Analysing recurrent events in exercise science and sports medicine

Ian Cook (BA (Phys Ed) Hons, BSc (Med) Hons)

Physical Activity Epidemiology Laboratory, University of Limpopo (Turfloop Campus), Polokwane

Abstract

Episodic or recurrent events are a class of data that is frequently reported in health sciences research. The purpose of this paper is to highlight the prevalence of published reports, especially within the South African context, that have used inappropriate statistical techniques when dealing with episodic events and to urge the use of appropriate univariate and multivariate techniques.

Introduction

Episodic or recurrent events are a class of data that is frequently described in sports medicine literature. However, the correct statistical techniques to deal with data containing recurrent events are not widely known within sports medicine and the exercise sciences. This is evidenced by the few papers in these specialist sciences that discuss the use of appropriate statistical techniques[1,2] and the preponderance of papers assuming event independence for recurrent events. For instance, in a recent paper[3] it is apparent that there is a trend in studies reporting injury incidences in rugby union players that need to be highlighted, namely the use of naïve statistical methods that treat recurrent events as independent observations. A number of references are cited (see[3] ref. 2, 11, 15-17) that also report injury incidence statistics in rugby union players, and as far as can be ascertained, treat recurrent or multiple injuries within the same individual as independent events. The purpose of this paper was first, on the basis of an example from the sports medicine literature, to contrast the effect of recurrent events on confidence intervals generated with unadjusted and adjusted univariate statistical techniques. Second, to demonstrate the implementation of a multivariate regression technique on data containing recurrent data and confounding variables, using data from the exercise sciences. Third, the use of two disparate examples should dispel the notion that the statistical techniques highlighted in this paper have limited application.

CORRESPONDENCE:

Ian Cook
Physical Activity Epidemiology Laboratory
University of Limpopo (Turfloop Campus)
PO Box 459
Fauna Park
0787 Polokwane
South Africa
Tel+fax: +27 15 268 2390
E-mail: ianc@ul.ac.za

Statistical concepts and considerations

For the purposes of this paper it is important to note that whether the injury occurs in the same or different anatomical structure does not influence how the event is considered in statistical terms; it is a recurrent event within the same individual. Consequently, even if the unit of analysis or outcome of interest is the injury count, the injury counts are clustered around the individual player. Injuries that occur in the same individual but at different anatomical sites can be correlated either through the mechanism of injury or via a common risk factor(s) to which the individual is exposed. Clustering can also occur at group level, for example school or team.[2] Importantly, whether clustering occurs at individual or group level, and the data are continuous, binary or count, appropriate univariate, non-model-based (e.g. *t*-test) and multivariate, model-based (e.g. regression) techniques are available that correct for clustered or correlated data.[2,4,5] Appropriate multivariate techniques adjust not only for confounders, but also for event dependence.[5] Moreover, for injuries at different anatomical sites in the same individual, a categorical variable can be created by grouping the different anatomical sites so that the risk for injury at different anatomical sites can be assessed adjusting for confounders and event dependence.[5] Whether the investigator has used univariate or multivariate statistical methods, it is essential to use appropriate formulae and statistical techniques to account for the increased variance that these recurrent events will have on the standard error and thus the confidence intervals (CI) of point estimates such as incidence rates (IR) and incidence rate ratios (IRR). Not doing so will result in artificially narrow CI.

If investigators are using the non-overlap of 95%CI to infer significant differences between IR, the adjustment for increased variance due to recurrent events is critical to avoid type I errors. Constructing adjusted 95%CI for univariate age-specific or age-adjusted rates can be implemented in a spreadsheet,[6] although it is recommended that suitable multivariate statistical techniques are invoked when analysing data sets with recurrent events.[1,2,5-12] Naïve statistical techniques either treat recurrent events as uncorrelated, or to avoid recurrent events only use the first event and ignore the subsequent events. In the former case, the CI are artificially narrow, in the latter case much information is lost. Appropriate statistical techniques include generalised estimating equations, survival analysis (Cox proportional hazards regression with robust variance estimation) and regression for count outcomes data (Poisson or negative binomial models with robust variance estimation).[13] Statistical software packages such as SAS, SPSS and Stata are required to implement these multivariate techniques. Importantly, the robust variance estimation yields IRR with unbiased 95%CI. Moreover, these are multivariate techniques which allow for the adjustment of relevant covariates and determination of risk for sub-groups.

Which multivariate technique to use will also be influenced by aspects such as whether the events are short or long lasting,

and whether the events occur at predefined intervals (recurring treatments in randomised controlled trials), or on a continuous basis (injuries or hospitalisation).[9] Also, data structure requirements can differ between techniques – multiple rows per person or one row per person.[9] If the recurrent events display event dependence (subsequent events are more or less likely to occur) and there is heterogeneity across individuals (cases with higher or lower event rates due to unaccounted for effects) then more complex models are required and statistical advice should be sought.[12]

The present discussion does not suggest that univariate techniques must be abandoned because statistical corrections are available for dealing with recurrent events and confounding.[2,5] What is being advocated in this paper is that researchers should consider the use of multivariate techniques which are more efficient than univariate techniques for datasets containing recurrent events and confounding variables.[5] Hence, statistical power is increased when using appropriate multivariate techniques in the presence of event recurrence and confounding.

Practical applications

Example 1: Sports medicine

It would appear from the methodological descriptions in Viljoen *et al*.[3] and the studies that they cited that univariate statistical techniques, which assume group independence,[14] were used to compare IR across two or more years or between training and match play (chi-square test for trend, z-test), and to construct crude IR 95%CI. In so doing, these studies have likely violated the statistical principle of independence of events to a greater or lesser degree, depending on the number of recurrent events. It is evident from Table I in their paper that there are recurrent events not only in the persistent injuries but also in the new injuries.[3] For example, from 38 injuries and 300 person-hours accumulated in the 2002 season (Table II),[3] the crude IR 95%CI is reported as 126.7 injuries per 1 000 person-hours (91.2 - 169.7 injuries per 1 000 person-hours). However, using standard statistical software (Stata/SE 11.0 for Windows, StataCorp LP, Texas, USA, 2009), the Poisson exact or Fisher's exact 95%CI is 89.6 - 173.9 injuries per 1 000 person-hours. If one assumes the new injuries (N=38) are evenly distributed in the 19 injured players during the 2002 period (Table II),[3] then there are 2 injuries per player. Once the increased variance has been taken into account, the crude IR 95%CI widens to 71.5 - 181.9 injuries per 1 000 person-hours (ideally the method employed here should be used for N>50).[6] Assuming that of the 19 injured players, 5 players have 3 injuries, 5 players have 1 injury and the remaining 9 players have 2 injuries each, the crude IR 95%CI widens even further; 67.7 - 185.6 injuries per 1 000 person-hours. It is evident that increasing recurrences have significant effects on the CI.

Example 2: Exercise science

Unpublished minute-by-minute, uni-axial accelerometry data (1 - 7 days) were collected in 263 rural and 16 urban women. The variable of interest was the number of bouts of ≥10 min of continuous moderate-to-vigorous activity the women accumulated (≥1 952 counts. min^{-1}). The question was whether urban women have greater odds of accumulating bouts of moderate-to-vigorous activity compared with rural women. Crude IR for the rural and urban women were 22.8 bouts per 1 000 person-hours and 31.9 bouts per 1 000 person-hours, respectively, and 170 women recorded more than one bout of moderate-to-vigorous activity. Using standard methods, which assume event independence, for calculating exact Poisson IR 95%CI yielded 21.2 - 24.4 bouts per 1 000 person-hours and 24.3 - 41.2

bouts per 1 000 person-hours, for rural and urban women respectively. Correcting for the increased variance due to episodic events by univariate means,[6] the IR for rural and urban women widened to 18.7 - 26.8 bouts per 1 000 person-hours and 8.3 - 55.5 bouts per 1 000 person-hours, respectively. A simple Poisson regression model, treating all the events as independent, produced an IRR of 1.40 (p=0.012, 95%CI: 1.08 - 1.83). On the basis of this superficial analysis we would conclude that urban women are significantly more likely (1.4-fold) to accumulate continuous bouts of moderate-to-vigorous activity, compared with rural women. However, by accounting for the recurrent events within individuals, the point estimate was no longer significant (IRR=1.40, p=0.281, 95%CI: 0.76 - 2.59). By extending the analysis and adding age, body mass index and subsistence level as covariates, while retaining the robust variance estimation option, the IRR increased to 1.80 (p=0.042, 95%CI: 1.02 - 3.16). We can now report that all reasonable analyses have been conducted on the dataset and can conclude that urban women are statistically more likely to accumulate bouts of continuous moderate-to-vigorous activity compared with rural women, adjusting for covariates.

Summary

Investigators reporting data which include recurrent events are urged to employ appropriate univariate and multivariate statistical techniques. Ignoring the valid methods available[1,2,5-12] can lead to conclusions being drawn which are at odds with the data.[9] Moreover, South African injury incidence data that have been analysed and reported, using naïve statistical methods, could be re-analysed using these univariate and multivariate statistical techniques and provide a more thorough understanding of the associated risks.

REFERENCES

1. Knowles SB, Marshall SW, Guskiewicz KM. Issues in estimating risks and rates in sports injury research. J Athl Train 2006;41:207-215.

2. Hayen A. Clustered data in sports research. J Sci Med Sport 2006;9:165-168.

3. Viljoen W, Saunders CJ, Hechter GD, Aginsky KD, Millson HB. Training volume and injury incidence in a professional rugby union team. S Afr J Sports Med 2009;21:97-101.

4. Ying G, Liu C. Statistical analysis of clustered data using SAS system. Available at: http://nesug.org/proceedings/nesug06/an/da01.pdf. Accessed July 2010.

5. Glynn RJ, Buring JE. Ways of measuring rates of recurrent events. BMJ 1996;312:364-367.

6. Stukel TA, Glynn RJ, Fisher ES, Sharp SM, Lu-Yao G, Wennberg JE. Standardized rates of recurrent outcomes. Stat Med 1994;13:1781-1791.

7. Kuramoto L, Sobolev B, Donaldson M. On reporting results from randomized controlled trials with recurrent events. BMC Med Res Methodol 2008;8:35.

8. Sturmer T, Glynn RJ, Kliebsch U, Brenner H. Analytic strategies for recurrent events in epidemiologic studies: background and application to hospitalization risk in the elderly. J Clin Epidemiol 2000;53:57-64.

9. Twisk JWR, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. J Epidemiol Comm Health 2005;59:706-710.

10. Thomsen JL, Parner ET. Methods for analysing recurrent events in health care data. Examples from admissions in Ebeltoft Health Promotion Project. Fam Pract 2006;23:407-413.

11. Gill DP, Zou GY, Jones GR, Speechley M. Comparison of regression models for the analysis of fall risk factors in older veterans. Ann Epidemiol 2009;19:523-530.

12. Box-Steffensmeier JM, De Boef S. Repeated events survival models: the conditional frailty model. Stat Med 2006;25:3518-3533.

13. Juul S. An Introduction to Stata for Health Researchers. 2nd ed. Texas: Stata Press, 2008.

14. Armitage P, Berry G, Matthews JNS. Statistical Methods in Medical Research. 4th ed. Massachusetts: Blackwell Publishing, 2002.