratio stability
- high_exploiter — ia_total ≥ 5 AND aas ≥ 0.60: credible movement exploiter
- low_exploiter — ia_total ≥ 5 AND aas ≤ 0.20: consistent tendency to overlook IA runs
- mid — ia_total ≥ 5 AND 0.20 < aas < 0.60: intermediate exploitation tendency

57% NaN in the `aas` column is expected and structurally correct. These players were never observed as ball-carrier when an IA run was being tracked. In a scouting presentation, NaN becomes *"insufficient tracked opportunity"* — a data coverage note, not an analytical failure.

A.5 Validation Suite

Three validation functions are implemented for AAS, designed to mirror the DQI validation philosophy while adapting to the sparse-count reality of ratio metrics.

validate_aas_reliability()

Standard split-half on a ratio metric with tiny counts is statistically meaningless. Instead, bootstrap 95% confidence intervals serve as the reliability proxy — the same principle as split-half, adapted for sparse binary outcomes.

Implementation: For each player with `ia_total ≥ min_ia` (default: 3), a binary array of `[1]*ia_exploited + [0]*(ia_total - ia_exploited)` is constructed. Bootstrap resampling (1,000 draws) over `np.mean` yields a 95% CI. CI width < 0.50 is the reliability threshold.

Output fields:

Term	Definition / Formula	Notes
playerId	Player identifier	
ia_total	Total IA opportunities for this player (n)	Determines bootstrap CI width — larger n yields narrower CI
ia_exploited	Number exploited (k)	
aas	ia_exploited / ia_total	
ci_lo / ci_hi	Bootstrap 95% confidence interval bounds	CI width = ci_hi - ci_lo
ci_width	CI width — primary reliability indicator	< 0.50 = reliable; = 1.0 = effectively uninformative
reliable	Boolean: ci_width < 0.50	Reliability flag for downstream filtering

Empirical result
AAS reliability: 0 / 16 reliable. Median CI width = 1.000.
A CI width of 1.0 means the interval spans the entire [0, 1] range — which arises when `ia_total` is 3 or 4 with binary outcomes. This is a sample size problem, not a methodological failure. With ia_total = 3, a 95% bootstrap CI on a proportion will always be near [0, 1]. The honest statement is: AAS is computable but not yet estimable with precision at this scale. A minimum of ia_total ≥ 10 is required to achieve CI width < 0.50.

validate_aas_cross_match_stability()

Mirrors the DQI `compute_rank_stability()` approach: compute AAS per match for each player, then assess Spearman rank correlation across match pairs. This tests whether a player's relative exploitation tendency is consistent across different opponents and contexts.

Implementation: For each match, a per-passer AAS is computed from the IA runs in that match using the same relatedEvents → pass linkage. Only players with `ia_total ≥ min_ia` (default: 2) within a match qualify. Spearman r is computed for all match pairs where ≥ 5 players appear in both matches. Target: mean r > 0.40 (lower than DQI due to sparsity).

Output fields:

Term	Definition / Formula	Notes
n_match_pairs	Number of match pairs with ≥ 5 common players and sufficient IA exposure	
mean_spearman_r	Mean Spearman correlation across all qualifying match pairs	Target > 0.40
median_spearman_r	Median Spearman correlation across all qualifying match pairs	
target_met	Boolean: mean_spearman_r > 0.40	

Empirical result
Cross-match stability: 0 pairs assessed, r = NaN.
Zero qualifying pairs means no player appeared in two or more matches with ia_total ≥ 2 in the same match. With 388 IA runs across 50 matches — approximately 8 per match — different passers are involved in each match. This is the sparsity problem in its starkest form. The correct framing: cross-match stability cannot be assessed at current IA run density. This confirms the Stage 2 design requirement that AAS gains discriminatory power only with substantially larger observation windows.

validate_aas_dqi_orthogonality()

This is the most important of the three AAS validation functions from a competition perspective — it tests whether AAS adds information beyond DQI, and whether the two metrics capture distinct cognitive dimensions of passing intelligence.

Implementation: Two tests are run on players with `ia_total ≥ min_ia` (default: 3) and non-null values for both `aas` and `mean_dqi_regret`:

- Spearman correlation between AAS and DQI_regret — should be low in magnitude (|r| < 0.40 = orthogonal)
- Mann-Whitney U test comparing mean DQI and UPDD across aas_exposure groups — tests whether high-AAS players are meaningfully different on quality and pressure dimensions

Output fields:

Term	Definition / Formula	Notes
aas_dqi_spearman_r	Spearman r between AAS and mean_dqi_regret over eligible players	Target: $\|r\| < 0.40$
aas_dqi_p	p-value of the Spearman test	Non-significant expected at n = 16 due to small eligible sample
orthogonal	Boolean: $\|aas\_dqi\_spearman\_r\| < 0.40$	Primary orthogonality flag

Empirical result — strongest AAS finding
AAS vs DQI Spearman $r = -0.179$, $p = 0.507$. Orthogonality CONFIRMED.
The p-value is non-significant, expected at n = 16. The direction is theoretically important: a slight negative tendency suggests that players who optimise among visible static options are not necessarily the same players who exploit off-ball movement. These are genuinely distinct cognitive dimensions — option selection quality versus movement awareness. Report as directional evidence with appropriate uncertainty, not as a strong causal claim.

A.6 Honest Scale Diagnosis

The three validation results combine into a coherent and honest diagnosis. AAS is not a failed metric — it is a metric that requires more data than this 50-match sample provides.

Diagnosis	Root Cause	Consequence	Honest Framing
23.9% of players have AAS = 1.0	ia_total of 1–2 creates a tautology: the relatedEvents link defines both numerator and denominator, so ia_exploited = ia_total by construction	AAS = 1.0 is a small-sample artefact, not elite performance	Flag as provisional; exclude from ranking
CI width = 1.000 for all 16 eligible players	ia_total < 10 for all players — binomial CI spans [0, 1] when n is 3–4	Reliability cannot be confirmed at this scale	Report CI alongside ratio; require ia_total ≥ 10 for credible estimates
Zero cross-match pairs	388 IA runs across 309 players = 1.25 opportunities per player on average; different passers each match	Temporal stability is untestable	AAS needs ≥ 200 IA runs per player across sessions to be rankable
r(AAS, DQI) = −0.179	Genuine cognitive independence, not confounded by completion tendency	AAS does capture a distinct dimension from DQI	The strongest finding; report as directional with n = 16 caveat

The scouting value at this scale is in the tails, not the average. A central midfielder with ia_total = 12 and AAS = 0.17 is a genuine red flag — repeatedly ignoring line-breaking run options. A player with ia_total = 9 and AAS = 0.78 is a credible off-ball exploiter worth further investigation. The middle of the distribution is not interpretable.

B. From AAS to ROER — The Transition

B.1 Why AAS Was Superseded

The AAS diagnosis in Section A.6 establishes that at 50-match scale, AAS cannot function as a ranking metric due to the tautology problem and extreme sparsity. However, the conceptual claim is sound and novel: a passer's willingness to reward intelligent movement is a distinct cognitive dimension from option selection quality (DQI). The challenge was purely one of data density.

The solution was architectural rather than parametric: instead of tuning AAS thresholds, the movement awareness question was re-operationalised using a data source that exists at MA36 density — passOption.player[i].activeRun.

The core architectural shift
AAS: Bridges MA58 → MA36 via relatedEvents. 388 qualifying events. 1.25 opportunities per player. Untestable for stability.
ROER: Uses passOption.activeRun natively within MA36. Up to 469K option-level observations across 51K pass events. Median 6 opportunities per player, 373 players with signal.
The ia_runs data from MA58 is not abandoned — it becomes a quality stamp on top of ROER's dense denominator. MA58 validates that the highest-ROER passes correspond to objectively run-created opportunities, not incidental movement.

Aspect	AAS
Data source	MA58 relatedEvents → 388 linked events
Players with signal	141 (57% NaN in player_metrics)
Typical opportunities per player	1–3 (mean 1.25)
AAS = 1.0 tautology problem	Yes — relatedEvents defines both numerator and denominator
MA58 connection	Primary (direct)
Ranking discriminability at 50 matches	None — all values cluster at 1.0
Cross-match stability testable	No — zero qualifying match pairs

Aspect	ROER
Data source	MA36 passOption.activeRun → up to 469K option observations
Players with signal	373 (from 309 in player_metrics population)

Typical opportunities per player	Median 6, mean 7.3, max 31
Tautology problem	No — running option must be strictly better than chosen (OVS gap ≥ tau)
MA58 connection	Via has_ia_run enrichment flag (quality validation layer)
Ranking discriminability at 50 matches	Sufficient — genuine spread, statistically testable
Cross-match stability testable	Yes — sufficient player-match co-occurrences

B.2 compute_roer() — Implementation

ROER is implemented as an extension of the existing DQI pipeline, deliberately reusing its core functions:

Code reuse from DQI pipeline
`add_progressive_gain()`, `compute_ovs()`, `aggregate_options_to_event()`, `identify_chosen_option_ovs()` — called with identical parameters. No duplication. The DQI pipeline already computes `n_active_run` (count of running options per event) as part of `aggregate_options_to_event()`, so that is available for free.

Opportunity definition

A pass constitutes a run opportunity if both conditions hold:

- n_active_run ≥ 1 — at least one available receiver was actively running at decision time (from passOption.activeRun)
- best_running_ovs > chosen_ovs + ROER_OPPORTUNITY_TAU — the best running option was strictly better than what was chosen, by at least tau = 0.02 OVS units. This excludes trivial dilemmas where the running option was barely distinguishable.

Note: `best_running_ovs` is computed from the running-only option subset. This is distinct from `best_ovs` (all options), which is already available from the DQI pipeline. One additional aggregation is needed.

Exploitation definition

A run opportunity is exploited if the chosen receiver was actively running at the time of the pass. This comes from `pass_targets.activeRun` — the *actual chosen receiver* with their observed movement state — rather than from the options table, which covers all available candidates.

`chosen_was_running_rate` is an unconditional bonus column: the raw proportion of passes where the chosen receiver was running, without requiring a 'better run existed' filter. This distinguishes stylistic preference for movement from disciplined exploitation of superior movement options.

MA58 enrichment

For each opportunity pass, the function checks whether an IA-confirmed run from MA58 was linked via `relatedEvents`. This produces two additional player-level counts: `roer_ia_opportunities` and `roer_ia_exploited`. These enable a two-tier narrative:

- ROER — broad movement-reward tendency (MA36-native, dense, rankable for 373 players)
- ROER on IA-confirmed runs — high-quality exploitation specifically of Opta-flagged high-value movements (MA58-grounded, sparse but precisely validated, 54 players)

B.3 Fields Added to player_metrics

Column	Source	Interpretation
roer	Computed: roer_exploited / roer_opportunities	Run-Option Exploitation Rate ∈ [0, 1]. Core movement awareness metric.
roer_opportunities	Count of passes with n_active_run ≥ 1 AND best_running_ovs > chosen_ovs + tau	Exposure denominator. Meaningful tier: ≥ 5 = provisional, ≥ 10 = sufficient.
roer_exploited	Count of opportunity passes where chosen receiver had activeRun = True	Exploitation numerator.
roer_ia_opportunities	Subset of roer_opportunities with an MA58 IA-confirmed run linked via relatedEvents	Quality-stamped denominator — smaller but Opta-validated.
roer_ia_exploited	Subset of roer_exploited from IA-confirmed opportunity passes	Quality-stamped numerator.
chosen_was_running_rate	Passes where passTarget.activeRun = True / total passes (unconditional)	Movement style indicator. Compare against ROER to distinguish style from decision quality.
roer_exposure	Categorical: insufficient / provisional / sufficient	Mirrors aas_exposure logic. Operationalises NaN for scouting use.
roer_tier	Categorical: high_exploiter / mid / low_exploiter / insufficient	Position-stratified tier label for shortlist and recruitment dashboard.

The `roer_exposure` tiers are assigned as: `roer_opportunities ≥ 10` → sufficient; `≥ 5` → provisional; `< 5` → insufficient.

B.4 ROER Critical Analysis — Correlation Audit

Four correlation findings required interrogation before ROER could be claimed as adding complementary information to DQI:

Raw correlations (unadjusted)

Pair	Spearman r	p-value	Interpretation
ROER vs DQI_regret	−0.328	< 0.0001	Negative and significant — apparently problematic
ROER vs completion %	−0.348	< 0.0001	Negative and significant — apparently problematic
ROER vs chosen_was_running_rate	+0.467	< 0.001	Moderate positive — ROER and unconditional running style share 22% of variance
ROER vs mean chosen xT	+0.406	< 0.001	Players who habitually choose higher-threat options also exploit running options more

The negative correlation with DQI was the most concerning finding. Two explanations were possible:

- Explanation A (genuine orthogonality with interesting structure): High-DQI players optimise for the highest static OVS option, which is often a safe, well-positioned receiver. Running receivers carry higher xT but lower xP — a riskier profile. High-DQI players are not ignoring runs; they are correctly preferring the safer option given their risk preference. ROER measures a distinct dimension: willingness to accept execution risk for movement-based threat.
- Explanation B (confound): Running receivers are harder to complete to (lower xP). Players in defensive or recycling roles have fewer running options and higher completion rates. ROER might be measuring positional context rather than cognitive choice.

Decisive result: partial correlation

Partial correlation — primary orthogonality finding
Partial r(ROER, DQI_regret | completion%) = −0.052, p = 0.487
After controlling for pass completion tendency, ROER and DQI are statistically independent. The apparent negative correlation between ROER and DQI was entirely explained by their shared relationship with completion%. Once the confounder is removed, the two metrics carry zero redundant information.
Partial r(ROER, DQI_regret | chosen_was_running_rate) = +0.065, p = 0.380
After controlling for raw movement tendency, ROER adds nothing to explain DQI variance, and vice versa. ROER is not a redundant encoding of running style — it captures the conditioned exploitation decision. The framework's two-dimension claim is empirically supported.
Explanation A is confirmed definitively. ROER and DQI are orthogonal cognitive dimensions. The completion% relationship is a confounder, not a structural link.

Positional gradient

The positional gradient confirms Explanation A and reveals expected role differences:

Position Group	Mean ROER	Interpretation
Defenders	0.226	Fewest forward running options by design; ROER ≥ 0.35 exceptional for this role
Central Midfielders	0.330	Moderate — many options, default to safest high-OVS choice
Left / Right Midfielders	0.316 – 0.399	Wide positions face more movement opportunities; higher exploitation expected
Forwards	0.380 – 0.452	Primary mechanism of value creation is movement exploitation

These are features, not confounds. A defender with ROER = 0.35 is exceptional. A forward with ROER = 0.25 is a red flag. ROER must be reported within position groups, exactly as DQI is stratified. Position-group z-scoring (`roer_z`) is the appropriate cross-position comparison tool.

B.5 The Tau Filter — Coherence with UPDD

The `ROER_OPPORTUNITY_TAU = 0.02` filter (running option must exceed chosen_ovs by ≥ 0.02 OVS units) excludes 7 players from the ROER-ranked population who had opportunities under the unfiltered definition. Analysis of these 7 players produced a critical joint coherence finding:

UPDD–ROER joint coherence
Mean UPDD: excluded players = 0.192; included players = 0.053 (Mann-Whitney $p = 0.018$)
The 7 excluded players have UPDD nearly four times the included mean, and three of the four with available UPDD scores sit above the pre-registered risk threshold of 0.15.
The mechanistic explanation: high-UPDD players make worse decisions under pressure. Under pressure, the option landscape collapses — fewer viable options, lower OVS spread, smaller gaps between options. This mechanically reduces the OVS gap between any running option and the chosen option, making it hard to exceed tau = 0.02. The tau is not randomly excluding these players — it is excluding them because their decision contexts are genuinely lower-OVS-spread situations where a clear movement dilemma rarely arises.
Tau exclusion reflects pressure-induced option collapse, not data poverty. ROER exclusion is correct and informative: the metric is undefined for these players not because of coverage gaps, but because their pressure-fragile decision contexts do not generate the structural conditions ROER requires.

For all 7 players individually, mean_ovs_gap is negative or negligible (range: −0.063 to +0.001), confirming the chosen option was on average better than the best running option available. Tau exclusion is justified for all 7.

B.6 AAS Directionality in ROER

The conceptual architecture of AAS survives its data density failure and becomes embedded as directionality within ROER. The two-tier structure gives AAS its role:

Tier	Data Source	Players	Role in Framework
ROER (broad)	MA36 passOption.activeRun	373 players	Ranking metric. Measures movement exploitation tendency at scale. The working implementation of what AAS was designed to do.
ROER on IA-confirmed runs	ROER denominator cross-referenced with MA58 relatedEvents IA flags	54 players	Quality validation layer. Confirms the highest-ROER passes correspond to Opta-flagged high-value movements. Bridges MA58 into the framework without relying on it as a primary signal.
AAS (legacy)	MA58 relatedEvents exclusively	133 players (57% NaN)	Directional concept. Retained as a theoretical construct; not used for player ranking at this scale.

Competition narrative: one sentence
"ROER operationalises movement awareness at MA36 density; the MA58 increasesAvailability flag validates that the highest-ROER passes correspond to objectively run-created opportunities, not incidental movement. AAS, as originally defined, provides the directional logic that motivates the metric — ROER is its scalable implementation."

C. Supplementary Sections

The following sections extend the corresponding sections in the main Synthesis Report with content specific to AAS, ROER, and their interaction.

C.1 Limitations — AAS and ROER

AAS limitations

- Tautology at small n: when relatedEvents defines both denominator and numerator, players with ia_total ≤ 2 will always produce AAS = 1.0. This is structural, not methodological.
- relatedEvents is conservative: Opta logs only runs where the event ID appears explicitly. Many valid spatial and temporal co-occurrences are not logged, so ia_total undercounts true exposure.
- 24.6% of runs have missing initiatorPlayerId: the relaxed constraint (use linkage when initiator is absent) introduces a small number of mis-attributed pass credits. The initiator check fires only when initiatorPlayerId is known and present.
- AAS cannot be validated for cross-match stability at 50-match scale: with 388 IA runs across 309 passers, no player accumulates enough per-match IA exposure to appear in two qualifying match windows. Minimum scale for stability testing: approximately 200 matches.
- AAS = 1.0 for 23.9% of players is a data artefact: these players had exactly one IA opportunity and exploited it. They cannot be distinguished from genuinely elite exploiters.

ROER limitations

- r(ROER, chosen_was_running_rate) = 0.467: ROER and unconditional movement style share 22% of variance. Always report both, clearly distinguished. chosen_was_running_rate measures style preference; ROER measures disciplined exploitation of superior options.
- 4,531 NaN chosen_ovs (8.8% of passes): these are Tier 3 events from identify_chosen_option_ovs() — no passTarget match and no proximity match. They are excluded from ROER opportunities entirely. ROER is computed on 91% of passes; document this coverage figure explicitly.
- Coverage trade-off from tau: applying ROER_OPPORTUNITY_TAU = 0.02 reduces the player set from 228 (≥5 opportunities without tau) to 183 (with tau). 7 players lose all opportunities. All 7 exclusions are structurally justified by elevated UPDD and negative mean_ovs_gap.
- IA-confirmed ROER is sparse: only 54 players have both ROER and an IA-confirmed opportunity. Report this subset as a validation confirmation, not as a separate ranking metric.
- chosen_was_running_rate is unconditional: it does not require a 'better run existed' filter, so a high value may reflect tactical system (teammates who run frequently) rather than individual choice.

C.2 Validation — AAS and ROER Summary

Validation Check	Metric	Result	Outcome
Bootstrap CI reliability (proxy for split-half)	AAS	0 / 16 reliable; median CI width = 1.000	FAIL — expected at this scale; documents scale requirement
Cross-match Spearman stability	AAS	0 qualifying pairs; r = NaN	FAIL — expected; confirms ≥200 matches required

DQI orthogonality (Spearman)	AAS	r = −0.179, p = 0.507; |r| < 0.40 CONFIRMED	PASS — directional evidence of distinct cognitive dimension
Positional gradient (Kruskal-Wallis)	ROER	Defender < CM < WM < Forward; gradient significant	PASS — expected role differences, not confounds
DQI orthogonality (raw Spearman)	ROER	r = −0.328, p < 0.0001 — apparently correlated	INVESTIGATE — resolved by partial correlation
DQI orthogonality (partial r | completion%)	ROER	r = −0.052, p = 0.487 — statistically independent	PASS — completion% was the confounder, not structure
DQI orthogonality (partial r | running_rate)	ROER	r = +0.065, p = 0.380 — statistically independent	PASS — confirms ROER is not redundant with running style
Tau exclusion coherence (UPDD comparison)	ROER	Excluded mean UPDD = 0.192 vs included 0.053, p = 0.018	PASS — exclusion is structurally justified, not arbitrary
Sample adequacy at ≥5 opportunities	ROER	183 players (59% of 309)	PASS — sufficient for comparison and ranking
Sample adequacy at ≥10 opportunities	ROER	71 players — reliable ranking tier	PARTIAL — sufficient for shortlist, limited for population claims

C.3 Key Results — AAS and ROER

AAS

- 133 players (43% of player_metrics population) have any AAS value; 176 have no_exposure
- Mean AAS across players with ia_total ≥ 1: approximately 0.87 — artificially inflated by tautology; not a population-level finding
- AAS is computable and the linkage mechanism works correctly; the metric is scale-limited, not design-flawed
- Orthogonality with DQI confirmed (r = −0.179): even with limited data, the conceptual claim of distinct dimensions survives

ROER

- 373 players have at least one ROER opportunity; mean ROER = 0.306 (31.2% exploitation rate)
- 183 players with ≥5 opportunities (provisional tier); 71 with ≥10 (sufficient tier)
- Positional gradient confirmed: Defenders (0.226) → Central Midfielders (0.330) → Wide Midfielders (0.316–0.399) → Forwards (0.380–0.452)
- Partial r(ROER, DQI | completion%) = −0.052: ROER and DQI are statistically independent after removing completion% confounder
- ROER vs chosen_was_running_rate r = 0.467: the two are related but distinct; 78% of variance is independent
- ROER vs mean chosen xT r = 0.406: risk appetite and movement exploitation are correlated — both reflect progressive decision style
- 54 players have IA-confirmed ROER data: MA58 quality stamp available for video-linkable audit

C.4 Recruitment Applications — AAS and ROER

AAS exposure in scouting workflow

At current scale, AAS functions as a coverage diagnostic, not a ranking signal. The `aas_exposure` column is the operationally useful output. In shortlist presentation:

- no_exposure → 'No tracked off-ball run opportunity in sample' — not a negative flag; simply no data
- low_exposure (ia_total < 5) → 'Insufficient sample for ratio stability' — report ia_total alongside the ratio
- high_exploiter (ia_total ≥ 5, AAS ≥ 0.60) → Credible movement exploiter — flag for further investigation
- low_exploiter (ia_total ≥ 5, AAS ≤ 0.20) → Consistent tendency to overlook availability-increasing runs — genuine red flag

ROER in the recruitment framework

ROER operates as the fourth layer in the recruitment filter hierarchy, downstream of DQI, UPDD, and JSD:

- **Layer 4 (Movement Intelligence): ROER ≥ position-group p75 = high_exploiter tier — confirms player rewards creative teammates. Low ROER in a forward (ROER < 0.25) is a recruitment risk flag even with high DQI.**

The five canonical player profiles from the AAS × DQI interaction analysis remain directly applicable to ROER × DQI, with higher statistical confidence:

Profile	DQI	UPDD	ROER	JSD	Verdict
Prime target	High ≥ 0.78	Low ≤ 0.10	High ≥ p75	Low vs buying club	Selects best option, pressure-robust, exploits movement, tactically compatible
Hidden value	High	Low	High	High	Elite decision-maker but tactical misfit — worth it if club adapts system
System product	High	High	Low / no exposure	Low	Good numbers in protective system; ignores off-ball movement; will struggle in high-press environment
Off-ball aware but fragile	Mid	High	High	Low	Sees and uses runs but collapses under pressure — development candidate, not first-team signing
Safe recycler	Mid	Low	Low	Low	Reliable but conservative — avoids run-exploitation passes; limits vertical threat
Red flag	Any	High ≥ 0.15	Low	High	Structurally fragile + stylistically misaligned + movement-blind — avoid at premium fee