Statistical Methods for Handling Observation Clustering in Sports Injury Surveillance

Avinash Chandran, PhD, MS*; Derek Brown, PhD, MS‡; Aliza K. Nedimyer, MA, ATC*†; Zachary Y. Kerr, PhD, MPH*

*Matthew Gfeller Sport-Related TBI Research Center, Department of Exercise and Sport Science and †Human Movement Science Curriculum, University of North Carolina at Chapel Hill; ‡Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston

Context: Advances in sports injury-surveillance methods have made it possible to accommodate non–time-loss (NTL) injury reporting; however, the analysis of surveillance data now requires careful consideration of the nuances of NTL injury records.

Background: Injury-surveillance mechanisms that record NTL injuries are more likely to contain multiple injury records per athlete. These must be handled appropriately in statistical analyses to make methodologically sound inferences.

Methods: We simulated datasets of NTL injuries using varying degrees of observation clustering and compared the inferences made using traditional techniques with those made after accounting for clustering in computations of injury proportion ratios.

Results: Inappropriate handling of even moderate clustering resulted in flawed inferences in 10% to 12% of our simulations. We observed greater bias in our estimates as the degree of clustering increased.

Conclusions: We urge investigators to carefully consider observation clustering and adapt analytical methods to accommodate the evolving sophistication of surveillance.

Key Words: injury proportion ratios, independence, sandwich estimator

---

Key Points

- Sophisticated sports injury-surveillance methods increase the likelihood of clustered observations in surveillance data.
- The sandwich estimator may be used when analyzing clustered data to minimize biased estimates and flawed inferences.

Sports injury surveillance has been critical in identifying patterns of injury incidence and outcomes among athletes in different sports.[1–5] As its popularity has increased, surveillance methods have been adapted and mechanisms have grown in complexity. One notable advance in recent years has been the accommodation of *non–time-loss (NTL) injuries*, which result in participation-restriction time <24 hours. This contrasts with older surveillance data, which typically examined only injuries resulting in ≥24 hours of participation-restriction time (and noted as *time-loss [TL] injuries*). The National Collegiate Athletic Association Injury Surveillance Program (NCAA-ISP)[6] has collected NTL injuries since 2009–2010; the National Athletic Treatment, Injury and Outcomes Network (NATION)[7] has also collected NTL injuries since its inception in 2011–2012. These methods have yielded findings suggesting that a substantial proportion of sports-related injuries in collegiate populations are NTL.[4]

The analysis of NTL injury data, however, poses a unique statistical challenge. Although injury-surveillance mechanisms that exclusively record TL injuries may contain multiple injury records per athlete, mechanisms that also record NTL injuries are more likely to demonstrate this phenomenon. This is commonly referred to as *observation clustering* (resulting in violation of the independence assumption) in the statistical literature and requires careful analytical consideration when measuring the burden of injury using effect estimates such as injury proportion ratios (IPRs). This problem, as it relates to IPRs estimated using sports injury data, was introduced by Knowles et al.[8,9] However, there has not yet been detailed discussion of it or of strategies to handle it in this context. We present, first, a short review of the statistical challenge posed by observation clustering in regard to NTL injuries and, second, a strategy for handling the challenge and making methodologically sound inferences.

Measuring the Burden of Injury Using Injury Proportion Ratios

In epidemiology, risk and incidence rate ratios or both are commonly used to quantify and compare differential morbidity and mortality between samples.[10] Analogously, injury rate ratios (IRRs) and IPRs are commonly used effect estimates in sports injury surveillance. To illustrate the

abovementioned analytical challenge posed by NTL injury clustering, we will focus on IPRs in this manuscript:

$$IPR = \frac{\left(\frac{\text{\# of Specific Injuries in Group X}}{\Sigma \text{ All Injuries in Group X}}\right)}{\left(\frac{\text{\# of Specific Injuries in Group Y}}{\Sigma \text{ All Injuries in Group Y}}\right)}$$

and is constructed as an estimation of differential risk based solely on a sample of injured individuals. The motivation for using the IPR in this context and the method for estimating it has been discussed previously.[8] Briefly, as seen in the equation, the ratio uses frequencies of injury observations in the effect estimation. A ratio >1 would imply that a higher proportion of a given type of injury was observed in group X compared with group Y. Similarly, a ratio <1 would imply that a lower proportion of a given type of injury was observed in group X compared with group Y. A simple application of this measure was demonstrated by Deits et al,[11] who used the IPR to illustrate sex differences in proportions of observed head injuries, facial injuries, concussions, etc, among ice hockey players. The proportion of injuries to the head was higher in female players than in male players (IPR = 2.22; 95% confidence interval [CI] = 1.78, 2.77).[11]

We note that it is not the effect estimate that is influenced by clustering, as the formula is based on a count of injuries. Instead, as we will discuss, it is the estimation of the standard error (SE) of the IPR and any inference made using it that is directly biased when clustering is present.

Clustering

When a set of injury observations is analyzed to estimate IPRs, it is inherently assumed that the observations are independent. In other words, standard analytical procedures are accompanied by an unstated understanding that they are applied to a set of distinct cases. This understanding is associated with a more critical assumption that the errors attached to any effect estimate are unrelated,[10] which directly relates to the process of computing the SEs associated with effect estimates. When multiple injury observations are linked to the same athlete, this assumption is clearly violated.

It has been suggested[9] that traditional estimates of the SE are still appropriate in this context, when clustering is not pervasive within a given set of records.[9] However, the inclusion of NTL injury reporting in surveillance increases the likelihood of observation clustering. It may be more common to observe >1 injury per athlete when records include both TL and NTL injuries rather than TL injuries only. Nonetheless, in circumstances of observation clustering, the fundamental condition associated with standard analytical procedures—that the errors are unrelated—no longer holds true. In these cases, the SE accompanying any effect estimate such as an IPR is biased.[10] As SEs attached to an estimate are routinely used to construct 95% CIs around the estimate and these CIs often are the basis for inferences regarding statistical significance (whether or not the null value of 1.00 is contained in the interval), failing to acknowledge the violation of the independence assumption can lead to flawed inferences. However, statistical techniques are available that can be used to address this problem.

Sandwich Estimator

A robust method for handling the clustering phenomenon is employment of the sandwich covariance estimator.[10,12,13] This technique is typically used to handle misspecification of model covariance, which is a consequence of the independence assumption being violated.[10,12,13] Compared with standard methods for computing model covariance, the sandwich estimator uses an adapted method that incorporates empirical data[10,13] and consequently adjusts the SE estimates. Practically, this technique is employed in statistical analyses that use a unique identifier to denote a given participant, which is then repeated each time an observation (an injury record in this context) in the dataset corresponds to the same participant. Ultimately, the data-collection and -management protocols would only need to ensure that the included athletes could be distinguished. Such identification could be done without compromising the deidentified nature of the data (eg, randomly generated alphanumeric expressions). However, for a number of surveillance systems that focus on data analysis at the aggregate level and not the individual level, such identification may not be feasible. At a minimum, such studies may benefit from disclosure of this limitation, noting the data were analyzed under the assumption of independence.

Simulations

To illustrate the effect of the sandwich estimator, we simulated separate datasets of NTL injuries (each containing 1000 NTL injury records) with various degrees of observation clustering. That is, we created datasets with the following amounts of clustering: no clustering (0%), low clustering ($\sim$25%), moderate clustering ($\sim$50%), considerable clustering ($\sim$75%), and high clustering ($>$95%). Thus, the proportion of observation clustering was used as a direct indication of the number of unique athletes who contributed injury records to each dataset. In each scenario, we then computed IPRs (using log-binomial regressions) to estimate sex differences in injury proportions across levels of event type (games and practices) and injury mechanisms (player-to-player contact, player contact with surface or equipment, and noncontact or overuse). In these IPRs, female participants served as the referent group. This approach is used extensively in descriptive epidemiology of sport-related injuries.[4,8,14–17]

For the estimated IPRs, we computed SEs and CIs using 2 methods. We first computed the SEs and CIs assuming that all records were independent and then repeated the computation while accounting for observation clustering using the sandwich estimator to produce the sandwich SEs (SSEs) and corresponding CIs. This simulation process was repeated 1000 times for each clustering scenario. Details regarding our simulation procedures and the code used for analysis may be found in the Appendices.

Results

A summary of results from our simulations is presented in the Table. The averages presented in this table were computed from 1000 simulated datasets under each of the clustering conditions described earlier. As previously mentioned, the estimates of the IPRs themselves should

Table. Summary of Results From Simulations Under Various Clustering Scenarios[a]

Variable Clustering	Average Injury Proportion Ratio Comparing Male Athletes With Female Athletes[b]	Average Standard Error[c]	Average Sandwich Standard Error[d]	Agreement Proportion[e]
Event type				
Game				
No	0.999	0.063	0.063	1.00
Low	0.997	0.064	0.077	0.95
Moderate	0.998	0.064	0.089	0.89
Considerable	0.996	0.064	0.098	0.85
High	0.994	0.064	0.102	0.83
Practice				
No	1.001	0.063	0.063	1.00
Low	1.003	0.063	0.077	0.96
Moderate	1.002	0.063	0.089	0.89
Considerable	1.004	0.064	0.098	0.85
High	1.006	0.064	0.103	0.83
Injury mechanism				
Player-to-player contact				
No	0.998	0.090	0.090	1.00
Low	0.995	0.090	0.109	0.95
Moderate	0.996	0.090	0.126	0.90
Considerable	0.992	0.090	0.139	0.84
High	0.998	0.090	0.145	0.81
Player contact with surface/equipment				
No	0.999	0.089	0.089	1.00
Low	0.997	0.090	0.109	0.95
Moderate	0.997	0.090	0.126	0.88
Considerable	1.006	0.090	0.138	0.84
High	0.998	0.090	0.145	0.84
Noncontact/overuse				
No	1.003	0.090	0.090	1.00
Low	1.008	0.090	0.109	0.94
Moderate	1.008	0.090	0.126	0.88
Considerable	1.001	0.090	0.138	0.85
High	1.004	0.090	0.146	0.85

[a] Clustering descriptor used as an indication of the number of unique athletes that contributed injury records to each dataset as follows: no clustering (0%), low clustering (~25%), moderate clustering (~50%), considerable clustering (~75%), and high clustering (>95%).

[b] Average injury proportion ratios computed from 1000 simulations containing 1000 non–time-loss injury records (1 injury proportion ratio estimated per simulated dataset).

[c] Average standard errors estimated from 1000 simulations using traditional techniques assuming independence.

[d] Average standard errors estimated from 1000 simulations using the sandwich estimator (correcting for violations of independence assumption).

[e] Comparison of agreement in rejecting the null based on 95% confidence intervals estimated using traditional technique and sandwich estimator.

not change drastically, as evidenced in the Table. However, we draw attention to the average estimated SE and SSE. Minimal changes in the SE occurred with different levels of clustering. Yet as the levels of clustering increased, so did the SSE. The differences between the SSE and SE increased as well. For example, while examining sex differences in the proportions of game-related injuries, we saw that the average SE (not appropriately accounting for clustering) ranged between 0.063 and 0.064, whereas the average SSE (appropriately accounting for clustering) ranged between 0.063 and 0.102. Moreover, we observed greater differences, on average, between the 2 SE estimations as the degree of clustering increased (Figures 1 and 2).

To highlight the effect of the differences between SEs on inferences related to statistical significance, we present an *agreement proportion* in the Table. This is a measure of the number of times (out of 1000) that an agreement was noted between the SE-based and SSE-based decisions with respect to the null hypothesis that the proportions for males and females would not be different. For example, while comparing sex differences in game-related injuries from 1 simulated dataset with high clustering (>95%), the SE-based 95% CI (0.75, 0.97) was constructed around an IPR of 0.85, while the SSE-based 95% CI around the same IPR was constructed as (0.69, 1.05). In an applied context, using the sandwich estimator results in failure to reject the null hypothesis, whereas the standard approach results in sufficient evidence to reject the null. Consequently, inferential disagreement was observed in this simulated dataset. As the degree of clustering increased, so did the proportion of disagreement in our simulations (Table).

Summary

Injury proportion ratios are a powerful method for comparing differential injury prevalence between groups.[8] However, the methods relied up on to draw inferences
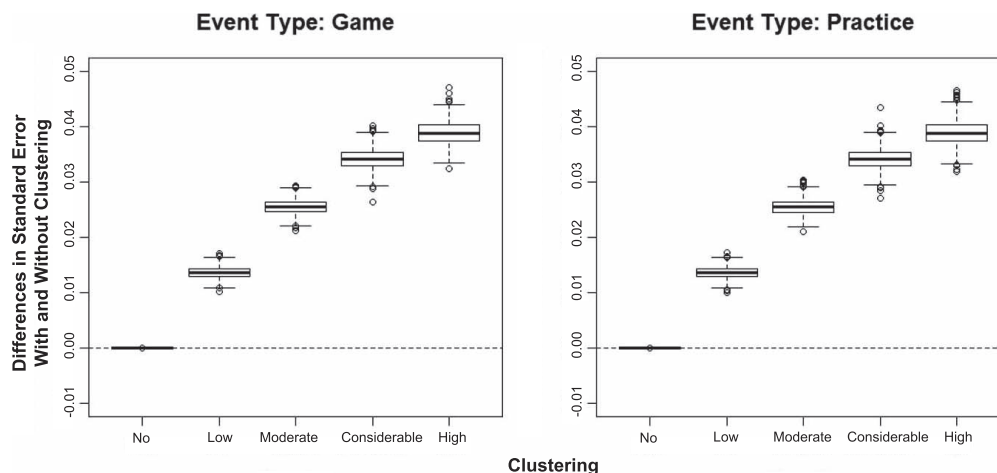
Figure 1. Differences between traditionally estimated standard errors and sandwich standard errors over varying degrees of observational clustering for event type.

related to an estimated IPR are predicated on the aforementioned assumption of independence. As presented here, the failure to handle violations of this assumption can lead to biased estimates of the SE and consequently flawed inferences. Although observation clustering and resultant violations of the independence assumption are difficult to avoid in injury surveillance, we present here a technique for handling such violations. The sandwich covariance estimator has been shown to be robust to similar misspecifications of model covariance.[10,12,13] We discuss here the specific application of this technique to injury surveillance and illustrated its value in this context. Although our discussions have been primarily limited to NTL injuries, we acknowledge the possibility that any combination of NTL and TL injuries may be clustered within a given set of surveillance data. As such, we note that investigators may apply the sandwich estimator as demonstrated here, even while conducting comparative analyses on consolidated sets of clustered TL and NTL injury observations.

Knowles et al stated[9] that the concern about correcting for clustering may be minimal if the average number of injuries per injured athlete is low. However, we urge investigators to use the sandwich estimator even when no clustering is suspected. Our results indicate that the SSE estimations are

identical to the standard or traditional SE estimations in cases of no clustering. We note that this is consistent with the statistical theory surrounding the sandwich estimator, and this property has been described in the context of presenting the mathematical derivation of the estimator.[10] Moreover, although the differences in SE estimations may seem negligible out of context, it is important to consider the inferential disagreement observed in our simulations. Inappropriate handling of even moderate clustering resulted in flawed inferences in 10% to 12% of our simulations. The degree of bias in the estimated SEs and the resultant flawed inferences depend on the extent of clustering as well as the nature of clustering. That is, not only the proportion of clustered observations (as a fraction of all observations) within a dataset but also the number of observations contributed by participants with multiple observations directly affect the observed results. Thus, we encourage investigators to carefully consider observation clustering in order to protect against inflated likelihoods of type I error. This will better inform clinicians using the literature as a foundation for evidence-based practice. Ultimately, the application of this technique may depend on the availability of requisite data (ie, unique identifiers for participants). However, as false rejections of the null hypothesis are
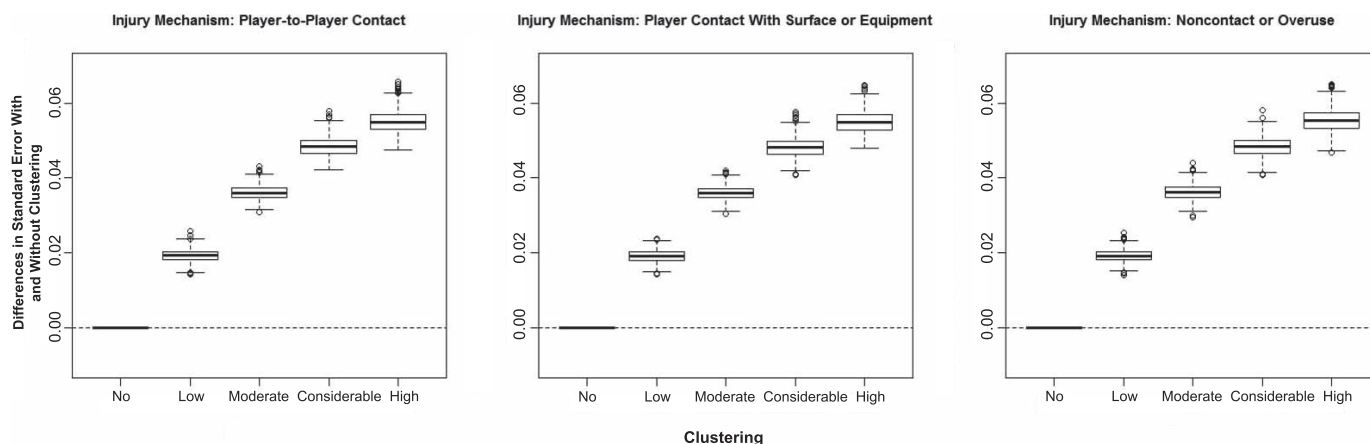

Figure 2. Differences between traditionally estimated standard errors and sandwich standard errors over varying degrees of observational clustering for injury mechanism.

generally considered to be egregious errors in observational and experimental science, it is advisable to use methods to minimize such errors whenever possible.

REFERENCES

1. Dick R, Putukian M, Agel J, Evans TA, Marshall SW. Descriptive epidemiology of collegiate women's soccer injuries: National Collegiate Athletic Association Injury Surveillance System, 1988–1989 through 2002–2003. *J Athl Train*. 2007;42(2):278–285.

2. Dick R, Sauers EL, Agel J, et al. Descriptive epidemiology of collegiate men's baseball injuries: National Collegiate Athletic Association Injury Surveillance System, 1988–1989 through 2003–2004. *J Athl Train*. 2007;42(2):183–193.

3. Dick R, Ferrara MS, Agel J, et al. Descriptive epidemiology of collegiate men's football injuries: National Collegiate Athletic Association Injury Surveillance System, 1988–1989 through 2003–2004. *J Athl Train*. 2007;42(2):221–233.

4. Roos KG, Wasserman EB, Dalton SL, et al. Epidemiology of 3825 injuries sustained in six seasons of National Collegiate Athletic Association men's and women's soccer (2009/2010–2014/2015). *Br J Sports Med*. 2017;51(13):1029–1034.

5. Marshall SW, Covassin T, Dick R, Nassar LG, Agel J. Descriptive epidemiology of collegiate women's gymnastics injuries: National Collegiate Athletic Association Injury Surveillance System, 1988–1989 through 2003–2004. *J Athl Train*. 2007;42(2):234–240.

6. Kerr ZY, Dompier TP, Snook EM, et al. National Collegiate Athletic Association Injury Surveillance System: review of methods for 2004–2005 through 2013–2014 data collection. *J Athl Train*. 2014;49(4):552–560.

7. Dompier TP, Marshall SW, Kerr ZY, Hayden R. The National Athletic Treatment, Injury and Outcomes Network (NATION): methods of the surveillance program, 2011–2012 through 2013–2014. *J Athl Train*. 2015;50(8):862–869.

8. Knowles SB, Kucera KL, Marshall SW. Commentary: the injury proportion ratio. What's it all about? *J Athl Train*. 2010;45(5):475–477.

9. Knowles SB, Marshall SW, Guskiewicz KM. Issues in estimating risks and rates in sports injury research. *J Athl Train*. 2006;41(2):207–215.

10. Lachin JM. *Biostatistical Methods: The Assessment of Relative Risks*. New York, NY: John Wiley & Sons; 2011.

11. Deits J, Yard EE, Collins CL, Fields SK, Comstock RD. Patients with ice hockey injuries presenting to US emergency departments, 1990–2006. *J Athl Train*. 2010;45(5):467–474.

12. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13–22.

13. Royall RM. Model robust confidence intervals using maximum likelihood estimators. *Int Stat Rev*. 1986;54(2):221–226.

14. Yard EE, Comstock RD. Injuries sustained by pediatric ice hockey, lacrosse, and field hockey athletes presenting to United States emergency departments, 1990–2003. *J Athl Train*. 2006;41(4):441–449.

15. Rechel JA, Yard EE, Comstock RD. An epidemiologic comparison of high school sports injuries sustained in practice and competition. *J Athl Train*. 2008;43(2):197–204.

16. Collins CL, Comstock RD. Epidemiological features of high school baseball injuries in the United States, 2005–2007. *Pediatrics*. 2008;121(6):1181–1187.

17. Yard EE, Comstock RD. Injuries sustained by rugby players presenting to United States emergency departments, 1978 through 2004. *J Athl Train*. 2006;41(3):325–331.

Address correspondence to Avinash Chandran, PhD, MS, Matthew Gfeller Sport-Related TBI Research Center, Department of Exercise and Sport Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-8700. Address e-mail to avinashc@email.unc. edu.