Risk Factors and Risk Statistics for Sports Injuries

Will G. Hopkins, PhD, Stephen W. Marshall, PhD,† Kenneth L. Quarrie,*‡ and Patria A. Hume, PhD**

Background: Risk factors for sports injuries include characteristics and behaviors of athletes and characteristics of sports and the environment that are associated with some measure of risk of injury.

Objective: To introduce risk statistics to clinicians evaluating studies of sports injuries.

Methods: Plain-language review of risk statistics and their practical application to sports injuries.

Results: The various measures of injury incidence are injury risk (proportion of athletes injured in a given period of training, playing, or other exposure time), injury rate (number of injuries per unit of exposure time), odds of injury (probability injury will happen divided by probability injury will not happen), injury hazard (instantaneous proportion injured per unit of time or mean injury count per unit of time), and mean time or mean number of playing exposures to injury. Effects of risk factors are estimated as values of effect statistics representing differences or ratios of one or more of these measures between groups defined by the risk factor. Values of some ratios and their sampling uncertainty (confidence limits) are estimated with specialized procedures: odds ratios with logistic regression, rate ratios with Poisson regression, and hazard ratios with proportional hazards (Cox) regression. Injury risks and mean time to injury in each group can also be estimated and can give a better sense of the effect of a risk factor. Risk factors identified in nonexperimental cohort and case-control studies are not always causes of injury; data from randomized controlled trials provide stronger evidence of causality.

Conclusion: Expressing risk statistics as meaningful numbers should help clinicians make better use of sports injury studies.

(*Clin J Sport Med* 2007;17:208–210)

Many factors are known or suspected to modify the risk or likelihood that an athlete will get injured during training or competition in sports and recreational physical activity. These so-called risk factors are usually subject characteristics and behaviors, such as age, sex, skill, use of protective equipment, playing position, and game strategies; they can also be sport or game characteristics, such as level of competition, playing surface, and weather.[1] Researchers use a variety of statistics to summarize the effects of risk factors. In this paper, we explain what these statistics mean by explaining how they arise from the various measures used to deal with incidence of injury.

Submitted for publication February 20, 2007; accepted March 6, 2007.
From the *Institute of Sport and Recreation Research NZ, AUT University, Auckland, New Zealand; †Injury Prevention Research Center, University of North Carolina, Chapel Hill, North Carolina; and ‡Research & Injury Prevention, New Zealand Rugby Union, Wellington, New Zealand.
No sources of funding were provided for this paper.
Reprints: Will G. Hopkins, PhD, Institute of Sport and Recreation Research NZ, AUT University, Private Bag 92006, Auckland, NZ (e-mail: will.hopkins@aut.ac.nz).

INJURY-INCIDENCE AND RISK FACTOR STATISTICS

When we speak of a risk factor for injury, we imply that different values of the factor are associated with differences in the incidence of injury. Several simple statistics are used as measures of injury incidence, including injury risk, injury rate, odds of injury, injury hazard, and time to injury. Effect statistics representing differences or ratios of these measures quantify the association between a risk factor and incidence of injury.

Injury Risk

The term *risk* is often used generically for all statistics related to incidence of injury, but it is also a specific statistic referring to the proportion of a study group with an injury or, equivalently, the probability of injury for any given subject. Risk can be expressed as a decimal fraction or as a percent of subjects injured (by multiplying by 100). For example, the risk of minor injury in 1 week of training and competing in a particular sport might be 0.15 or 15% for those with the condition of extreme flexibility known as hypermobility and 0.10 or 10% for those with normal flexibility. Such differences in risk are summarized into a single statistic that represents the effect of the risk factor on injury risk. In this example, 2 obvious such statistics are the *risk difference* of 0.05 or 5% (= 15–10) and the *risk ratio* of 1.50 (= 0.15/0.10 or 15/10). The term *relative risk* is often used instead of risk ratio, but relative risk can also refer to the comparison of other measures of injury incidence.

Risks are useful, because they answer a question of great interest to an athlete, coach, or parent of a young athlete: what is the probability the athlete will get hurt playing a particular sport this season? The risk is the average probability of injury for an athlete. As with any of these statistics, some athletes may have a different risk, for example by being overly flexible. Risks generally increase with time: the risk of injury over 10 games will be higher than the risk of injury over one game. Thus, the time period for each risk should always be stated.

Risk differences and ratios are good measures for a researcher to use when playing and practice times are similar between the groups being compared. For example, the risk ratio of 1.50 given above implicitly assumes that hypermobile and normal athletes have the same exposure time and, therefore, the same opportunity for injury. If hypermobile athletes play less or more than normal athletes, then the risk ratio is misleading. Rates are one way to overcome this limitation.

Injury Rate

An injury rate is the number of injuries over a period of time divided by a measure of the total exposure to sport during this period. The measure of exposure is athlete-time at risk: the total number of athletes (injured and uninjured) multiplied by their average participation time, expressed as the number of practices and/or games or as actual time spent in these activities. For example, a team of 25 athletes that holds 10 games and 100 practices over a season would total 250 athlete-games and 2500 athlete-practices of exposure to sport. The games and practices are sometimes added together and referred to as athlete-exposures, and the injury rates are expressed per 1000 athlete-exposures. Authors also report injury rates per 10,000 player-minutes, typically by multiplying the number of athlete-practices and athlete-games by the average number of minutes in a practice and game, then adding the two together. The term *incidence density* is sometimes used to refer to injury rate, but it is unnecessary. Irrespective of its name or units, an injury rate indicates how many injuries occur per number of episodes of exposure to sport.

The advantage of rates over risks for the researcher is that they allow us to adjust for differences in exposure to sport between groups. Rate differences and ratios are derived from the injury rates in the groups. In the hypermobility example, risks of 15 and 10 injured athletes per 1000 athlete-exposures in the hypermobility and normal groups each would give a rate ratio of 1.5. The risk ratio and the rate ratio are the same in this example, because the amount of playing time (athlete-exposure) is the same in both groups.

Odds of Injury

Odds are a less intuitive measure than risks or rates. The odds of injury is the probability the injury will happen divided by the probability it won't happen. In the hypermobility example, the odds of injury for hypermobile and normal athletes would be respectively 0.15/(1–0.15) = 0.18 and 0.10/(1–0.10) = 0.11. Notice that the odds are different from the risks! Odds approximate risks only when the risk of injury is low (below 10%).

The effect statistic from a comparison of odds of injury in 2 groups is the odds ratio, which here is 0.18/0.11 = 1.59. This value is a little higher than the risk ratio (1.50). As with the odds themselves, the odds ratio approximates the risk ratio closely when both risks are small (<10%). In this situation, the odds ratio can be interpreted as a risk ratio, but it is otherwise difficult to interpret and should not be discussed as if it were a risk ratio. The odds ratio of 1.59 does not mean that hypermobile athletes are 1.59 times as likely to be injured; over the time frame of the study they are in fact 1.50 times as likely to be injured. For this reason, we discourage the use of odds and odds ratios unless the injury risk is low. Odds ratios are widely reported in the biomedical literature, largely because some statistical procedures are based on analysis of odds.

Injury Hazard

In the example above, we have imagined calculating the injury rates and rate ratios over a reasonable interval of time, such as a season. But we can also compute the injury rate over very short periods of time, such as hours or minutes or even seconds. When we evaluate rates over a period of time that is so small as to be essentially zero, the rate becomes a new statistic known as a *hazard*: the instantaneous risk per unit time. The value of the hazard changes with the scale of the unit of time; for example, the minor-injury hazard for the hypermobile athletes is 0.023 per day, 0.16 per week, 0.70 per month, 1.0 per 43 days, 8.5 per year, and so on. One way to understand that all these values represent the same hazard is to think of the hazard as the number of injuries the athlete would sustain on average over the unit of time (under the assumption that the hazard does not change over that period). Thus, a hazard of 0.023 per day implies about two-hundredths of an injury per day or, more meaningfully, one injury per 43 days or 8.5 injuries per year.

The hazard ratio is the instantaneous risk ratio. In the hypermobility example, a hypermobile athlete and a normal athlete both have very small risk of injury over any brief period of playing time, but the hypermobile athlete is still at greater risk of injury. The ratio of these very small risks, the hazard ratio, is 1.54; in other words, the hypermobile athlete is 1.54 times more likely to be injured than a normal athlete in any moment of play.

The hazard ratio, like the odds ratio, approximates the risk ratio when the actual risks over a finite monitoring period are <10%. As the proportions injured increase over longer periods of monitoring, the risk, hazard, and odds ratios diverge such that risk ratio < hazard or rate ratio < odds ratio. By the time almost all athletes in both groups would be injured (and ignoring any issues arising from re-injury), the risk ratio would converge on 1.0, the odds ratio would tend to infinity, but the hazard ratio would still be meaningful and useful.

Mean Time to Injury

The mean time to injury is a practical measure of risk for the individual, although it is rarely reported. Time can be the number of games and practices, rather than clock or calendar time; mean time to injury then becomes mean number of games and/or practices to injury.

If the hazard is constant during a period of monitoring, the mean time to injury is simply the inverse of the hazard. In the preceding example, the mean time to injury for hypermobile and normal athletes would be 6.2 and 9.5 weeks. The effect statistics based on time to injury are the difference in mean times and the ratio in mean times. The difference of 3.3 weeks is the extra playing time before injury that the average normal athlete would get in comparison with the average hypermobile athlete. The ratio of mean times to injury (normal/hypermobile) is simply the hazard ratio, indicating here that on average normal athletes go for 1.54 times as long as hypermobile athletes before getting injured.

UNCERTAINTY IN RISK FACTOR STATISTICS

All estimates of injury incidence are based on a sample of athletes, so they are only approximations of the true incidence–the incidence that the researcher would observe with a very large sample. Researchers sometimes show the sampling uncertainty for the different values of incidence in different groups, but an inference about the true effect of a risk factor has to be based on the uncertainty in the effect statistic. Some researchers address this problem by performing a hypothesis test involving a *P* value and a statement about statistical significance or nonsignificance of the effect statistic, in the mistaken belief that this approach resolves the issue of whether or not the effect is "real."[2] This test-based approach is no longer acceptable in epidemiologic studies.[3] Researchers should now present the sampling uncertainty directly as confidence limits or a confidence interval, usually but not necessarily at the 95% level. The confidence interval represents a range in which the true value (the very large-sample value) of the effect is likely to be found.

In the hypermobility example, the 95% confidence interval for the risk ratio of 1.50 might be 0.91 to 2.48. Thus, the true risk of injury from hypermobility is probably somewhere between 0.91 (slightly less risk of injury than women experience) and 2.48 (more than twice the risk that women experience). The researcher must then decide whether the uncertainty in the magnitude of the risk factor provides sufficient evidence that the factor is important. The decision needs to take into account the value for the factor that would represent the smallest clinically important increase or decrease in risk. The values of such differences and the way they should be taken into account depend on how the data are to be used and are topical issues.[4,5]

A confidence interval tells us about the uncertainty or error in a risk statistic due to sampling variation, but there are frequently other sources of substantial error, such as selection bias (if the sample is not representative of the population of interest) and confounding (by unknown or unmeasured risk factors). Authors usually discuss but seldom quantify these errors. Methods for including them in the confidence interval are under development.[6]

ESTIMATING RISK FACTOR STATISTICS

Crude estimates of injury incidence can and should be derived simply from counts of injured and uninjured athletes, but sophisticated procedures are required to calculate confidence limits and to properly adjust for and estimate the separate contributions of several risk factors. The procedures in widespread use are logistic regression, Poisson regression, and proportional-hazards (or Cox) regression. The nature of the data determines which procedure is most appropriate. Researchers usually present the outcome statistic provided by the procedure: odds ratios for logistic regression, count ratios or rate ratios for Poisson regression, and hazard ratios for proportional hazards regression. As explained above, with certain assumptions these outcome statistics are all interchangeable. Note, however, that odds ratios from properly designed case-control studies are actually statistically equivalent to hazard ratios, so conversion of such odds ratios to relative risks[7] is not appropriate. Note also that injury risk, time to injury, and the differences and ratios of these measures of injury incidence are not analyzed directly by the regression procedures. These statistics are arguably the best for communicating the magnitude of the effect of a risk factor and should be estimated via the regression procedures. When authors do not provide these statistics, you may have to consult a statistician to work out at least approximate values.

RISK FACTORS AND CAUSALITY

We should be cautious about assuming that a characteristic or behavior identified as a risk factor for injury is a cause of injury. Almost all studies of injury risk factors have a cross-sectional, prospective cohort or case-control design, and these non-experimental studies provide evidence only of an association between the factor and risk of injury. A properly designed controlled trial or intervention, in which athletes are randomized to experimental and control treatments, provides stronger evidence that a known or suspected risk factor represented by the difference between the treatments causes different risks of injury during the follow-up period. Such studies are currently rare in the sports injury literature, partly for logistical or ethical reasons, and well-designed and analyzed non-experimental studies usually provide good evidence for programs of injury prevention.

CONCLUSION

Risk and risk factor statistics can be confusing for practitioners and researchers. It is important to transform the statistics into numbers that convey magnitudes of risk in meaningful realistic terms, such as chances of injury in a season or mean playing time before an injury occurs. Uncertainty in the estimate of a risk factor statistic provided by a study also needs to be taken into consideration before deciding whether a behavior or characteristic represented by the risk factor is important. Finally, risk factors identified in non-experimental studies are not necessarily causes of injury.

REFERENCES

1. Meeuwisse WH. Assessing causation in sport injury: a multifactorial model. *Clin J Sport Med*. 1994;4:166–170.
2. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ*. 2001;322:226–231.
3. Poole C. Low p-values or narrow confidence intervals: which are more durable? *Epidemiology*. 2001;12:291–294.
4. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Sportscience*. 2005;9:6–12.
5. Hopkins WG. Magnitude matters. *Sportscience*. 2006;10:58.
6. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int J Epidemiol*. 2004;33:1389–1397.
7. Shrier I. Understanding the relationship between risks and odds ratios. *Clin J Sport Med*. 2006;16:107–110.