

Football Injury Analysis - Predicting Long-Term Durability from Early Career Load.”

In recent months, many football fans and analysts have expressed concern that **Lamine Yamal’s extremely high match load at a young age** could lead to frequent injuries in the future.

To explore whether early overuse truly predicts later injury risk, I conducted a data-driven analysis titled **“Football Injury Analysis - Predicting Long-Term Durability from Early Career Load.”**

The study focuses on **15 elite wingers who were consistently active in the 21st century**.

The central question is:

“Does playing too many matches before turning 22 increase the risk of injuries and missed days later in a player’s career?”

All data was **manually collected from Transfermarkt**, considering **only club-level appearances** (no national teams), and reflects statistics **up to 23 October 2023**.

The dataset includes the following **key variables**:

- Matches before 22 and after 22
- Number of injuries (before / after 22)
- Missed days due to injuries (before / after 22)
- Position (Left Winger / Right Winger)
- Average injury duration and total career matches

Each player’s data was calculated manually, then cleaned and visualized in Python using **pandas, matplotlib, and seaborn**.

The analysis computed ratios such as **injury growth factor** and **missed-days growth factor**, and examined overall relationships through a **correlation matrix** to uncover possible links between early workload and later durability.

Note: All data was manually collected and may contain minor rounding or transcription inaccuracies, though overall trends remain valid.

Visual 1 . Career Match Distribution: Before vs After Age 22

What it shows:

This stacked bar chart compares how many club matches each winger played before turning 22 versus after 22. The **blue section** represents early-career workload, while the **yellow section** shows the remainder of their career. Each player's total bar height reflects their overall match volume, highlighting who was heavily used at a young age and who accumulated most of their games later.

What it means:

The chart shows clear contrasts in development pace. For example, **Neymar** and **Eden Hazard** had unusually high early match counts, while **Cristiano Ronaldo** and **Arjen Robben** built their totals more gradually.

This pattern indicates that some wingers faced **very intense workloads before 22**, which could later relate to higher injury growth or missed-day counts. Visual 1 sets the stage for the following analyses that explore whether that early exposure correlates with long-term durability.

Visual 2. Injury Growth Comparison: Before vs After Age 22

What it shows:

This bar chart compares the total number of injuries each winger experienced **before** and **after** turning 22.

The **pink sections** represent injuries sustained before 22, while the **black sections** show those that occurred later in their careers. The visualization makes it clear how injury frequency tends to rise as players age and accumulate physical load.

What it means:

The results highlight a sharp contrast between early-career and post-22 injury rates.

Players such as **Frank Ribéry**, **Arjen Robben**, and **Eden Hazard** show exceptionally high numbers of injuries after 22, confirming their reputations for repeated setbacks. Meanwhile, **Cristiano Ronaldo** and **Mohamed Salah** stand out for their durability and minimal growth in injury count.

Overall, this visual reinforces one of the study's main findings that **most top wingers experience a significant surge in injuries after the age of 22**, suggesting that accumulated early workload and intensity may play a role in long-term physical wear.

Visual 3. Workload vs Durability: Early Matches vs Later Missed Days

What it shows:

This scatter plot explores the relationship between early career workload and later physical durability.

Each point represents one winger, where:

- **X-axis:** Matches played before turning 22
- **Y-axis:** Missed days due to injuries after 22
- **Dot size:** Number of injuries after 22
- **Color:** Player's position (Right Winger in black, Left Winger in pink)

Player names are labeled to show how their early exposure relates to long-term injury outcomes.

What it means:

The visualization indicates a mild positive relationship: wingers with **heavier early workloads** tend to have **more missed days after 22**.

For instance, **Neymar**, **Eden Hazard**, and **Arjen Robben** appear on the upper-right side, suggesting that extensive early play may correlate with greater long-term injury burden. Conversely, **Cristiano Ronaldo** and **Mohamed Salah** demonstrate exceptional durability despite high match volumes, showing that outliers can maintain longevity through conditioning or play style.

Overall, this visual forms the analytical core of the project; it connects early workload to later injury costs and reinforces the broader idea that **early physical intensity may influence long-term durability**.

Visual 4. Correlation Matrix: Workload and Injury Metrics

What it shows:

This heatmap displays the correlation coefficients between key workload and injury variables. Each cell represents the degree of relationship between two factors — from **-1 (strong negative)** to **+1 (strong positive)**.

The red tones indicate positive correlations, while the blue tones represent negative ones.

The variables analyzed include:

- Matches Before 22
- Matches After 22
- Injuries After 22
- Missed Days After 22

What it means:

The most notable relationship is the **strong positive correlation (0.87)** between **Injuries After 22** and **Missed Days After 22**, confirming that players who sustain more injuries also tend to miss significantly more time.

A weaker positive correlation (0.30) exists between **Matches Before 22** and **Missed Days After 22**, suggesting that a heavier early workload may slightly contribute to more absences later on. Interestingly, there is a **negative correlation (-0.53)** between **Matches Before 22** and **Matches After 22**, implying that players who were highly active early in their careers often see reduced match participation later, possibly due to injury accumulation or fatigue.

This matrix statistically reinforces the trends observed in the earlier visuals — that **early workload, while beneficial for development, may be associated with greater long-term injury risk and reduced availability in later years.**

Visual 5. Best of the Best: Messi vs Ronaldo

What it shows:

This radar chart provides a direct durability and workload comparison between Lionel Messi and Cristiano Ronaldo across four normalized dimensions:

- Matches Before 22
- Matches After 22
- Injuries After 22
- Missed Days After 22

The polygon shape of each player reveals their overall balance between workload and durability. Messi is represented in blue, while Ronaldo is shown in gold.

What it means:

The chart highlights two distinct athletic profiles.

Cristiano Ronaldo exhibits consistently high match counts both before and after 22, reflecting his longevity and physical conditioning. His polygon remains compact, indicating fewer missed days and injuries relative to workload.

In contrast, **Lionel Messi's** shape extends further on the **"Injuries"** and **"Missed Days"** axes, showing slightly higher vulnerability in those areas despite comparable workload levels.

The visual emphasizes that while both athletes maintained elite performance levels for two decades, Ronaldo's long-term durability appears marginally superior, underscoring his exceptional recovery and conditioning discipline.

Interpretation note: The radar layout helps visualize balance and resilience a smaller, symmetrical polygon represents better long-term consistency, while extensions indicate areas of relative weakness.

Visual 6 Lamine Yamal: Single-Player Durability Projection

What it shows:

This visual provides a personalized durability forecast for Lamine Yamal, based on his real early-career data.

The two bar charts compare his performance before turning 22 and his projected post-22 values.

The left plot displays injury counts, while the right plot shows total missed days due to injuries. Predictions are calculated using the average injury-growth and missed-days-growth factors derived from the elite-winger dataset analyzed earlier.

What it means:

The projection suggests that if Yamal continues his current match intensity, his long-term injury load could rise substantially:

- Injuries increasing from 8 → 35 ($\approx 4.4\times$ growth)
- Missed days increasing from 200 → 860 ($\approx 4.3\times$ growth)

These values mirror the average post-22 trajectories observed among high-usage wingers such as Neymar, Hazard, and Robben.

While not deterministic, the visual serves as a data-driven warning that extraordinary early workloads may accelerate physical wear.

It highlights the importance of managed rotation, recovery time, and injury prevention programs for young high-minute players.

Interpretation note:

The projection is statistical, not medical.

It assumes Yamal follows the same injury-growth trends as the studied

winger group and should be interpreted as an illustrative risk scenario, not a medical forecast.

Conclusion

The analysis of fifteen elite wingers from the 21st century demonstrates a consistent pattern between **early match workload** and **later injury frequency**.

While not an exact causal link, the data supports the hypothesis that players with heavier match exposure before the age of 22 tend to experience greater physical strain, leading to more injuries and missed days as their careers progress.

Key insights include:

- Most wingers show a **×4 to ×5 increase** in injury counts after age 22.
- Players with intense early workloads, such as **Neymar, Robben, and Hazard**, exhibit the steepest post-22 growth in missed days.
- In contrast, highly durable athletes like **Cristiano Ronaldo** and **Mohamed Salah** sustain both high workloads and exceptional long-term availability, indicating that conditioning and recovery management can offset early fatigue risks.
- The **correlation matrix** confirms that early workload and long-term missed days have a modest positive relationship, while injury frequency and missed days are strongly correlated.

Finally, the **Lamine Yamal projection** illustrates these findings in a real-world scenario.

If his current rate continues, statistical models predict a notable rise in injury burden after 22, a reminder of the importance of **load management, rest, and long-term planning** in modern football.

In summary: early success and exposure come at a measurable physical cost. Balancing development with protection may be key to sustaining performance and extending peak years in professional football.