TrackTempo AI – RPR Signal Analysis Log

2025-03-27 | Author: J. D. Farrell

Objective

To clarify the usage, trustworthiness, and post-race behavior of the Racing Post Rating (RPR) in modeling workflows. Based on direct observation, controlled scraping, and forum-sourced domain insights.

Summary of Updated Findings

- **RPRs are assigned before a race** based on known conditions (going, field size, trainer form).
- **Minor changes may occur** post-race, especially for:
  - Late jockey changes
  - Non-runner adjustments
  - Misreported or unannounced going shifts

- **Changes are typically small** and do not shift the runner rankings substantially.
- **RPRs seen in result pages** may be slightly adjusted vs those on the morning of the race.
- Post-forum observation confirms: *"...changes are not that common and even if there are, there is not much change to the ratings positions."*

Current Strategy

- Continue scraping racecards **before the race** to capture pre-official RPRs.
- Label data sources where possible: `rpr_source = "pre_race_card"`
- Do not train on **historically scraped result-time RPRs** unless pre-race fidelity is confirmed.
- Accept that **slight differences** post-race do not necessarily mean leakage — they're part of the update cycle.

RPR vs OR Recap

- **RPR** – Analyst-assigned, pre-race but updated slightly after race based on final conditions.
- **OR** – Assigned by BHA; changes weekly; what you see pre-race is stable and safe for modeling.

Recommended Practices

- Archive scrapes on the morning of the race (before ~11:30am UK).
- Track data provenance in modeling datasets (e.g., rpr_source, scrape_time).
- Test downstream impact of **marginal RPR updates** to assess real leakage risk.

Generated by TrackTempo AI | Race Intelligence Modeling System | J. D. Farrell

RPR Scraping Timing & Lifecycle

To maximize feature integrity and predictive value, it is essential to scrape RPRs at the optimal time on race day.

Daily RPR Signal Timeline

Time	Modeling Safety	Reason
5am–7am	Safe	Cards are freshly published. Ratings reflect overnight expert form review with stable assumptions.
7am–9am	Safe	Likely contains slight refinements; still purely pre-race and analytically sound.
9am–11:30am	⚠ Caution	Some late non-runners or going updates may trigger editorial tweaks. Minimal risk, but growing.
After 11:30am	Risky	Live info starts influencing analyst actions. Ratings may be shifted based on field volatility or bets.
Post-Race	Do not use	Backfilled RPRs encode hindsight. These are result-driven and leak-prone.

Recommendation

- Scrape all racecards between **5:00am and 9:00am** on race day.
- Store `scrape_time` or `rpr_source` flags for data audits.
- Ignore post-result RPRs for model training — use only for validation or backtesting experiments.

This guidance is based on direct analysis, forum-sourced insights, and professional modeling best practices.