



Sports Injury Incidence and Risk Factor Dataset among Collegiate Athletes: A Data Paper

Dr. Sanjeev Yadav

Assistant Professor, Department of Physical Education, Kalinga University, Raipur, Chhattisgarh, India, Email:
sanjeev.yadav@kalingauniversity.ac.in

Abstract: *This data paper presents a structured dataset documenting sports injury incidence and associated risk factors among collegiate-level athletes across multiple sporting disciplines. The dataset captures athlete demographics, training load, injury characteristics, playing surface, and preventive practice indicators such as warm-up routines and prior injury history. It is intended to assist sports scientists, physiotherapists, coaches, and physical education researchers in analysing injury patterns and evaluating risk factors relevant to injury prevention programmes in a university sporting context. The paper describes the data collection protocol, variable structure, technical validation steps, and appropriate usage guidance for the dataset.*

Keywords: *Sports Injury, Injury Surveillance, Collegiate Athletes, Risk Factors, Physical Education, Injury Prevention, Training Load.*

1. Background & Summary: Sports injuries remain a significant concern in collegiate athletics, with implications for athlete health, training continuity, and competitive performance. Injury surveillance systems that systematically document incidence, type, and contributing risk factors are essential for designing effective prevention strategies, yet structured and openly documented injury datasets from Indian university sporting programmes remain scarce.

This data paper addresses that gap by presenting an injury surveillance dataset compiled from athletes participating in institutional sports programmes across multiple disciplines. The dataset is designed for reuse in secondary analyses such as identifying discipline-specific injury patterns, examining the relationship between training load and injury occurrence, and assessing the protective role of structured warm-up practices.

2. Methods

2.1 Participants: Data were compiled from athletes aged 17–25 years participating in institutional-level competitive sport across disciplines including athletics, football, basketball, kabaddi, and volleyball. Athletes were enrolled through the sports department's season-long injury reporting mechanism, with participation contingent on informed consent.

2.2 Data Collection Instruments

- A standardised injury report form completed by team physiotherapists/coaches at the time of injury occurrence

- A pre-season questionnaire capturing training hours, warm-up practice, and prior injury history
- Anthropometric measurement of height and weight for BMI computation
- Match and training exposure logs to contextualise injury onset (acute vs overuse)

2.3 Procedure: Injury data were recorded prospectively across a full competitive season. Each reportable injury — defined as one resulting in restriction from training or competition — was logged at the point of occurrence, including body part affected, injury type, playing surface, and estimated days lost. Pre-season questionnaire data were linked to each athlete’s injury record using the anonymised athlete identifier.

3. Data Records: The dataset is organised such that each row corresponds to one athlete, with injury-specific fields populated where an injury was reported during the season. Table 1 describes the variables included.

Table 1: Variable description of the dataset

Athlete_ID	Unique anonymised identifier assigned to each athlete	Alphanumeric	AT001–AT300
Sport_Discipline	Primary sport played by the athlete	Categorical	Athletics/Football/Basketball/Kabaddi/Volleyball
Age	Chronological age of athlete at time of reporting	Years (integer)	17–25
Gender	Self-reported gender of athlete	Categorical (M/F)	—
Training_Hours_Week	Average structured training hours per week	Hours	4–20
Injury_Occurred	Whether the athlete sustained a reportable injury during the season	Binary (Yes/No)	—
Injury_Type	Nature of injury sustained, where applicable	Categorical	Sprain/Strain/Fracture/Contusion/Other
Body_Part_Affected	Anatomical region of the reported injury	Categorical	Ankle/Knee/Shoulder/Back/Hamstring/Other
Injury_Onset	Mechanism of injury onset	Categorical	Acute/Overuse
Severity_Days_Lost	Number of training/competition days lost due to injury	Days (integer)	0–90

Surface_Type	Playing surface on which the injury occurred	Categorical	Turf/Hard court/Track/Mat/Other
Warm_Up_Practice	Whether athlete regularly followed a structured warm-up protocol	Binary (Yes/No)	—
Previous_Injury_History	Whether athlete had a prior injury to the same body region	Binary (Yes/No)	—
BMI	Body Mass Index of the athlete	kg/m ²	17–29

4. Descriptive Summary of Injury Patterns: Table 2 presents an illustrative season-level summary of injury incidence and characteristics. These figures are indicative and should be recomputed against the final compiled dataset before formal reporting or publication.

Table 2: Summary of injury incidence and characteristics

Overall injury incidence	31.4% of athletes reported ≥ 1 injury	Illustrative, season-level figure
Most affected body part	Ankle (26%), Knee (21%)	Combined across all disciplines
Most common injury type	Sprain (38%)	Followed by strain (29%)
Mean days lost per injury	9.6 ± 7.8 days	Mean $\pm$ SD
Acute vs Overuse onset	62% Acute / 38% Overuse	—
Athletes with prior injury history	44%	Self-reported

5. Technical Validation

- Injury report forms were completed by qualified physiotherapists/coaching staff at the time of occurrence to minimise recall bias.
- Pre-season questionnaire responses were cross-checked against training attendance records where available.
- Ambiguous or incomplete injury reports were flagged and excluded from the final dataset during digitisation.
- A random 10% subsample of records was reviewed by a second staff member to verify coding consistency for injury type and body part classification.

6. Usage Notes: Users of this dataset should note that injury reporting relies on athlete and staff disclosure, and mild injuries not resulting in restricted participation may be under-represented. The binary warm-up and injury-history variables are self-reported and should be interpreted as indicative rather than clinically verified. The dataset is well suited for descriptive epidemiology and exploratory risk-factor analysis, but causal claims regarding injury prevention should be substantiated through controlled prospective study designs.

7. Conclusion: This data paper documents a season-long sports injury surveillance dataset from collegiate athletes across multiple disciplines, detailing injury type, severity, and associated risk factors such as training load and warm-up practice. Making the data structure and collection protocol transparent is intended to support reuse of the dataset in injury-prevention research and to encourage broader adoption of systematic injury surveillance within university physical education and sports departments.

References

- Bahr, R., & Krosshaug, T. (2005). *Understanding injury mechanisms: A key component of preventing injuries in sport. British Journal of Sports Medicine, 39(6), 324–329.*
- Fuller, C. W., Ekstrand, J., Junge, A., et al. (2006). *Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. British Journal of Sports Medicine, 40(3), 193–201.*
- Van Mechelen, W., Hlobil, H., & Kemper, H. C. G. (1992). *Incidence, severity, aetiology and prevention of sports injuries: A review of concepts. Sports Medicine, 14(2), 82–99.*
- World Health Organization (2020). *WHO Guidelines on Physical Activity and Sedentary Behaviour. World Health Organization, Geneva.*

Citation: Yadav, S., (2026) “Sports Injury Incidence and Risk Factor Dataset among Collegiate Athletes: A Data Paper”, *Bharati International Journal of Multidisciplinary Research & Development (BIJMRD)*, Vol-4, Issue-07(1), July-2026.