Statistical modelling for recurrent events: an application to sports injuries

Shahid Ullah,[1] Tim J Gabbett,[2,3] Caroline F Finch[4]

[1]Flinders Centre for Epidemiology and Biostatistics, Faculty of Health Sciences, Flinders University, Adelaide, South Australia, Australia
[2]School of Exercise Science, Australian Catholic University, Brisbane, Queensland, Australia
[3]School of Human Movement Studies, The University of Queensland, Brisbane, Queensland, Australia
[4]Australian Centre for Research into Injury in Sport and its Prevention (ACRISP), Monash Injury Research Institute (MIRI), Monash University, Melbourne, Victoria, Australia

Correspondence to
Professor Caroline F Finch, Australian Centre for Research into Sports Injury and its Prevention (ACRISP), Monash Injury Research Institute (MIRI), Building 70, Monash University Clayton Campus, Melbourne, VIC 3800, Australia; caroline.finch@monash.edu

ABSTRACT

Background Injuries are often recurrent, with subsequent injuries influenced by previous occurrences and hence correlation between events needs to be taken into account when analysing such data.

Objective This paper compares five different survival models (Cox proportional hazards (CoxPH) model and the following generalisations to recurrent event data: Andersen-Gill (A-G), frailty, Wei-Lin-Weissfeld total time (WLW-TT) marginal, Prentice-Williams-Peterson gap time (PWP-GT) conditional models) for the analysis of recurrent injury data.

Methods Empirical evaluation and comparison of different models were performed using model selection criteria and goodness-of-fit statistics. Simulation studies assessed the size and power of each model fit.

Results The modelling approach is demonstrated through direct application to Australian National Rugby League recurrent injury data collected over the 2008 playing season. Of the 35 players analysed, 14 (40%) players had more than 1 injury and 47 contact injuries were sustained over 29 matches. The CoxPH model provided the poorest fit to the recurrent sports injury data. The fit was improved with the A-G and frailty models, compared to WLW-TT and PWP-GT models.

Conclusions Despite little difference in model fit between the A-G and frailty models, in the interest of fewer statistical assumptions it is recommended that, where relevant, future studies involving modelling of recurrent sports injury data use the frailty model in preference to the CoxPH model or its other generalisations. The paper provides a rationale for future statistical modelling approaches for recurrent sports injury.

INTRODUCTION

Sports injuries are often recurrent in that some people experience more than one sports injury over time. There is wide recognition that subsequent injury (of either the same or a different type) can be strongly influenced by previous injury occurrences.[1–4] Such recurrent injuries are unlikely to be statistically independent, and appropriate statistical methods need to be used to analyse such data accurately.[5–8] While different modelling approaches have been used to report recurrent event data, such as modelling the within-person total number of events or time to the first event, they have often been naïve in the statistical sense in that they do not take correlation between events into account or have excluded important detailed information about the subsequent events.[9] Over the last decade, there have been some significant statistical advances in the modelling of recurrent event data.[7] [10–12] While there has been some application to health data,[9] [13] these methods are yet to be reported in sports medicine applications. This means that many models of the likelihood of recurrence of sports injury, or for understanding causal relationships when conditions can be recurrent, could be flawed, leading to incorrect information being used to inform prevention priorities and programmes.

A key statistical challenge inherent in analysing recurrent injury data is that the probability of injury occurrence is likely to be influenced by earlier injuries, even when they are not of exactly the same type; this can be manifest as an injury either raising or lowering the rate of further injury. This is important because analyses that incorrectly treat different within-person injuries as statistically independent run the risk of generating misleading results. Ignoring potential within-person event dependency leads to reported greater precision than is warranted and possible biasing of results away from the null. A second statistical issue is that many naïve statistical approaches implicitly restrict the baseline probability of injury, and the influence of covariates on this, to be the same across all injuries when, in fact, they vary across people and different injury types. Across people, this variability implies that some will have inherently higher or lower rates of different subsequent injuries. Together, these statistical issues mean that in any recurrent injury dataset there will be different within-person correlations across people and that the within-person injury times will be dependent. Any correlation among injuries (whether produced by event dependence or variability) will violate assumptions that the timing of injuries is independent, and result in problems of estimation and incorrect inference if not properly taken into account.

Despite many studies documenting the incidence of sports injuries, and recognition of the recurrent nature of many injuries,[14] appropriate statistical modelling for recurrent sports injuries has largely been absent from published studies. In general, subsequent sports injury has been handled statistically in one of three ways. The majority of cohort studies have reported Poisson counts and calculated injury rates as the total number of injuries per unit time, even when many players contribute more than one injury occurrence to the numerator. Inherently, such calculations treat all injuries within given players as independent. When these studies have recognised that injury history can predict injury risk, they have adjusted for it in regression models by including a dichotomous predictor representing 'previous injury history? (yes/no)'. On the rare occasion when researchers have recognised within-player injury dependency, they have only modelled the time to first injury and have excluded

valuable information about any subsequent injuries from consideration.[15][16] To progress recurrent sports injury epidemiology, there is a need for guidance in the most appropriate statistical models for these data.

Several event history model variations based on the Cox proportional hazards (CoxPH) model[17] have been proposed for the analysis of repeated events but their application leads to different results because of the different assumptions they make about the data they are modelling.[5][12][18–24] In practice, the choice of the most appropriate model depends upon the: (a) distribution of subsequent event times; (b) within-person correlation of subsequent events; (c) frequency of the recurrent events; and (d) the specific research question being posed at the time (eg, estimation of population-level effects of covariates as averaged across people or describing event dependency within people).[9][25] A major statistical consideration is therefore how to address both the players at risk and the subsequent injuries appropriately.

In the general recurrent event literature, extensions of the CoxPH model are popular because they enable all events for each individual to be analysed. Application of four prominent regression models (Andersen-Gill (A-G),[26] frailty,[27] Wei-Lin-Weissfeld total time (WLW-TT) marginal model,[28] Prentice-Williams-Peterson gap time (PWP-GT) conditional model[29]) yield different results because of their different underlying assumptions. To our knowledge, these models have not been previously applied and compared in sports injury epidemiological studies and so it is currently unknown which of these is the most suitable for modelling recurrent sports injuries.

The aim of this paper is to (a) summarise the issues that need to be considered when modelling recurrent sports injury data where the time before/between injuries is of interest and (b) to assess and compare the suitability of the CoxPH model and its extensions for modelling such data. The methods and model comparison are demonstrated on Australian National Rugby League (NRL) injury data to provide defensible guidance on how to appropriately model recurrent sports injury data.

METHODS
The data
To demonstrate and compare the applicability of different extensions of the CoxPH model to a real-world data example, injury data were obtained on 35 players from a professional rugby league club competing in the 2008 Australian NRL competition. Injury and participation data were collected from 29 matches (including all trial, fixture and finals matches). Injuries were defined as conditions associated with pain or disability that occurred during match participation, irrespective of the need for first aid, medical attention or time loss.[30] In the context of this paper, a recurrent injury was said to have occurred if a person sustained more than one injury over the 29 matches, irrespective of whether or not it was to the same body region or of the same type. For this paper, only data on all contact injuries (defined as those resulting from tackling, being tackled, collision and accidental contact) were extracted. All players received a clear explanation of the study, including the risks and benefits of participation and written consent was obtained. The study was approved by the University of Queensland Human Ethics Committee.

The models
The CoxPH model and four recurrent event generalisations were applied to the sports injury data. The time variable was taken to be the match number (range 1–29). Major statistical challenges with this sort of data are how to address the number of recurrent events and the number of players at risk appropriately. Four components were considered for the recurrent event model:[5] (a) risk interval which defines when a player is at risk of having an injury along a given timescale and determines whether a model is either marginal or conditional; (b) risk set or the number of players included in the set at a given point in time; (c) an event-specific or common baseline hazard; and (d) handling of within-subject correlation. Table 1 summarises how these components are defined for each of the four models considered in this paper.

Table 1 Statistical specifications and assumptions in relation to the risk interval, risk set, baseline hazard and within-person correlation in the extended Cox proportional hazards (CoxPH) models

Components	Andersen-Gill (A-G)	Frailty	Wei-Lin-Weissfeld total time (WLW-TT) marginal	Prentice-Williams-Petersen gap time (PWP-GT) conditional
Risk interval	Duration since starting observation	Duration since starting observation	Duration since starting observation	Duration since previous injury
Risk set for injury k at time t	Independent injuries (any given injury occurrence is not affected by previous injuries)	A random effect (or frailty) term is used to account for the within-player correlation between injuries to enable modelling of the phenomenon by which some players are intrinsically more or less susceptible to experiencing a given injury than others are	All players who have not experienced injury k at time t	All players who have experienced injury k−1, and have not experienced injury k at time t
Baseline hazard	Common/same baseline hazard across all injuries	Heterogeneity is directly incorporated via a random effect so that the baseline hazard is allowed to vary with each injury	Common baseline hazard for all injuries within a player	Stratifies the data by injury so that the baseline hazard is allowed to vary with each injury
Within-person correlation	The within-person injuries are independent	Captures within-person correlation due to both injury dependence and heterogeneity	The within-person injuries are independent	The current injury is unaffected by earlier injuries that occurred to the player
Comment	A-G model is recommended when there is no injury dependence and no covariate/injury effects	The frailty approach accounts for heterogeneity. The random effect (the frailty) has a multiplicative effect on the baseline hazard function and the mixture of individuals with different injury risks	At any time point (matches), WLW-TT describes all players who have not experienced k injuries are assumed to be at risk for the kth injury which is not realistic in the sports setting injury data	PWP-GT model takes into account the ordering of events

The following three risk intervals were considered: (a) gap (or interoccurrence) time representing the time from the prior injury event and not relative to the actual timeline of observation; (b) calendar time which uses the same timescale for all events, referenced to a fixed point in time, but does not allow an overlap in risk periods across events for a given player; and (c) total time representing the time from the start of the player follow-up. In each case, the interval ends with the current injury. In both the gap time and calendar time representations, the player is at risk for the same length of time. Gap and calendar time models are conditional since a player is at risk of a new injury, conditioned on having sustained a previous injury. For total time, the clock does not reset for each event and the beginning of each event is at the same point in the observation timeline; risk periods for different events for the same player overlap. Total time models are marginal since the player is at risk from the start of play, independent of any previous injury. Irrespective of the definition, the risk interval for the first injury is the same.

Figure 1 describes these risk intervals in more detail, through the specific examples of three players (figure 1A). In the gap time representation, after an injury event, the player resumes play again at time 0 and the time to the next event corresponds to the number of matches that it takes for that player to experience the next event. The occurrence of all events after the first is modelled on a timescale relative to the prior event and not relative to the actual timeline of observation. Thus, the gap time (figure 1B) for our example indicates that player A is at risk of his first injury during 0–2 matches, and of his second, third and subsequent injuries during 0–18 matches, 0–6 matches and 0–1 match, respectively. In the calendar time (figure 1C), player A is at risk for his first injury event during 0–2 matches, and his second, third and subsequent injuries during matches 2–20, 20–26 and 26–27, respectively. The total time (figure 1D) indicates that player A is at risk for his first, second and subsequent injuries during 0–2, 0–20, 0–26 and 0–27 matches, respectively.
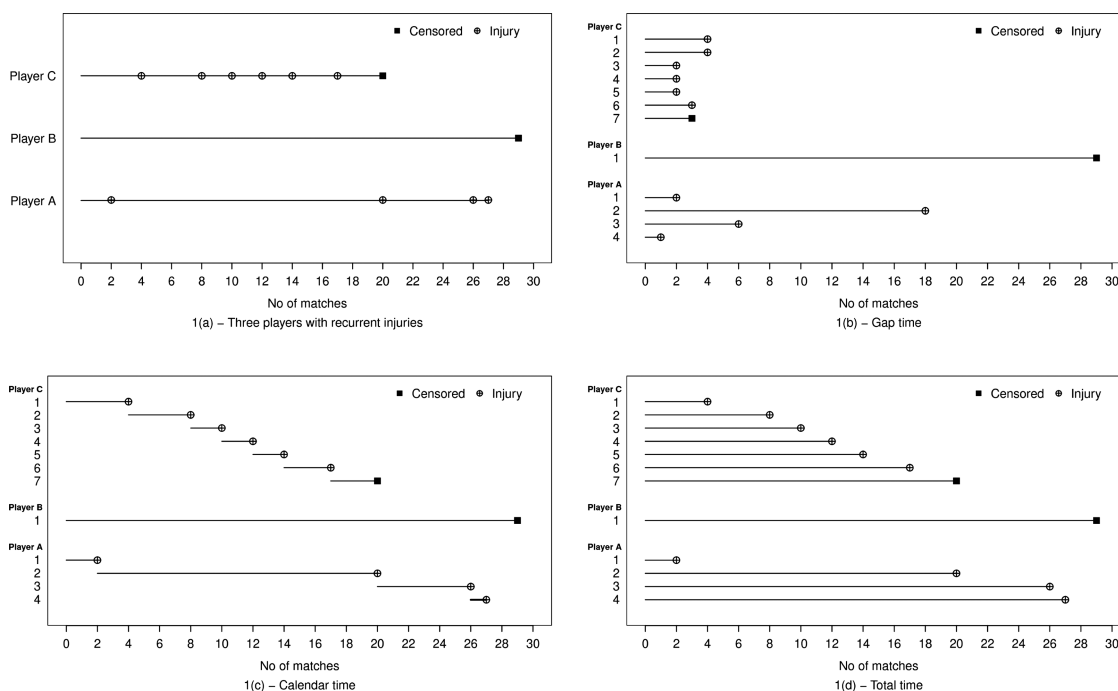
The Kaplan-Meier (K-M) method is used to estimate the survival function non-parametrically from observed (censored and uncensored) survival times.[31] The CoxPH model with time to first injury event as the outcome variable is a regression model used to estimate the survival probability after adjusting for both baseline hazard and explanatory variables. This model counts the players at risk at the time of this first event, after which they are no longer considered to be at risk. The result is an estimation of the probability of remaining free of injury for a given point in time based on the observed injuries. The steps in the K-M curves show changes in the probabilities of remaining free of injury for various matches across the group of players, when first injuries occur in new players.

The A-G model is a simple extension of the CoxPH model where players contribute to the risk set for an event as long as they are under observation at the time the injury occurs and share the same baseline hazard function. However, the A-G model requires the strongest statistical assumptions including that of an independent increment in which any given injury occurrence is not affected by previous injuries, that is within players, injuries are independent. This restriction means that injury dependence cannot be included and the A-G model inherently assumes that injuries do not change the player and that the player does not learn from previous injuries. Moreover, this model does not allow investigation of effects that might change based on injury-specific covariate effects, but there is the possibility of incorporating injury dependence via time-dependent covariates. Given these limitations, the A-G model is recommended when there is no injury dependence and no covariate/injury effects.

Analysis of recurrent injury data frequently assumes that the player injury histories are all statistically independent (at least conditionally on observed time-fixed covariates) so that the interoccurrence times appear in an independent manner. However, some players are intrinsically more or less susceptible to experiencing an injury than others. The frailty model is

Figure 1 Illustrations of the risk interval formulations: (A) three players with recurrent injuries; (B) gap time; (C) calendar time; (D) total time. A circle (●) indicates an injury event and a solid square (■) indicates censoring. Each time to an event or censoring is a separate risk interval.

characterised by its inclusion of a random effect, or frailty, term can account for the within-player correlation between injuries. If the frailty is less than 1, a player tends to experience the injury at a later time than another player, whereas the opposite occurs if the frailty is greater than 1.

The WLW-TT model is a marginal model and assumes a common baseline hazard for all injuries within a player. Marginal models consider the marginal distribution of each failure time and impose no particular structure of dependence among distinct failure times on each player. Each recurrence is modelled as a different stratum and each stratum is treated as marginal data. This model is marginal with respect to the risk set since each player is at risk from the beginning of the study and can be at risk for several events simultaneously.

The PWP-GT model is a conditional model which allows for event dependence via stratification by event number so that different events can have different baseline hazards. The main difference to the marginal model is that a player cannot be at risk for the later injury until a prior event occurs. This conditional model preserves the order of sequential injuries in the creation of the risk set and therefore incorporates injury dependence. The PWP-GT model is estimated with the data organised in interoccurrence/gap time (ie, gap time risk set or time since the previous injury).

Model estimation and evaluation

The outcome being modelled is the probability of remaining injury-free over the 29 matches. As shown in table 1, different model formulations handle the time variable in relation to injury occurrence differently. All models were fitted using the *cph* function of the Design package within R (Version 2.12.2).[32][33] The strata, cluster and frailty functions were used to fit the extended CoxPH models. The proportional hazard assumption test was performed using the *cox.zph* command. All models were adjusted by age, match experience and body mass of the players as known confounders of injury risk in NRL players. The R code is available from the authors, upon request.

K-M curve representations of the observed probability of remaining free of injury were used to provide a visual comparison of each model fit. The log likelihood (LL), Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare the goodness of fit of the fitted models in terms of fitting the observed data.[34][35] A lower AIC or BIC indicates a better fit to the observed data and two models can be compared by comparing the differences in the AIC or BIC, with preference being given to the model with the smallest criterion measure.[36] A simple rule of thumb is that models are not different if the difference in AIC is less than 2; there is minor evidence of difference when the AIC ranges from 2 to 4, and there is strong evidence for a difference with the AIC difference is more than 10. When comparing BIC, differences ≤2 are considered weak, those >2 but ≤6 are positive, those >6 but ≤10.0 are strong and BIC differences >10 are very strong.[37]

Model accuracy

The most common criterion for evaluating the performance of a statistical model is its accuracy in terms of data fit. In this sense, the model accuracy is an assessment of the closeness of estimates to the exact (or observed) value and can be computed on a point-by-point basis. The most widely used measures of accuracy are the mean-squared error (MSE), the root MSE, the mean absolute error and the mean absolute percentage error.[38] Smaller values of each of measure indicate more accurate and

reliable models. Further details about these measures can be found in Hyndman and Koehler.[39]

Comparing the models

Three test criteria were used to compare the fitted models: likelihood ratio (LRT), F[40] and bootstrap tests.[41][42] All tests compare two models where one model is an extension to the other (ie, the models are nested, with the simpler model being contained as a subset of the more complex one). For example, the A-G model is nested within the frailty model and comparison of the two models can test if there are random effects components for recurrent events that need to be modelled (as considered by the frailty model, but not the A-G model).

As an example, the LRT begins with a direct comparison of the likelihood scores of the two models and tests whether the frailty is necessary for analysing recurrent sports injury events. A significant LRT suggests that a random effect (frailty) accounts for the within-player correlation between injuries. A similar approach is used for the F-ratio test.

In the bootstrapping procedure, a large number of random samples are generated.[41] The observed test statistic is then compared with the test statistics calculated from the bootstrap samples. Although there are many ways to use the bootstrap for hypothesis testing, the method of Walters and Campbell[42] for computing a bootstrap p-value corresponding to the observed value of a test statistic T was used.

Simulation framework

Finally, a simulation approach for calculating the size and power of the models was undertaken.[43] One hundred sets of injury data were simulated from the exponential distribution and the bootstrapping procedure was applied 100 times to each generated dataset to obtain the significance level of the test. Within the context of model selection, power and size estimates are based on the proportion of replications that indicate acceptable fit, with greater numbers of replications resulting in smaller CIs (higher power, more accuracy) around the estimates. Simulations were run on a bi-processed Pentium 4 machine with a 3.20 GHz processor and 2.0 Gb RAM memory. The data-generating process was performed using the SimSurv function of the prodlim package[44] from R version 2.12.2,[33] operating on a Windows XP professional platform.

RESULTS

Overall, 47 contact injuries were sustained by the 35 players during a total of 557 player appearances. The median follow-up time was 18 matches (range 1–29 matches). More than half of the players (54.3%) sustained 1–6 injuries, with 40% sustaining >1 injury over the 29 matches (table 2). The most common site of injury was the head and neck (26% of all injuries). The incidence of injury was similar for the shoulder, thigh, calf and knee (13%). Sprains (32%), contusions (26%) and haematomas (17%) were the most common type of injury. The majority of injuries occurred while tackling or being tackled (47%).

Figure 2 shows the timing of the incidence of each injury event in relation to the total number of matches (each of 80 min or 1.33 h duration). Censoring sometimes occurred when a player (a) had not experienced the relevant outcome, by the end of the season; (b) was lost to follow-up; or (c) experienced a different event that made further follow-up impossible. The data structure shows the complex nature of the recurrent injuries in that, in some players, several injuries occurred and the time between injuries also differed across players.

Table 2 The distribution of number of injuries sustained by 35 National Rugby League players, the respective number of matches with a number of injuries and the injury incidence rates per 1000 matches

Number of injuries	Number of players	Total number of injuries	Proportion of players	Total number of matches	Injury incidence rates
0	16	–	45.7	134	–
1	5	5	14.3	107	46.7
2	7	14	20.0	133	105.3
3	2	6	5.7	55	109.1
4	4	16	11.4	108	148.1
5	–	–	–	–	–
6	1	6	2.9	20	300.0
Total	35	47		557	

Figure 3 shows the K-M survival curves as a means of comparing the CoxPH model and its generalisations. The K-M curves estimate the probability of remaining free of injury at a given point in time based on observed recurrences. All players were free of injury at the beginning of the season and survival rates were lower after every match, during which injuries occurred. Almost 95% of players were injured by the end of 29 matches. The Cox proportional regression model provided the best fit for the first few matches and the worst fit for the remaining matches. This is not surprising given that the CoxPH model only considers the time to the first injury. The fit was also relatively poor for the WLW-TT and PWP-GT models. The A-G[26] and frailty models provided the best fit for modelling recurrent sports injury data.

Table 3 shows the LL, AIC and BIC criteria for the fitted models. The AIC and BIC results provide strong evidence that both the A-G and frailty models perform better than the WLW-TT, PWP-GT and CoxPH models. Although the differences in LL were indistinguishable, the A-G and frailty models showed minor AIC differences but strong BIC differences in fitting the recurrent events.

Table 4 compares the fit of the five models to the injury data according to the four accuracy measures. On all measures, the CoxPH model had a poorer fit than each of its extensions. Although there was little difference in fit between the two, the frailty model performed a little better than the A-G model.
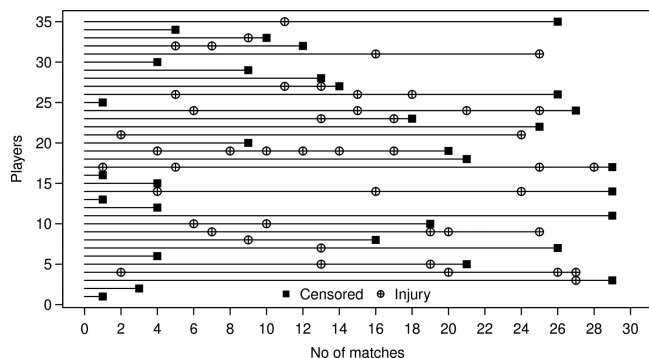

Figure 2 Recurrent injury history of 35 professional rugby league players. The event of interest is any contact-injury sustained by a player, which is denoted by a circle (○). Censored data which arise when the outcome injury status is either not-injured or unknown is denoted by solid squares (■).

The p-values from the pairwise LRT, F and bootstrap model comparison tests are shown in table 5. The estimated p-values comparing the A-G and frailty models, all being >0.05, show that these models are indistinguishable and either model could be used for analysing the recurrent sports injuries in this paper. There was no significant difference between the A-G and PWP-GT models but the A-G model was statistically different from the WLW-TT model. The CoxPH was significantly different only to the frailty model.

Table 6 shows that the bootstrap simulation tests were performed satisfactorily for each pair of models. However, the actual sizes and powers were slightly different from the simulated model sizes and powers. For example, the simulated model size superseded the actual size and simulated model power preceded the actual power at 10% level of significance, when the WLW-TT and PWP-GT models were considered.

DISCUSSION
Knowing how to choose the best model for analysing recurrent events in sports injury settings is important for the generation of accurate and reliable information to guide priority setting for targeting of intervention investments to tackle the sports injury problem. Although there are some guidelines on how to appropriately model injury count data,[45 46] little prior attention has been paid to the analysis of recurrent injury data. A recent conceptual model has described how and why recurrent injuries are a problem in the sports injury context, but gives no guidance on how to analyse such data.[14] Analysis of recurrent sports injury is complex and researchers interested in this are advised to collaborate with a statistician.

The need to correctly statistically model recurrent events occurs in many clinical trials, longitudinal epidemiological studies and sociological research.[13 47–51] Sports injury studies often report recurrent events because players can experience more than one injury event over a playing season.[14] Sports injury prevention is dependent on players' ability to tolerate repeated exposures to injury risks while being active in their sport. In terms of injury risk, it is likely that some of the risk factors for a subsequent injury will also be implicated in the initial injury. However, these injuries could also occur because an injured player continues to participate in their sport with some modification of their techniques, physical adaptation or mal-adaptation, complete/incomplete recovery from injury or a combination thereof. This means that their risk of further injury will no longer be the same as for their first injury.[52]

In the sports injury literature to date, recurrent injuries have been considered from a clinical management and return-to-play (or time away from sport to recover from injury) perspective.[1–3 15 53 54] Sports injury surveillance guidelines and several conceptual papers describe the complex issues associated with properly classifying injuries as recurrent, re-injury, exacerbations or overuse.[4 55–58] None of this prior work, however, has discussed recurrent injuries from a statistical viewpoint, and so adequate recognition of the various dependencies both within and between injured players is lacking in the sports medicine literature.

In the case of injury count data, sports injury counts have been most commonly analysed in the literature as Poisson counts. When players would reasonably be expected to sustain more than one injury, it would be more correct to apply negative binomial models, as we have shown when modelling falls in older people.[46] The present study offers a comprehensive approach to guide the choice of different survival models when
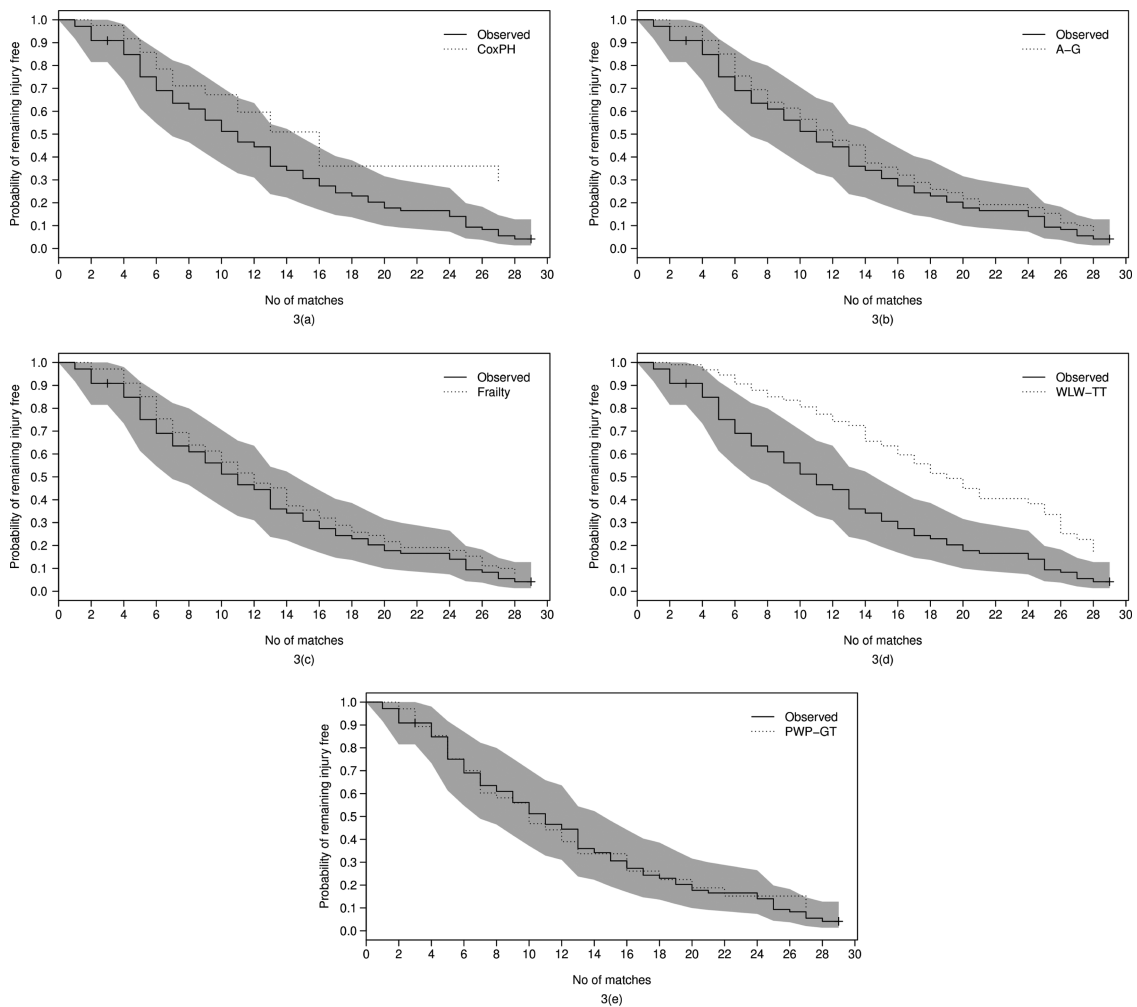
Figure 3 Standard Kaplan-Meier (K-M) curves for probability of remaining free of injury for 35 professional rugby league players. Actual and fitted survival curves from (A) CoxPH model, (B) A-G model, (C) frailty model, (D) WLW-TT model and (E) PWP-GT model. The grey shaded regions are 95% CIs for the fitted survival curves. Models were adjusted by age, match experience and body mass of the players.

modelling recurrent sports injury data when the time to events is of prime interest, rather than only an overall count.

Although the CoxPH model is the most commonly used approach for analysing time-to-event data, it fails to take into account the extra variability of the recurrent events and, as this paper has shown, provides only poor fit to recurrent sports injury data. This is perhaps not surprising given that it only considers the time to the first injury and discards the remaining

injuries. This is a critical limitation because it means that important information about injury occurrence and associated risk factors is potentially excluded from current models which only consider the time to first injury.

Each of the four tested generalisations of the CoxPH model (A-G, frailty, WLW-TT and PWP-GT models) provided a substantial model improvement over the CoxPH model. In general, the A-G and frailty models performed best and provided better data

Table 3 Model selection criteria (log likelihood (LL), Akaike information criterion (AIC) and Bayesian information criterion (BIC)) of the fitted models for sports injury recurrent data*

Model	Model selection criteria		
	LL	AIC	BIC
Andersen-Gill (A-G)	135.0	275.9	355.6
Frailty	134.9	277.9	378.0
Wei-Lin-Weissfeld total time (WLW-TT) marginal	158.1	334.2	487.6
Prentice-Williams-Petersen gap time (PWP-GT) conditional	154.8	327.7	481.1

*The LL, AIC and BIC were not reported due to the small estimated likelihood for the CoxPH model for only the first event.

Table 4 Mean square error (MSE), root mean-squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) of the fitted models for sports injury recurrent data

Model	Model accuracy measures			
	MSE	RMSE	MAE	MAPE
Cox proportional hazards (CoxPH)	0.04	0.19	0.15	0.64
Andersen-Gill (A-G)	0.001	0.04	0.03	0.13
Frailty	0.001	0.03	0.03	0.12
Wei-Lin-Weissfeld total time (WLW-TT) marginal	0.03	0.18	0.15	0.64
Prentice-Williams-Petersen gap time (PWP-GT) conditional	0.01	0.10	0.09	0.47

Table 5 Pairwise goodness-of-fit (likelihood ratio test (LRT), F-ratio (F) and bootstrap (BS)) p-values for comparing the Cox proportional hazards (CoxPH) model, Andersen and Gill (A-G) model, frailty model, Wei-Lin-Weissfeld total time (WLW-TT) marginal model and Prentice-Williams-Petersen gap time (PWP-GT) conditional model for sports injury recurrent data

Comparison of models	Goodness-of-fit p values		
	LRT*	F	BS†
CoxPH vs A-G‡	–	–	–
CoxPH vs frailty	–	<0.001	–
CoxPH vs WLW-TT	–	0.67	–
CoxPH vs PWP-GT	–	0.99	–
A-G vs frailty	0.84	0.50	0.85
A-G vs WLW-TT	0.03	0.03	0.02
A-G vs PWP-GT	0.20	0.08	0.14
Frailty vs WLW-TT	0.02	0.02	<0.001
Frailty vs PWP-GT	0.03	0.02	0.01
WLW-TT vs PWP-GT‡	–	–	0.78

*LRT test is based on log likelihood and is not appropriate for comparing first event model (CoxPH model) and recurrent events models (Cox extension models).
†The resampling procedure was based on the CoxPH model in the BS test and hence the extended models were not fitted for first event only when compared with the CoxPH model.
‡Models are not nested.

fits to the recurrent sports injury data when compared to both WLW-TT and PWP-GT models.

There was no statistical difference between the A-G and frailty models when applied to the NRL recurrent injury data analysed in this study in terms of model selection, goodness of fit or accuracy. This was confirmed with the simulation substudy. Nonetheless, as the frailty model requires fewer data assumptions than the A-G model and it does allow investigation of effects that might change based on injury-specific covariate effects (which the A-G model does not), it is recommended that the frailty model be adopted when analysing recurrent sports injury data in the future, when this is consistent with the research question.

The A-G model is the most simple variance-corrected model and incorporates robust variance estimators, which have good statistical properties under some circumstances. This model can

also be used to adjust for covariate effects. The frailty model includes a random effect (frailty) to account for the within-subject correlation between injuries and so is a more general model, with fewer assumptions. In the case where there is significant within-person correlation (as applies to our injury data), Kelly and Lim[5] recommend the use of frailty models, which incorporate random effects because they fit the data better than the PWP-GT model.

The statistical model comparison was only conducted on a small injury sample, and it is possible that different conclusions may arise when applied to other injury contexts. We have recently applied the frailty model to other rugby league injury data, including for the purposes of risk factor identification, indicating its likely robustness for this sort of recurrent injury data.[59 60]

Although the frailty model has offered the best fit to the rugby league recurrent injury count data in this study, this does not guarantee that this model would offer the best fit for other sports injury data sets, and this would need further exploration. Nevertheless, the fitting procedures presented in this paper, and the various model selection criteria, may be used as guidelines for modelling recurrent injury data in other sports injury contexts.

In conclusion, sports injury data characterised by recurrent events due to repeat or subsequent injuries over a period of time need to be appropriately analysed to take into account the different likely dependences within the data. Such data can be appropriately analysed by either the A-G or frailty model, with the frailty model representing a marginally better fit than A-G model. The strength of the frailty model is that it considers individual baseline injury risks for different players, makes fewer statistical assumptions and also is able to model time-varying covariates.

What this study adds

- ▶ A summary of the important statistical considerations when analysing recurrent injury data.
- ▶ Guidance on the best statistical model to use for analysing recurrent sports injuries.

Table 6 Simulated estimates (based on 100 simulation replications) of the size and power of the test to compare Andersen and Gill (A-G) model, frailty model, Wei-Lin-Weissfeld total time (WLW-TT) marginal model and Prentice-Williams-Petersen gap time (PWP-GT) conditional model fitted to sports injury recurrent data*

Comparison of models	Simulated model size			Simulated model power		
	Pr(P>α)=α			Pr(P>β)=1−β		
	α=0.01	α=0.05	α=0.10	1−β=0.99	1−β=0.95	1−β=0.90
A-G vs frailty	0.02	0.04	0.08	0.99	0.99	0.92
A-G vs WLW-TT	0.03	0.08	0.13	0.97	0.93	0.88
A-G vs PWP-GT	0.03	0.05	0.11	0.99	0.96	0.93
Frailty vs WLW-TT	0.02	0.05	0.10	0.96	0.90	0.86
Frailty vs PWP-GT	0.01	0.04	0.09	0.98	0.93	0.89
WLW-TT vs PWP-GT	0.01	0.04	0.18	0.98	0.90	0.80

*The re-sampling procedure was based on the Cox model in the bootstrap test and hence the extended models were not fitted for first event only when compared with the Cox regression model.

REFERENCES

1. Hagglund M, Walden M, Ekstrand J. Previous injury as a risk factor for injury in elite football: a prospective study over two consecutive seasons. *Br J Sports Med* 2006;40:767–72.
2. de Visser H, Reijman M, Heijboer M, et al. Risk factors of recurrent hamstring injuries: a systematic review. *Br J Sports Med* 2012;46:124–30.
3. Swenson D, Yard E, Fields S, et al. Patterns of recurrent injuries among US high school athletes, 2005–2008. *Am J Sports Med* 2009;37:1586–93.
4. Hamilton G, Meeuwisse W, Emery C, et al. Subsequent injury definition, classification and consequence. *Clin J Sport Med* 2011;21:508–14.
5. Kelly P, Lim L-Y. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Stats Med* 2000;19:13–33.
6. Cai J, Schaubel D. Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Anal* 2004;10:121–38.
7. Schaubel D, Zeng D, Cai J. A semiparametric additive rates model for recurrent event data. *Lifetime Data Anal* 2006;12:389–406.
8. Ye Y, Kalbfleisch J, Schaubel D. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 2007;63:78–87.
9. Twisk J, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. *J Epidemiol Commun Health* 2005;59:706–10.
10. Schaubel D, Cai J. Semiparametric methods for clustered recurrent event data. *Lifetime Data Anal* 2005;11:405–25.
11. Sun L, Park D-H, Sun J. The additive hazards model for recurrent gap times. *Stat Sin* 2006;16:919–32.
12. Hosmer D, Lemeshow S. *Applied survival analysis: regression modelling of time to event data*. New York: Wiley, 2008.
13. Glynn R, Buring J. Ways of measuring rates of recurrent events. *BMJ* 1996;312:364–7.
14. Meeuwisse W, Tyreman H, Hagel B, et al. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clin J Sport Med* 2007;17:215–19.
15. Hollis S, Stevenson M, McIntosh A, et al. Mild traumatic brain injury among a cohort of rugby players: predictors for time to injury. *Br J Sports Med* 2011;45:997–9.
16. Gabbe B, Finch C, Bennell K, et al. Risk factors for hamstring injuries in community level Australian football. *Br J Sports Med* 2005;39:106–10.
17. Cox DR, Oakes D. *Analysis of survival data*. New York: Chapman and Hall; 1984.
18. Clayton D. Some approaches to the analysis of recurrent event data. *Stat Methods Med Res* 1994;3:244–62.
19. Lin DY. Cox regression analysis of multivariate failure time data: the marginal approach. *Stats Med* 1994;13:2233–47.
20. Gao S, Zhou X-H. An empirical comparison of two semi-parametric approaches for the estimation of covariate effects from multivariate failure time data. *Stats Med* 1997;16:2049–62.
21. Therneau TM, Hamilton SA. rhDNase as an example of recurrent event analysis. *Stats Med* 1997;16:2029–47.
22. Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. *Stats Med* 1997;16:833–9; discussion 41–51.
23. Box-Steffensmeier J, ed. A Monte Carlo analysis for recurrent events data. *Annual Political Methodology Meetings*; 2002.
24. Klein JP, Moeschberger ML. *Survival analysis: techniques for censored and truncated data*. New York: Springer, 2003.
25. Lim H, Liu J, Melzer-Lange M. Comparison of methods for analyzing recurrent events data: application to the emergency department visits of pediatric firearm victims. *Accid Anal Prev* 2007;39:290–9.
26. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Math Stats* 1982;10:1100–20.
27. McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. *Biometrics* 1991;47:461–6.
28. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 1989;84:1065–73.
29. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981;68:373–9.
30. Gabbett TJ. Incidence, site, and nature of injuries in amateur rugby league over three consecutive seasons. *Br J Sports Med* 2000;34:98–103.
31. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457–81.
32. Ihaka R, Gentleman RR. A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996;5:299–314.
33. R Development Core Team. *A language and environment for statistical computing*. Vienna, Australia: R Foundation for Statistical Computing, 2004. http://www.R-project.org.
34. Cameron AC, Trivedi PK. *Regression analysis of count data*. Cambridge, UK: Cambridge University Press; 1998.
35. Dayton CM. Model comparisons using information measures. *J Mod Appl Stat Meth* 2003;2:281–92.
36. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Meth Res* 2004;33:261–304.
37. Hardin J, Hilbe J. *Generalized linear models and extensions*. Texas: Stata Press, 2007.
38. Makridakis S, Wheelwright SC, Hyndman RJ. *Forecasting: methods and applications*. New York: Wiley, 1998.
39. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecasting* 2006;22:679–88.
40. Shen Q, Faraway J. An F test for linear models with functional response. *Stat Sin* 2004;14:1239–57.
41. Box-Steffensmeier J, De Boef S. Repeated events survival models: the conditional frailty model. *Stats Med* 2006;25:3518–33.
42. Walters SJ, Campbell MJ. The use of bootstrap methods for estimating sample size and analysing health-related quality of life outcomes. *Stats Med* 2005;24:1075–102.
43. Davidson R, MacKinnon JG. The power of bootstrap and asymptotic tests. *J Econometrics* 2006;133:421–41.
44. Thomas AG. Product Limit estimation for event history and survival analysis. R package version 1.3.1, 2012 http://CRAN.R-project.org/package=prodlim.
45. Khan A, Ullah S, Nitz J. Statistical modelling of falls count data with excess zeros. *Inj Prev* 2011;17:266–70.
46. Ullah S, Finch CF, Day L. Statistical modelling for falls count data. *Acc Anal Prev* 2010;42:384–92.
47. Lin DY, Wei LJ, Yang I, et al. Semiparametric regression for the mean and rate functions of recurrent events. *JRSS(B)* 2000;62:711–30.
48. Wang M-C, Chen Y-Q. Nonparametric and semiparametric trend analysis for stratified recurrence times. *Biometrics* 2000;56:789–94.
49. Fong D, Lam K, Lawless J, et al. Dynamic random effects models for times between repeated events. *Lifetime Data Anal* 2001;7:345–262.
50. Wang M-C, Qin J, Chiang C-T. Analyzing recurrent event data with informative censoring. *J Am Stat Assoc* 2001;96:1057–65.
51. Metcalfe C, Thompson SG, Cowie MR, et al. The use of hospital admission data as a measure of outcome in clinical studies of heart failure. *Eur Heart J* 2003;24:105–12.
52. Bangdiwala S. Methodological considerations in the analysis of injury data: a challenge for the injury research community. In: Mohan D, Tiwari G. *Injury prevention and control*. London: Taylor & Francis, 2000:35–48.
53. Croisier J-L. Factors associated with recurrent hamstring injuries. *Sports Med* 2004;34:681–95.
54. Hammond L, Lilley J, Pope G, et al. Considerations for the interpretation of epidemiological studies of injuries in team sports: illustrative examples. *Clin J Sport Med* 2011;21:77–9.
55. Fuller C, Ekstrand J, Junge A, et al. Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Br J Sports Med* 2006;40:193–201.
56. Fuller C, Molloy M, Bagate C, et al. Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. *Br J Sports Med* 2007;41:328–31.
57. Fuller CW, Bahr R, Dick RW, et al. A framework for recording recurrences, reinjuries, and exacerbations in injury surveillance. *Clin J Sport Med* 2007;17:197–200.
58. Bahr R. No injuries, but plenty of pain? On the methodology for recording overuse symptoms in sports. *Br J Sports Med* 2009;43:966–72.
59. Gabbett TJ, Ullah S. Relationship between running loads and soft-tissue injury in elite team sport athletes. *J Strength Cond Res* 2012;26:953–60.
60. Gabbett TJ, Ullah S, Finch CF. Identifying risk factors for contact injury in professional rugby league players—application of a frailty model for recurrent injury. *J Sci Med Sport* 2012; in press. doi.org/10.1016/j.jsams.2012.03.017.