The mathematical relationship between the number of events in which people are injured and the number of people injured

Daan G. Uitenbroek MA

Research Unit in Health and Behavioral Change, University of Edinburgh, Edinburgh, UK

In this paper the mathematical relationship between the number of times people sustained injuries and the number of people injured is discussed. The number of injury-causing events in a population will be larger than the number of people injured as people can be injured on more than one occasion. The Poisson distribution is used as a way to study the relationship between the two different sets of numbers. A practical example is given of injuries sustained while running. From this it seems that injury proneness needs to be considered in studying those who were injured on more than one occasion while running during the past year.

Keywords: injury rate, accident rate, running injuries, Poisson distribution, injury proneness

A complicating factor in the measurement of the number of injuries and the establishment of injury rates is that two different ways of counting can be used. The counting of occasions on which people were injured (accidents or events) using a particular time frame, and the counting of people injured in a similar time frame, are two methods which give very different results from a mathematical perspective. For example, if one takes throughput in an accident and emergency department as a basis for calculating injury rates then one counts the number of individual events, i.e. the number of injuries. This is because if the same person reported to the accident and emergency department the next day having sustained another injury, he or she would be counted again, as a new case. However, if patient files are used as a basis for calculating injury rates this is done by taking the proportion of files in which reports of injuries are made. In this case the number of people injured will be counted. A lower sum total than for 'events' will be found as a number of those counted will have sustained an injury on more than one occasion. A similar problem arises in retrospective surveys in which the question asked is: 'how many patients or respondents were injured in the previous

month or year?'. In this case, again, it is the number of injured that is counted, and we have to multiply the number of injured with the average number of events per injured individual to arrive at the total number of events for the population.

This problem has been recognized by researchers and the number of occasions on which people were injured within the period under study is often established by using follow up questions. A second method is to relate the number of injured and the number of events observed in a particular time frame using mathematical theory. This is relatively easy and can be done using classical and well tested statistical methods. However, one has to be able to support some of the assumptions to do these calculations reliably and validly. Two assumptions are considered to be particularly important: (1) that any member of a population has an equal chance of being involved in an 'event' and therefore to sustain an injury; and (2) that the chance of sustaining an injury on another occasion is not influenced by having had an injury recently. On the basis of these assumptions, the chance of getting an injury can be viewed as a random process with replication; the victim is being put back in the population immediately after being 'drawn'. It will not always be possible to support these assumptions. For example, a serious or fatal injury would lower the victim's chance of sustaining an injury on another occasion quite dramatically. In the psychological literature there is mention of the injury-prone or clumsy personality, and of 'risk-seekers', these individuals have a higher than random chance of sustaining injuries¹⁻³.

However, although the assumptions stating that the injury-causing 'events' are a random process where replication might not be fully met, mathematical models can be used in the description of injuries and can be used to see if sustaining injuries is a random process, or whether other factors also need to be taken into consideration.

The Poisson distribution

The first basic rule to consider is that the chance of an individual sustaining at least one injury ($p_{\geq 1}$) is 1

Address for correspondence: Daan G. Uitenbroek, Research Unit in Health and Behavioral Change, University of Edinburgh, 24 Buccleuch Place, Edinburgh EH8 9LN, UK

minus the chance of this individual sustaining no injuries (p_0) thus:

$$p_{>1} = 1 - p_0$$

The proportion of individuals in a population who sustained injuries on exactly one, two, three or more occasions is given by a multiplication rule where the proportion of people who sustained one more injury (p_{n+1}) is calculated by multiplying a particular number of injuries (p_n) with a factor. The number of people sustaining injuries on nil, one, two, three or more events is related as determined by the multiplication factor. The Poisson distribution can be best used to express the relation between the number of events and the number of injured:

$$p_{n+1} = p_n \cdot -\ln(p_0)/(n+1) \quad (n = \text{number of events})$$

The basic rule of statistics of course applies:

$$p_0 + p_1 + p_2 + \dots + p_{\infty-1} + p_{\infty} = 1$$

On the basis of this rule the proportion of individuals who were injured on more than a particular number of injury-causing occasions ($p_{>n}$) is 1 minus the chance of being involved in nil, one, two, up to the number in which one is interested:

$$p_{>n} = 1 - p_0 - p_1 - p_2 - \dots - p_n$$

These basic rules of statistics can be applied to study the distribution of injuries in a population and to see how many people sustain an injury on one occasion, how many on two, or more occasions, and how many sustain no injuries. These rules ensure that the results obtained in mathematical calculations are also 'reasonable' in a more tentative way. For example, under these rules there will always be a lucky few who sustain no injuries, while, on the other hand, it is possible that there will be some unlucky individuals who are injured on a very large number of occasions. Further, these rules ensure that in the calculations the number of victims of injuries does not exceed the size of the population. However, the total number of occasions on which people are injured can be larger than the size of the population. This is the case if on average the number of 'events' per head of the population is higher than 1, which would inevitably happen if injury-causing events are studied using a long time frame. The expected average number of events per head of the population (μ) is related to the chance of not sustaining an injury (p_0) according to the following formula:

$$\mu = -\ln(p_0) \text{ and } p_0 = \exp(-\mu)$$

Of course, the total number of events in a population is the average number of events per head of the population multiplied by the size of the population. Lastly, two formulae to recalculate data are used: (a) using the month as a frame of reference; and (b) using the year as a frame of reference. First, to obtain the average number of events per head of the population per year one takes the sum of the averages for the 12 months:

$$\mu_{(\text{year})} = \mu_{(\text{jan})} + \mu_{(\text{feb})} + \mu_{(\text{mar})} + \dots + \mu_{(\text{dec})}$$

However, to obtain the proportion of individuals who sustained no injury in a year, the product of the proportions found for each month needs to be considered, thus:

$$P_0(\text{year}) = P_0(\text{jan}) \cdot P_0(\text{feb}) \cdot P_0(\text{mar}) \cdot \dots \cdot P_0(\text{dec})$$

A practical example

Blair *et al.*⁴ report data regarding running injuries sustained by members of a health and fitness club in Dallas, Texas. The data concerns 438 runners. Runners who had to stop running for at least 7 days in the last 12 months due to an injury were classified as injured. For each injured individual the number of occasions on which they were injured was also registered. Table 1 shows the distribution of running injuries reported by Blair *et al.* and the predicted number of injuries on the basis of the Poisson distribution. The Chi-square for the difference between the observed and predicted scores, calculated for the injured only, equals 62.7, which gives a *P*-value smaller than 0.001 by two degrees of freedom. Thus, the observed distribution of injuries is non random. Comparing the observed with the expected distribution, the over-estimation of high numbers of injuries sustained and the under-estimation of low numbers of injuries sustained points to the fact that the injured runners who have a higher number of injuries are injury prone, it is not only a question of bad luck to be injured more than once. Blair *et al.* discuss stretching habits, age, time and place of running, weight relative to squared height (BMI), average speed of running and average distance run per week as factors which might possibly be related to sustaining running injuries. Only BMI and average distance run were found to be related to an elevated risk of being injured. However, the non-randomness observed is very pronounced: seven times more runners who had more than two injuries were observed than could be expected had this number of injuries been due to random chance. Therefore other explanations, such as the above mentioned psychological factors, deserve future study. The chronic nature of many running injuries should also be considered. It might be that those who stop running due to frequent injury, do so because the same injury caused repeated problems. In this last case calculating an average number of newly sustained injuries on the basis of the data of Blair *et al.* would have overestimated the actual number of newly sustained injuries.

Table 1. Observed and predicted number of running injuries in a Dallas fitness club

Number of injuries	Observed (%) [†]	Predicted (%)
0 (p_0)	76.0	76.0
1 (p_1)	16.8	20.9
2 (p_2)	5.0	2.9
> 3 ($p_{>2}$)	2.2	0.3

[†]Source: Blair SN, Kohl HW, Goodyear NN. Rates and risks for running and exercise injuries: studies in three populations. *Res Qua Exerc Spt* 1987; 58: 221–28.

Conclusion

In this paper the mathematical relationship between the number of injury-causing events and the number of injured in a particular time frame was studied. These are two very different figures which have different mathematical properties. In particular, if a relatively long time frame is used in counting injury-causing events and injured individuals the nature of the data one collects needs to be carefully considered. The practical example given showed that running injuries are not randomly distributed, injury proneness could be observed, i.e. some personal or situational factor must be related to suffering more injuries. Comparing a distribution of running injuries with a random distribution can in this way be helpful in interpreting the collected data.

Acknowledgements

The Unit is funded by the Chief Scientist office of the Scottish Home and Health Department; by the Health Education Board Scotland (HEBS); and by the Economic and Social Research Council (ESRC: Grant # XA44250009). The opinions expressed in this communication are those of the author, not of the funding bodies.

References

- 1 Kelly MJ. Psychological risk factors and sport injuries. *J Spt Med Phys Fitn* 1990; **30**: 202–21.
- 2 Kerr JH. Arousal-seeking in risk sport participants. *Per Ind Diff* 1991; **12**: 613–16.
- 3 Yaffé M. Sports Injuries: psychological aspects. *Brit J Hosp Med* 1983; **29**: 224–32.
- 4 Blair SN, Kohl HW, Goodyear NN. Rates and risks for running and exercise injuries: studies in three populations. *Res Qua Exerc Spt* 1987; **58**: 221–8.