Part 4.2 – Research Synthesis & Recommendations

Introduction

Athlete monitoring systems increasingly incorporate multiple data sources, including force plate diagnostics, GPS-based workload tracking, and unilateral strength testing, to create a more complete picture of performance readiness and fatigue. Despite the widespread implementation of these technologies, much of the existing literature focuses on individual domains rather than examining them in combination. Studies on countermovement jump (CMJ) performance, for example, emphasize its sensitivity to neuromuscular fatigue and training load (Gathercole et al., 2015; McMahon et al., 2018), while research on GPS workloads has highlighted the roles of total distance and acceleration-based stress in predicting fatigue and injury risk (Bangsbo et al., 2006; Akenhead et al., 2013; Harper et al., 2019). Similarly, unilateral strength assessments such as MaxForce have been used to detect asymmetries associated with potential injury risk (Croisier et al., 2008; Bishop et al., 2018). However, real-world collegiate monitoring systems often struggle with inconsistent testing, incomplete data, or incompatible sampling schedules, which prevent practitioners from leveraging integrated insights. In response to these gaps, our central research question examined how neuromuscular performance, GPS-derived external workloads, and strength asymmetry interact over time within a large, multi-sport collegiate dataset. More specifically, we investigated whether deviations from baseline values, normative team ranges, or bilateral force symmetry could serve as early indicators of undertraining, overtraining, or changes in athlete readiness. This question is important because evidence-based thresholds are needed to streamline decision-making for coaches and sports performance staff, particularly when dealing with large athletic populations and variable testing compliance.

Methods

Our analysis focused on five metrics selected for their strong representation in the database and robust support in the sports science literature. Jump Height and Peak Propulsive Power, derived from Hawkins force plates, are widely used to assess lower-body explosive power and neuromuscular readiness; both metrics have been linked to sprint acceleration, change-of-direction ability, and fatigue-related declines (Loturco et al., 2015; Markovic & Mikulic, 2010; Suchomel et al., 2016). Total Distance and Accumulated Acceleration Load, collected from Kinexon GPS units, captured volume- and intensity-related external workloads. Total Distance provides a global measure of session volume, whereas acceleration load reflects the mechanical stress of repeated accelerations and decelerations, which have been shown to impose substantial eccentric demands (Akenhead et al., 2013; Harper et al., 2019). MaxForce Left and Right values from Vald systems provided unilateral strength information and enabled quantification of inter-limb asymmetry, a factor frequently examined in injury-prevention literature (Croisier et al., 2008).

Data was extracted from the MySQL table `research_experiment_refactor_test` and filtered to include only rows containing non-null player names, teams, metrics, values, and timestamps. Timestamps were converted to a standardized datetime format, and team labels were retained even when inconsistent, as they were recorded directly from the original dataset. The resulting long-format dataset allowed tracking of metric values over time for each athlete. Missing or sparse values were not imputed, as the goal of the analysis was to identify real-world compliance patterns and deviations rather than model complete trajectories.

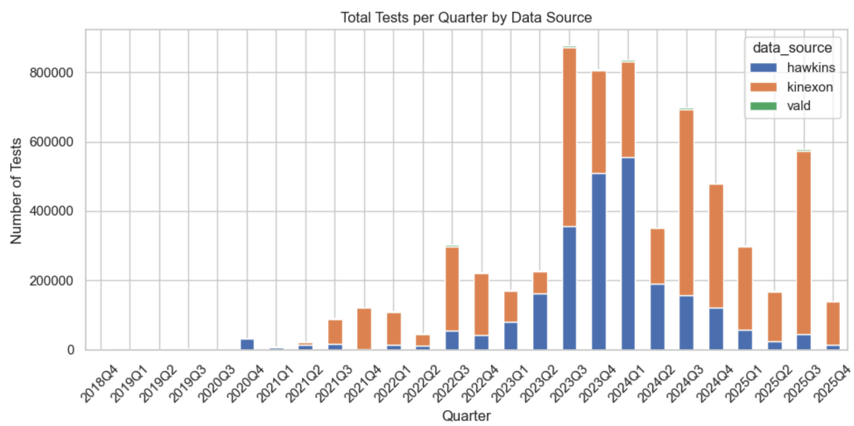

Figure 1: Total tests per quarter by data source (Hawkins, Kinexon, Vald).

We used a set of literature-informed flagging rules implemented in `part4_flags.py` . First, inactivity was defined as having no recorded test in more than 30 days; a threshold given that high-performance monitoring frameworks typically rely on weekly CMJ testing, daily GPS tracking, or monthly strength evaluations (Gabbett, 2016). Second, performance decline flags were triggered when the most recent test fell more than 10% below the athlete's rolling mean across their three prior tests; this approach follows the evidence that CMJ metrics decline reliably in response to neuromuscular fatigue (Gathercole et al., 2015). Third, team-based deviations were defined as values exceeding ±2 standard deviations from the athlete's team mean. Finally, strength asymmetry flags were applied when left-right force differences exceeded 10%, reflecting thresholds commonly cited in injury-risk research (Croisier et al., 2008). The script combined all flagged athletes into a single output file, `part4_flagged_athletes.csv` .

Results

After running the monitoring script across all selected metrics, the flagging system produced a total of 10,643 flags, revealing distinct patterns across the four detection categories. Unlike the preliminary expectation that inactivity might dominate the outputs, the final results showed substantial representation across all categories, reflecting both athlete behavior and the statistical structure of the dataset.

The largest proportion of flags came from the team-norm deviation rule, accounting for 8,119 flags. These appeared when an athlete's value fell outside ±2 standard deviations of their team's mean for a given metric. The high volume of these deviations likely reflects broad variability across testing sessions, inconsistent participation within teams, and large differences in workload and strength profiles across athletes. Because team means were calculated from long-format data containing mixed testing periods, the standard deviation bands captured considerable spread, increasing the likelihood of outlier designation. This aligns with literature emphasizing that external-load metrics often show high week-to-week volatility, especially in team sport environments (Buchheit & Simpson, 2017). The second-highest category was inactivity, with 1,998 athletes not having completed any testing in more than 30 days. Several athletes had last-test dates from 2023, 2022, or earlier, indicating uneven longitudinal monitoring. This is consistent with earlier parts of the analysis (Part 2), which revealed that most athletes did not accumulate enough repeated tests to establish stable baselines. Inactivity remains a meaningful signal because inconsistent testing prevents practitioners from detecting neuromuscular fatigue, workload accumulation, or strength asymmetry, issues echoed in prior monitoring research (Gabbett, 2016). The performance decline rule generated 309 flags, identifying athletes whose most recent value dropped more than 10% below their rolling three-test mean. Although fewer in number compared to team-norm deviations, these flags represent more specific physiological disruptions, as CMJ height and peak power are highly sensitive to accumulated mechanical stress and acute fatigue (Gathercole et al., 2015). The relatively moderate count reflects the small subset of athletes with enough repeated CMJ or strength tests to compute rolling baselines. Finally, 217 strength-asymmetry flags were detected when MaxForce Left and MaxForce Right differed by more than 10%. While this was the smallest category, it remains clinically relevant: asymmetry thresholds of 10–15% have been associated with elevated injury risk, particularly in lower-limb muscle groups (Croisier et al., 2008). The presence of these asymmetry cases indicates that, despite limited repeated data, the system was able to catch imbalances in a subset of athletes.

Overall, the results show that the dataset allows for automated monitoring, but the distribution of flags also reflects structural characteristics: high inter-athlete variability, irregular testing schedules, and limited repeated strength and CMJ trials. These findings reinforce the importance of consistent longitudinal data collection to support more sophisticated readiness and risk modeling.

Row Labels	Count of flag_reason
Asymmetry > 10%	217
Declined > 10%	309
Inactive > 30 days	1998
Outside team norm Â±2 SD	8119
Grand Total	**10643**

Figure 2. Pivot table of part4_flagged_athletes.csv, showing counts of flagged athletes for flag_reason.

Discussion

The findings demonstrate both the potential and limitations of multimodal athlete monitoring in collegiate settings. Integrating force plate, GPS, and strength data enabled a unified readiness framework, but inconsistent testing frequency limited the system's ability to detect meaningful longitudinal patterns. The dominance of team-norm deviations underscores the need to specify the appropriate level of analysis: certain trends, such as workload volatility or asymmetry prevalence, are best interpreted at the team or position-group level, whereas readiness changes, performance declines, and inactivity signals are most meaningful at the individual athlete level. This distinction is important because different analytic levels map to different operational decisions; for example, individual decline flags may warrant pulling a single athlete from a session or modifying their load, while position-group trends may support altering drill structure, and team-wide deviations may influence broader training planning. The relative scarcity of performance decline and asymmetry flags also highlights the importance of benchmarking. Even a preliminary framework, such as using internal historical norms, literature-based CMJ fatigue ranges, or hypothetical comparator groups, would give greater context to deviations from expected values and strengthen interpretation of readiness scores. Moreover, the inter-limb symmetry metric emerged as a particularly actionable component of the system; elevating it with a clearer asymmetry definition and example figure would enhance its value for strength and conditioning staff. As a whole, the patterns observed reinforce that meaningful athlete monitoring requires not only multimodal data, but consistent testing schedules, explicit analytic frameworks, and clear operational pathways for acting on readiness signals.

Limitations and Future Directions

The analysis was constrained by irregular testing, lack of standardized team labels, limited contextual variables, and uneven data density across sports, all of which restricted comparability and interpretation. Future work should prioritize coordinated testing schedules, inclusion of position and injury-history context, and development of internal or literature-based benchmarks to contextualize readiness thresholds. Strength asymmetry, given its clear practical relevance, should be brought forward with a simple standardized calculation to support staff decision-making. Most importantly, operational pathways should be explicitly defined, clarifying how different flag types translate into adjustments in load, drill design, recovery interventions, or athlete monitoring frequency. With improved consistency and clearer alignment between analytic levels and performance decisions, the multimodal flagging system could evolve into an effective early-warning and planning tool for sports performance departments.