Injuries in Professional and Collegiate Sports: Identifying Athletic Predisposition and Susceptibility to Injury and Analyzing Modern Incorporation of Biomedical Technology

By Peter Chapman

STAT468: Senior Research Project

University of Delaware

December 10, 2025

Introduction and Research Questions

- **Problem**: High multitude of injuries that occur to athletes
- Led to an increase in research on identifying injury risk and prevention within professional and collegiate athletics
- Popular research topic in football, basketball, soccer, and baseball
- **Research Question 1**: Are there predispositions to certain injuries for athletes of a certain size, position, sport, or activity? If so, what is the full extent of this susceptibility?
- Several biomedical advancements within these sports have been made in previous seasons to address the rise in injuries
- **Research Question 2**: Are modern advancements in injury risk and prevention effective in reducing the number of injuries?
- Leagues analyzed: NFL, NHL, EPL, NCAA sports

Motivation and Hypotheses

- **Goal #1:** Help identify potential causes of injuries in significant sports leagues
- **Goal #2:** Evaluate technological advancements that were recently implemented into collegiate and professional athletics
- **Hypothesis #1:** Certain external and internal factors contribute to determining likelihood of injury
  - External factors: Playing surface, player position, quality of protective equipment
  - Internal factors: Age, previous injuries, height and weight
- **Hypothesis #2:** Biomedical and biomechanical advancements in athletics have been effective in reducing injuries, especially with common injuries and in sports with a high number of injuries
  - Common injuries: Torn UCL for pitchers in baseball, concussions for running backs in football, etc.
  - Leagues with many injuries: NFL, NHL, EPL, NCAA sports
- Less confidence in the second hypothesis, as advancements are very recent and online research and data may be limited

Literature Review

External and Internal Factors

Internal Factors

- Several internal factors affect likelihood and severity of injuries (Wentao)
  - Age: Older age -> Higher risk
  - Health status: Bad health -> Higher risk
  - Physical condition: Height, weight, bone density, gender, flexibility
  - Psychological factors: Self-esteem and self-identity
- Similar findings and other significant internal factors (Taimela, et al.)
  - Injury history: Signficant if not properly treated
  - Weight: Heavier weight -> Higher risk
  - Height: Taller height -> Higher risk

External Factors

- Many characteristics can be considered external risk factors (Bahr and Holme)
  - Human factors
    - Referee decisions
    - Opponent behavior
    - Team playstyle
  - Protective factors
    - Equipment designed to protect athletes
    - Shin guards, helmets, etc.
  - Sports equipment
    - All other equipment that is used
    - Gloves, sticks, baseball bats, etc.
  - Environment
    - Weather: Rain, snow, ice, etc.
    - Field or floor type, quality of the playing surface

Multifactorial Nature of Injury

- Heavily researched and debated topic within sports
- Definition: Many factors could be correlated with the occurrence of an injury
- Cause of a specific injury to an athlete cannot be isolated to one factor (Bahr and Holme)
- Internal/external factors are correlated with athletic injuries
- **Statistical Application**: Multivariate approach is necessary for any predictive model that attempts to assess injury risk
- Suggestion: Logistic regression model with multiple predictor variables must be considered

$$\ln\left[\frac{P(injury)_{Athlete}}{1-P(injury)_{Athlete}}\right]$$

$$= \beta_0 + \beta_1 Factor1_{Athlete} + \beta_2 Factor2_{Athlete} + \cdots + \beta_k Factork_{Athlete}$$

Biomechanical and Biomedical Advancements in Athletics

- Rapid development as technology improves, especially in the 21st century
- Some advancements target a specific injury as opposed to a certain sport or athlete type
  - Non-contact knee injuries (ACL, PCL, meniscus injuries)
    - Research into playing surface and field upkeep (artificial or natural field)
    - Improvements in footwear (cleats, basketball shoes, skates)
  - Head and concussion injuries (Nath, et al.)
    - High-tech materials (carbon fiber and polycarbonate) used for hockey and football helmets
    - Creation of the Guardian Cap in professional and collegiate football
- Other advancements are general across all sports (Nath, et al.)
  - Smart fabrics that collect vital signs, assess fatigue, and motor functions
  - Wearable technology such as GPS trackers, LPS trackers, augmented reality
  - Embedded sensors that collect movement patterns, forces, athlete posture
  - Artificial intelligence can make predictions and recommendations from data

Methodology

English Premier League Model for Injury Days Missed

- Data collected from Kalpesh Kolambe on Kaggle; scraped using 'worldfootballR' package
- Relevant internal factors: Age, height, weight, BMI
- Relevant external factors: Minutes played, position, matches played, number of days missed due to injury in the player's previous season
- Dependent variable of interest: Number of days missed due to injury
- Binary variables created to represent player position; goalkeepers are reference group
- Standard multiple linear regression model used to predict injury days

$$
\begin{aligned}
&SeasonDaysInjured_{Player,Year} \\
&= \beta_0 + \beta_1 SeasonMinutesPlayed_{Player,Year} + \beta_2 Age_{Player,Year} + \beta_3 BodyMassIndex_{Player,Year} \\
&+ \beta_4 DaysInjuredPreviousSeason_{Player,Year} + \beta_5 isForward_{Player,Year} \\
&+ \beta_6 isMidfielder_{Player,Year} + \beta_7 isDefender_{Player,Year} + \varepsilon_{Player,Year}
\end{aligned}
$$

Binary Logistic Regression for Injury Likelihood in NCAA Training Sessions

- Data collected from ‘Ziya’ on Kaggle; protects identities of athletes to address HIPAA
- Relevant internal factors: Heart rate, respiratory rate, blood oxygen level, respiratory rate
- Relevant external factors: Impact force of training session, session duration, training type, sport type
- Dependent variable of interest: Whether or not student-athlete suffered an injury during training
- Binary variables created for level of contact (mild, heavy, or none) and intensity of activity
- Binary logistic regression to quantify likelihood of student-athlete injury

$$\ln\left[\frac{P(injury)_{Athlete,Session}}{1-P(injury)_{Athlete,Session}}\right] =$$
$$= \beta_0 + \beta_1 HeartRate_{Athlete,Session} + \beta_2 RespiratoryRate_{Athlete,Session}$$
$$+ \beta_3 SkinTemperature_{Athlete,Session} + \beta_4 BloodOxygenLevel_{Athlete,Session}$$
$$+ \beta_5 ImpactForce_{Athlete,Session} + \beta_6 CumulativeFatigueIndex_{Athlete,Session}$$
$$+ \beta_7 SessionDuration_{Athlete,Session} + \beta_8 isHeavyContact_{Athlete,Session}$$
$$+ \beta_9 isMildContact_{Athlete,Session} + \beta_{10} isIntense_{Athlete,Session}$$

NFL Playing Surface and Injury Severity Binary Logistic Regression Model

- Data collected from ‘JasonZivkovic’ on Kaggle; confirmed lower-limb NFL injuries
- Exclusively external factors: Playing surface that the injury occurred on (synthetic or natural), body part that was injured by the football player (foot, ankle, or knee)
- Series of binary threshold variables for injury days missed; 1 day. 7 days, 28 days, 42 days
- Dependent variable of interest: Whether or not the football player missed a minimum of 42 days due to injury; allows for a study on the most severe injuries
- Binary logistic regression with two indicator variables
  - Playing surface during injury was synthetic (reference: natural surface)
  - Injury was in the knee (reference: foot or ankle)

$$\ln\left[\frac{P(Min42DaysMissed)_{Athlete,Injury}}{1-P(Min42DaysMissed)_{Athlete,Injury}}\right] = \beta_0 + \beta_1 isSynthetic_{Athlete,Injury} + \beta_2 isKnee_{Athlete,Injury}$$

NHL Injury Trends by Position

- Data collected and provided by an online database called NHL Injury Viz
- Position group recorded for each injury (defenseman, forward, goaltender)
- Injured body part provided for each observation
  - Very specific injuries: Appendicitis, bronchitis, dehydration, COVID-19
  - More broad injury location: Upper-body, lower-body, mid-body, arm, leg, foot, head
- Data from 2000-01 to 2025-26, including playoffs
- Both number of days missed and count of each injury provided
- Data visualizations created using Tableau Public software
  - Heat map showing frequency of injury type by the three positions
  - Line graph showing injury trends over time by position group

Results and Discussion

EPL Injury Days Missed Model Outcomes

- Model R-squared of 11.39% reinforces idea of multifactorial nature of injury
- Statistically significant factors: Season minutes played, days injured in the previous season
- Clinically significant factors: Age, variables representing player position
- Defenders are predicted to miss the most days due to injury, followed by midfielders, forwards, and then goalkeepers
- Season minutes played and BMI have negative relationships with days missed
- Age and days missed in previous season have positive relationships

Variable	Coefficient	P-Value
Intercept	88.569744	0.01616
Season Minutes Played	-0.027323	<0.0001
Age	0.969747	0.10979
Body Mass Index (BMI)	-0.496373	0.75396
Days Injured in Previous Season	0.086429	0.00467
isForward	10.634257	0.26964
isMidfielder	12.337696	0.16918
isDefender	17.0408	0.05643

NCAA Training Sessions Binary Logistic Regression Model Results

- Multicollinearity existed between two pairs of factors; one variable removed from each pair
  - Blood oxygen level percentage and skin temperature in degrees Celsius
  - Heart rate (beats per minute) and respiratory rate (breaths per minute)
- Statistically significant factors: Impact of force (N) of session, cumulative fatigue index, blood oxygen level percentage
- Clinically significant factors: Contact level of sport played

Variable	Coefficient	P-Value
Intercept	13.315	<0.0001
Heart Rate (BPM)	-0.00279	0.3334
Blood Oxygen Percent Level	0.1196	<0.0001
Impact of Force (N)	-0.00597	0.0001
Cumulative Fatigue Index	-1.9031	<0.0001
Duration of Session (min)	-0.00162	0.6241
isHeavyContact	0.3394	0.0606
isMildContact	-0.0294	0.8452
isIntense	-0.124	0.365

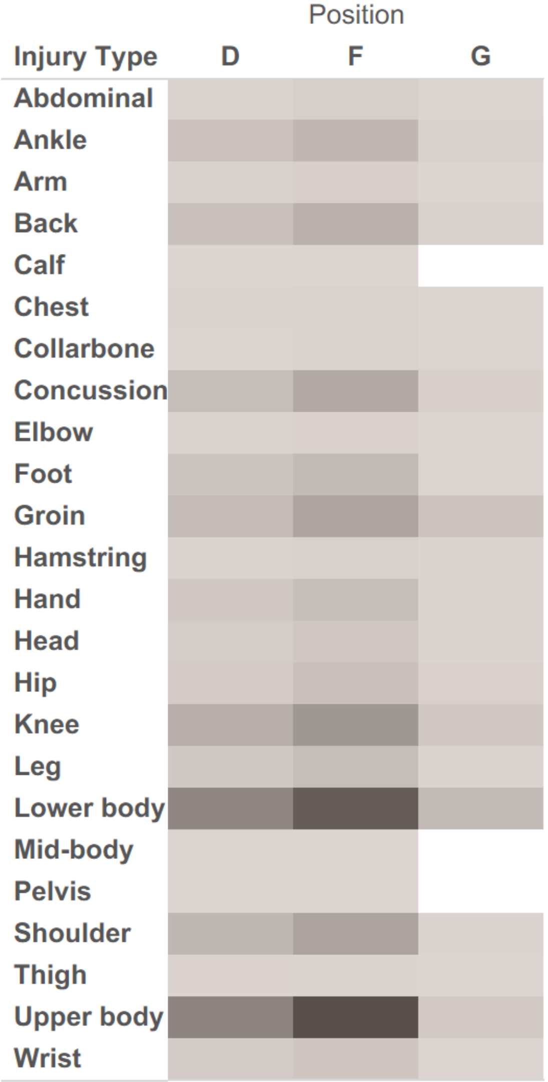
NFL Playing Surface and Injury Severity Binary Logistic Regression Model

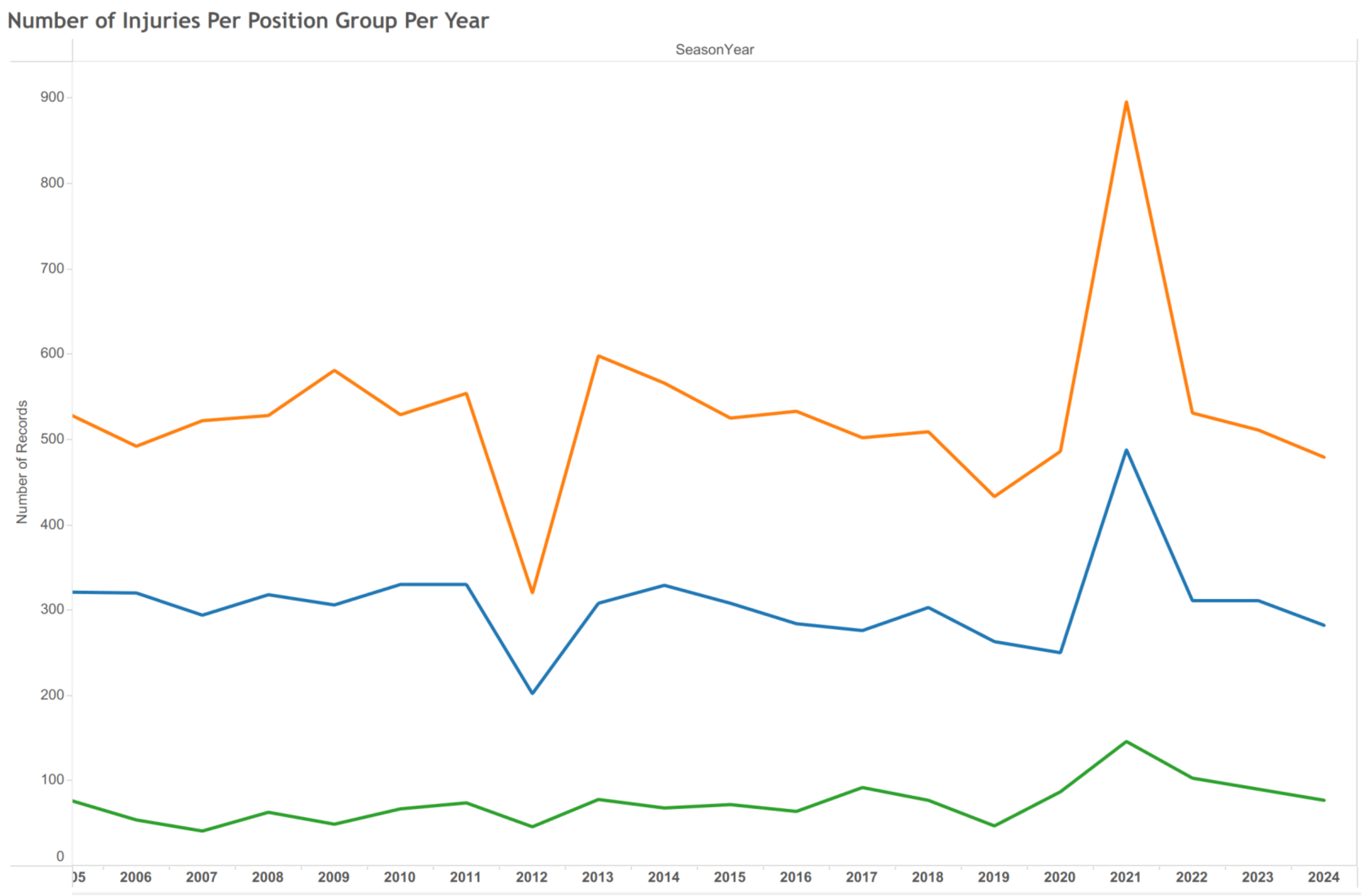
$$\ln\left[\frac{P(Min42DaysMissed)_{Athlete,Injury}}{1-P(Min42DaysMissed)_{Athlete,Injury}}\right]$$
$$=\ 0.7832 - 0.1642 isSynthetic_{Athlete,Injury} + 0.3992 isKnee_{Athlete,Injury}$$

- Both variables statistically insignificant at the 5% level
- May both hold clinical significance due to health and safety setting
- Interpretations can be developed from the odds ratio estimates
  - Knee injury increases likelihood of missing 42 days by 49.1%
  - Synthetic surface is associated with a 15.1% decrease in the likelihood of missing 42 days due to injury

Variable	**Coefficient**	**P-Value**
Intercept	0.7832	<0.0466
isSynthetic	-0.1642	0.7119
isKnee	0.3992	0.3682

Heat Map of Injury Frequency by Position Group





Impact of Guardian Caps on Concussions

Collegiate Level (NCAA)

- Study from MIT’s collegiate team (Carratu)
- Two different tests to evaluate
  - Visual response test
  - Physical perception test
- Players naturally grouped into two categories
  - Treatment group: Wore Guardian Caps
  - Control group: Did not wear Guardian caps
- Two conclusions made from the study
  - No statistically significant difference between the two groups
  - Guardian Cap group had less variation for response speed, eye response constriction, tracking ability

Professional Level (NFL)

- Laboratory testing of the Guardian Cap NXT (Funk, et al.)
- Setting: Preseason practices from 2018 to 2023
  - Excludes the 2020 preseason
  - 2022: Linemen, linebackers, and tight ends required
- Controls included to isolate effect of Guardian Caps
  - Practice intensity (pads or no pads)
  - Player position (offensive lineman, defensive lineman, linebacker, tight end)
- Two main takeaways
  - Concussion rate for Guardian cap players decreased 54% after helmets were introduced
  - Could not be solely attributed to the Guardian Cap NXT

Conclusion and Future Research

- Statistical findings are best to be interpreted by sport as correlations
  - Professional soccer: Minutes played, days missed due to injury in previous season
  - Collegiate sports: Contact level, blood oxygen level percentage, cumulative fatigue index, and total impact of forces during the training session
  - National Football League: Higher risk of severe injury on natural playing surface
- **Conclusion**: Mix of internal and external factors (multifactorial nature of injury)
  - Results vary across sport, level, and physical activity
  - Injury likelihood and severity cannot be generalized across all sports
- Future research can make this study even more accurate
  - Full populations are largely better than small samples of injuries
  - Regression estimates would become more accurate
  - Model would hold higher relevance for predicting injuries and their severity

Limitations

1. Difficulty in finding free, public data
   a. Many leagues are very private with their injury data
   b. Most data sets are behind paywalls or have restrictions on user access
   c. Leagues excluded: Major League Baseball (MLB), National Hockey League (NHL), National Association of Stock Car Auto Racing (NASCAR)
2. Collected data sets were samples, not full populations of injury
   a. Required assumption: Samples are representative of full populations of injuries
   b. Statistical effects are estimates of true population effects
3. Recency of biological and technological advancements
   a. Limited amount of available data for evaluating effectiveness
   b. Advancements and progress are constantly being made

Acknowledgements

- **Professor Bryan Crissinger**
  - University of Delaware Department of Applied Economics and Statistics
  - STAT468: Senior Research Project faculty advisor for this semester-long project
  - Provided advice on topic, research questions, data collection and transformation, statistical and data analysis, rough draft, and presentation
- **Professor Patrick DeFeo**
  - University of Delaware College of Agriculture and Natural Resources
  - STAT611: Regression Analysis Instructor
  - Key concepts: Regression techniques, binary logistic regression, multicollinearity among predictors
- **Professor Jack Davis**
  - Associate Director of Business Intelligence, Analytics, and Strategy at the University of Delaware Athletic Department
  - SPAX402: Predictive Analysis with Athletics Data Instructor
  - Key concepts: Microsoft Excel, Microsoft Power Query, data collection and transformation methods

Literature Review Works Cited

Bahr, R, and Holme, I. “Risk Factors for Sports Injuries -- a Methodological Approach.” *British Journal of Sports Medicine*, vol. 37, no. 5, 1 Oct. 2003, pp. 384–392, bjsm.bmj.com/content/37/5/384, https://doi.org/10.1136/bjsm.37.5.384. Accessed 27 Sept. 2025.

Carratu, Christopher. “Guardian Cap Effectiveness on Reducing Impact Accelerations and Enduring Player Health at the Collegiate Football Level.” *Mit.edu*, 23 June 2025, dspace.mit.edu/handle/1721.1/162566, https://hdl.handle.net/1721.1/162566. Accessed 7 Dec. 2025.

Funk, James R., et al. “An Analysis of Guardian Cap Use and Changes in the Concussion Rate in National Football League Preseason Practices from 2018 to 2023.” *The American Journal of Sports Medicine*, 2 July 2025, journals.sagepub.com/doi/full/10.1177/03635465251351288, https://doi.org/10.1177/03635465251351288. Accessed 7 Dec. 2025.

Nath, Debraj, et al. “View of Progress of Protective Gear in Preventing Sports Injuries.” *Brazilianjournals.com.br*, 2025, ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/67830/48261. Accessed 6 Dec. 2025.

Taimela, Simo, et al. “Intrinsic Risk Factors and Athletic Injuries.” *Sports Medicine*, vol. 9, no. 4, Apr. 1990, pp. 205–215, link.springer.com/article/10.2165/00007256-199009040-00002, https://doi.org/10.2165/00007256-199009040-00002. Accessed 27 Sept. 2025.

Wentao, Zhao. “Analysis of the Causes of Sports Injuries in Sports Training.” *Frontiers in Sport Research*, vol. 6, no. 1, 2024, francis-press.com/uploads/papers/GABe79pq5P7HyzaqpGLsIxYGW4JZwvH1lJHfA1eT.pdf, https://doi.org/10.25236/fsr.2024.060106. Accessed 27 Sept. 2025.

Data Sources

1. https://www.kaggle.com/datasets/ziya07/college-sports-injury-detection
2. https://www.kaggle.com/datasets/kolambekalpesh/football-player-injury-data?resource=download
3. https://www.kaggle.com/code/jaseziv83/an-analysis-of-nfl-injuries
4. https://nhlinjuryviz.blogspot.com/p/index-page.html