Adding confidence to our injury burden estimates: is bootstrapping the solution?

Sean Williams ,[1] Joseph W Shaw,[2,3] Carolyn Emery ,[4] Keith A Stokes[1,5]

INTRODUCTION

Injury burden is a composite measure of injury incidence and mean severity that can be used to understand the overall impact of injuries and help identify priority areas for injury prevention.[1] Injury burden has been used within rugby union epidemiological studies since the early 2000s,[2] but it is now recognised and recommended within other sports, including the most recent International Olympic Committee consensus statement for the recording and reporting of epidemiological data on injury and illness.[3] Injury burden is normally reported as athlete days absence per 1000 athlete-hours and is derived from the product of injury incidence (expressed as injuries sustained/1000 athlete-hours) and severity (expressed as the mean severity of injury in days).[1] While the value of injury burden as an output measure from injury surveillance studies is evident, there appears to be some confusion in the literature regarding its calculation. For instance, some authors have used median severity to calculate injury burden rather than mean severity, as discussed in a recent critical review.[4] In addition, there appears to be no clear guidance within the sports medicine literature regarding the most appropriate way to calculate confidence intervals (CIs) for this metric.

ESTIMATING UNCERTAINTY

CIs convey the uncertainty about a point estimate (in this case, an injury burden rate) and are thus crucial for interpreting results and informing decision-making. Injury burden estimates have typically been modelled using the Poisson distribution.[5] Using this approach, the precision around the burden estimate is driven entirely by 'N' in Equation 1.[5]

Equation 1.

$$Lower\ 95\%CI = Burden\ rate \div \left(Exp^{\left(1.96 \div \sqrt{N}\right)}\right)$$

$$\text{Upper 95\% CI} = \text{Burden rate} \times \left(Exp^{\left(1.96 \div \sqrt{N}\right)}\right)$$

However, several different values for 'N' in Equation 1 have been used within the literature. Many studies, including the International Olympic Committee consensus statement,[3] use the total number of athlete days absence, while others have used the injury burden rate itself[6] or the injury count[7]; each approach produces a different CI width (table 1) and could therefore hugely impact on the interpretation of the burden rate. For example, using the total number of injury days as the 'N' value produces a CI that is extremely narrow when the total number of injury days is large, greatly increasing the likelihood of making a type I error. Of these approaches, using the injury count as the 'N' value appears most appropriate as it produces a CI width that changes proportionately in line with the sample size (ie, the number of injury events). Yet, there is no consensus on this in the sports medicine literature. Moreover, the appropriateness of assuming a Poisson distribution when modelling injury burden data may be questioned given the extreme positive skew that is often present, and this approach does not account for the multiple sources of random error (ie, from both injury incidence and average injury severity).

ENTER BOOTSTRAPPING

Bootstrapping is a statistical resampling technique that involves sampling with replacement from an original dataset (at least 1000 iterations are recommended), with confidence limits derived from percentiles of the sampled values.[8] Bootstrapping enables the estimation of the sampling distribution of a statistic without making assumptions about the underlying population distribution. As such, it appears a promising solution for producing robust CIs around an injury burden estimate. Bootstrapping is now readily available for descriptive statistics in statistical programmes such as SPSS and R (see online supplemental materials for worked examples). Note, a burden value is required for each injury record in the dataset, which may be obtained by dividing each injury severity value by the athlete's exposure time (and then multiplying by 1000 to express as days absence per 1000 athlete-hours); where individual exposure time is not available, the total/team exposure time should be divided by the total injury count. As shown in table 1, the bootstrapping method changes proportionately in line with the sample size and produces the most conservative CI width. By avoiding the need to make assumptions about the distribution from which the data were obtained, we believe the bootstrapped CIs are the most trustworthy estimates of the precision around the burden estimate.

The 'Delta Method' is another viable solution for obtaining a valid SE that accounts for both sources of uncertainty when calculating a burden value.[8] Further considerations are needed when modelling injury burden as an outcome variable in regression models (eg, the use of a negative binomial distribution and/or bootstrapping approaches to account for clustering),[5] but this is beyond the scope of the current editorial.

MOVING FORWARD WITH CONFIDENCE

The uncertainty associated with output measures from injury surveillance studies should always be presented and used to guide decisions (eg, whether a meaningful change has occurred between seasons). Although a full methodological review of the proposed methods is warranted, bootstrapping appears to provide a robust and appropriate estimate of uncertainty for burden rate estimates, and this key methodological point should be considered by sports injury researchers.

Twitter Sean Williams @statman_sean, Joseph W Shaw @JosephShaw_ and Carolyn Emery @CarolynAEmery

Contributors SW drafted the manuscript. KAS, JWS and CE provided input. All authors edited and reviewed the final version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

[1]Centre for Health, Injury and Illness Prevention in Sport, University of Bath, Bath, UK
[2]Faculty of Sport, Technology and Health Sciences, St Mary's University Twickenham, London, UK
[3]Ballet Healthcare, Royal Opera House, London, UK
[4]Sport Injury Prevention Research Centre, University of Calgary, Calgary, Alberta, Canada
[5]Rugby Football Union, London, UK

Correspondence to Dr Sean Williams, Centre for Health, and Injury & Illness Prevention in Sport, University of Bath, Bath, BA2 7AY, UK; S.Williams@bath.ac.uk

Table 1 Injury burden CIs calculated using various approaches to derive CI width

Method	Injury count	Exposure (h)	Total injury days	Incidence rate (injuries/1000 hours)	Mean severity (days)	Injury burden (days/1000 hours)	95% CI	CI width (Upper CI–Lower CI)
Bootstrapping	200	20 000	6000	10	30	300	228 to 382	154
N=Injury count	200	20 000	6000	10	30	300	261 to 345	83
N=Injury burden rate	200	20 000	6000	10	30	300	268 to 336	68
N=Total injury days	200	20 000	6000	10	30	300	293 to 308	15
Bootstrapping	20	2000	600	10	30	300	106 to 545	439
N=Injury count	20	2000	600	10	30	300	194 to 465	271
N=Injury burden rate	20	2000	600	10	30	300	268 to 336	68
n=Total injury days	20	2000	600	10	30	300	277 to 325	48

'N' refers to the value used in Equation 1.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bjsports-2023-107496).

To cite Williams S, Shaw JW, Emery C, *et al*. *Br J Sports Med* 2024;**58**:57–58.

Accepted 14 November 2023
Published Online First 28 November 2023

Br J Sports Med 2024;**58**:57–58.
doi:10.1136/bjsports-2023-107496

ORCID iDs
Sean Williams http://orcid.org/0000-0003-1460-0085
Carolyn Emery http://orcid.org/0000-0002-9499-6691

REFERENCES

1 Fuller CW. Managing the risk of injury in sport. *Clin J Sport Med* 2007;17:182–7.
2 Brooks JHM, Fuller CW, Kemp SPT, *et al*. Epidemiology of injuries in English professional Rugby Union: part 1 match injuries. *Br J Sports Med* 2005;39:767–75.
3 Bahr R, Clarsen B, Derman W, *et al*. International olympic committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020 (including STROBE extension for sport injury and illness surveillance (STROBE-SIIS)). *Br J Sports Med* 2020;54:372–89.
4 Fuller CW. Why median severity and ordinal scale severity values should not be used for injury burden results: a critical review. *Int J Sports Med* 2023;44:313–9.
5 Verhagen E, van Mechelen W. Sports injury research. In: *Sports injury research*. Oxford University Press, 2010.
6 Sewry N, Verhagen E, Lambert M, *et al*. Seasonal time-loss match injury rates and burden in South African under-16 Rugby teams. *J Sci Med Sport* 2019;22:54–8.
7 Robertson CM, Williams S, West SW, *et al*. Influence of playing surface on match injury risk in men's professional Rugby Union in England (2013–2019). *Scand J Med Sci Sports* 2022;32:1615–24.
8 LaFlair GT, Egbert J, Plonsky L. A practical guide to Bootstrapping descriptive statistics, correlations, t tests, and Anovas. In: *Advancing Quantitative Methods in Second Language Research* 46. 2015: