METHODOLOGIC ISSUES

Methodologic issues in injury case-control studies

Ian Roberts

Abstract

In this paper some methodological problems particularly relevant to case-control studies of injury are illustrated by reference to previous childhood injury case-control studies. In contrast to studies of disease, where 'person time' constitutes the observational experience of interest, in injury studies person time engaged in a particular activity is often more appropriate. The implications for the definition of the study base are discussed. The potential for hospital admission bias in injury case-control studies is considered along with potential strategies for avoiding it. The importance of errors in exposure measurement, including those arising from inappropriate induction time assumptions, are illustrated. Finally, the potential for bias resulting from the combination of etiologically unrelated injury outcomes into a single outcome measure is illustrated and discussed.

(Injury Prevention 1995; **1**: 45–48)

Keywords: epidemiologic methods, case-control studies, accidents.

The application of case-control methodology to the identification of risk factors for unintentional and intentional injury has been an important and timely advance in the field of childhood injury control.[1] Although the conceptual basis of the case-control study readily accommodates the examination of risk factors for injury, there are nevertheless a number of methodological issues which are particularly pertinent in studies of injury. In this paper these issues are discussed. For the purpose of this discussion the case-control study is viewed as involving three distinct stages. First, the definition of the study base. Second, the collection of data on injury events and person time exposed and unexposed in the study base and finally, the calculation of effect estimates.

Defining the study base

Department of Epidemiology, McGill University, Montreal, Canada

Correspondence to:
Dr Ian Roberts, Community Paediatric Research, C538, Montreal Children's Hospital, McGill University, 2300 Tupper, Montreal, Quebec H3H 1P3, Canada.

Every case-control study takes place within a theoretical cohort, the study base. The study base is the segment of person time in which diseased (or injured) subjects become cases in the study.[2] For the valid conduct of a case-control study, cases and controls must be representative of the same base experience.[3]

Incidence rates for disease are conventionally expressed with person time in the denominator. Person time 'living' constitutes the observational experience in which disease onsets are counted.[4] As a consequence, the study base in most case-control studies of disease is a specific segment of person time living.[5] However, in studies of injury, the observational experience of interest is often not simply person time living, but person time engaged in a particular activity. For example, an appropriate measure of injury occurrence for motor vehicle occupant injuries might be injuries per unit of time spent as a motor vehicle occupant. The study base for a case-control study of motor vehicle occupant injuries might therefore be a specific segment of person time spent as an occupant of a motor vehicle. The control group would be required to provide an estimate of the exposure distribution in such a study base.

The definition of the study base depends on the nature of the research question. Consider a study designed to quantify the protective effect of bicycle helmets. Cycle helmets are believed to prevent head injury only in the event of a cycle crash. If a cohort study were conducted, the numerators of the incidence rates for the exposed and unexposed would be the numbers of head injuries sustained in cycle crashes. The denominators for the incidence rates would be person time spent crashing a bicycle, or since crashing can be considered as instantaneous, the number of persons involved in bicycle crashes. Consequently the study base in a case-control study would be defined as the bicycle crashing experience of a defined population. A control group would be selected which would provide an estimate of the distribution of exposed (helmeted) and unexposed (unhelmeted) within this study base. Provided that the wearing of cycle helmets is not associated with the propensity to injure body regions other than the head, a control group comprising cyclists presenting with injuries other than to the head would therefore comprise a suitable control group.[6][7]

A similar approach to defining the study base was taken in a case-control study of the effectiveness of wrist guards in preventing wrist injuries to in-line skaters.[8] Cases were all persons in a defined population who sustained a wrist injury while in-line skating. Controls were all persons in the same population who sustained in-line skating injuries to sites other than the wrist. In this example, the study base can be conceptualised as the in-line skating 'accident' experience of the study population.

The case-control data below describe the association between socioeconomic status and

the risk of child pedestrian injuries occurring on the school-home journey. The methods have been described in detail elsewhere.[9] If the study base were defined as all school aged children in a temporally and geographically defined population, then cases would be all pedestrian injury events occurring on the school-home journey and controls would be a random sample of all schoolchildren in the study population. The first row in table 1 shows the distribution of subjects and the odds ratios associated with socioeconomic position that were obtained in this study. For the purpose of these analyses socioeconomic position was defined according to the New Zealand based Elley Irving scale.[10] There were strong associations with socioeconomic position. However, pedestrian injuries occur only while walking. If the aim of the study was to determine whether poor children are exposed to a more dangerous urban traffic environment, then it would be necessary to separate out the effect of the increased pedestrian exposure of poor children.[11] The second column of the table shows the distribution of subjects and odds ratios obtained when the study base was defined as all children who walk to and from school in the same temporally and geographically defined population. The association with socioeconomic position is now considerably weaker. Comparing the control groups in the two rows it can be seen that the proportion of children who walk in the lowest socioeconomic stratum is over twice that of children in the highest stratum. As a result, the effect of socioeconomic position found in row one was a mix of the effects of socioeconomic position of the propensity to walk[12][13] and the effect of socioeconomic position on the risk of occurrence of pedestrian injury while walking. The association with socioeconomic position was confounded by the propensity to walk. Restricting the study base to children who walk to and from school effectively prevents the propensity to walk from confounding the association with socioeconomic position.

Data collection — injury events

Having defined the study base, data on injury events arising in the study base are then collected. This will usually involve some method of injury surveillance. While ascertaining all injury events arising in the study base is not a prerequisite for validity, it is a requirement that the cases ascertained are a random sample of the study base with respect to the exposures of

interest.[2] One potential concern in studies where the case group comprises injury subjects requiring hospitalisation, is that at lower levels of injury severity, the likelihood of hospitalisation may be dependent on factors other than the injury itself and some of these factors may be the exposures of interest in the study.[14] It is conceivable that both the decision made by the patient to attend hospital and the decision of the doctor to admit the patient may be influenced by factors such as the ease of access to the hospital, admission policies, and the socioeconomic circumstances of the patient. Thus injured poor children may be more (or less) likely to be admitted to hospital than injured children who are not poor at the same level of injury severity. Since differential admission is likely to be more of a problem at low injury severity, some investigators have chosen to restrict the case group to subjects with injuries over and above a given level of injury severity.[15] An alternative approach would be to define the case group as all children admitted to hospital with the injury of interest and then to score the injuries according to an injury severity scoring system.[16] This would then permit an examination of effect estimates with cases stratified according to injury severity.

Data collection — measurement of exposure

Once case and control subjects have been identified they must be classified as exposed or unexposed to the potential risk factor under investigation. Errors in exposure measurement may result in subjects being misclassified with respect to the potential risk factor. Misclassification is non-differential when the sensitivity and specificity of exposure classification is the same for cases and controls. Misclassification is differential when the sensitivity and/or specificity of exposure classification varies according to case/control status. For dichotomously measured exposures, non-differential misclassification will bias the odds ratio towards the null value, whereas differential misclassification may bias the odds ratio either towards or away from the null value.[17]

Some degree of non-differential misclassification with consequent bias in the odds ratio towards the null value is inevitable in case-control studies. It is, however, particularly important to consider the potential for non-differential misclassification in studies which find no effect of the risk factor in question, since a real effect may have been obscured. On the other hand, in studies reporting a strong risk factor effect, it cannot be argued that exposure measurement error and consequent non-differential misclassification invalidate the findings, since the resulting bias would have been towards the null.[4] For example, in a case-control study conducted to determine the effect of left handedness on the risk of unintentional injury in childhood, handedness was determined using a questionnaire administered to children attending an emergency department. Although some misclassification of handedness would appear inevitable in this situation, this would not invalidate the study finding that left

Table 1   Numbers, odds ratios (OR), and 95% confidence intervals (CI) for socioeconomic position with varying definitions of the study base

Socioeconomic position	Cases	Controls	OR (95% CI)
(1) All schoolchildren controls			
I, II, III	6	80	1·00
IV, V	20	143	1·85 (0·68 to 5·90)
VI, VII	22	86	3·41 (1·25 to 10·76)
(2) Walking controls			
I, II, III	6	26	1·00
IV, V	20	71	1·22 (0·41 to 4·13)
VI, VII	22	58	1·64 (0·56 to 5·53)

handedness is a risk factor for unintentional injury (odds ratio = 1·80, 95% confidence interval 1·20 to 2·72).[18]

The presence of differential misclassification may be more problematic. In a case-control study of child pedestrian injuries, parental questionnaire reporting of variables describing the fencing of the children's play area from the roadway and the driveway were validated against direct observations made by the research staff.[19] When cases injured in residential driveways were compared with controls it was found that there was differential recall. The parents of injured children overstated the extent to which their driveways were unfenced. In this validation study, the odds ratio based on the questionnaire data overestimated the protective effect associated with a fence separating the driveway from the play area compared with the odds ratios based on direct observation. As a result of the bias demonstrated in this validation study, exposure measurement for the main study was based on direct observation.

As in all case-control studies, exposure measurement requires an assumption about the etiologically relevant timing between exposure and injury.[20] Inappropriate assumptions regarding the timing of exposure are a potent source of exposure misclassification. For example, in a case-control study of traffic volume and the risk of child pedestrian injury, when the exposure measure was defined as the average traffic volume on the day of injury for the case and a comparable day for the control, an odds ratio of 5·56 was estimated. However, when exposure was defined as the average traffic volume for the three hour period around the time of injury for the case and a comparable time for the control the odds ratio was 14·45.[21] Presumably both exposure measurement periods included the etiologically relevant period of traffic volume, however when volume was averaged over the entire day irrelevant exposure would have been included with relevant exposure, resulting in exposure misclassification and dilution of the effect estimate. In this example the etiologically relevant period of traffic volume is likely to have been the traffic volume at the instant the child entered the roadway, an induction time of almost zero, however, the induction period for exposures in injury studies need not necessarily be instantaneous. For example, in a study investigating the association between beer drinking and the risk of motor vehicle crash, the etiologically relevant period of beer drinking would depend on the rate of absorbtion of alcohol and duration of its effect.

Effect estimation

Once the case-control dataset has been assembled, effect estimation is the process of calculating the epidemiologic measures of interest, which in case-control studies is the odds ratio. At this stage, a potent source of misclassification in injury case-control studies arises from the grouping of etiologically unrelated injury outcomes into a single outcome measure. Broad injury categories such as 'childhood drowning' contain a wide range of

Table 2   Case-control data, odds ratios (OR), and 95% confidence intervals (CI) describing the association between visual defects on the risk of child pedestrian injury

| | Visual defects | | | | | |
| | (1) | | (2) | | (3) | |
	Yes	No	Yes	No	Yes	No
Case	9	191	9	144	9	121
Control	4	396	4	396	4	396
OR (95% CI)	4·67 (1·42 to 15·30)		6·19 (1·88 to 20·40)		7·36 (2·23 to 24·33)	

etiologically distinct scenarios.[22] The risk factors for drowning in public swimming pools are likely to be quite different from the risk factors for drowning in domestic pools. The combination of etiologically distinct outcomes is a form of outcome misclassification and will bias the odds ratio towards the null value.

The case-control data shown in table 2 describe the association between a visual defect and the risk of child pedestrian injury. The first column (1) in the table shows the distribution of subjects by exposure status and odds ratio for a case-control study where cases were all child pedestrian injuries occurring in a defined population. Close to one quarter of these pedestrian injuries involved either toddlers reversed over in the residential driveway or children backed into in car parks. It is likely that a visual defect is etiologically unrelated to this mechanism of injury. The data in column (2) shows the odds ratios when these cases were excluded. Of the remaining injuries, close to 16% occurred when children were struck by vehicles failing to stop at pedestrian crosswalks. Again a visual defect may not be etiologically important for this mechanism of injury. The third column shows the odds ratios for the association between visual defect and the risk of injury excluding these cases. The association, although imprecise, is considerably stronger. While it may be tempting to combine categories in order to obtain greater precision, the gain in precision is at the expense of bias.[4] Although this type of outcome misclassification is not unique to injury studies, because of the wide diversity of scenarios in which injuries occur, injury case-control studies are likely to be particularly vulnerable to this bias.

Summary

In the preceding examples a number of methodologic issues particularly relevant to case-control studies of injury have been considered. None of the issues raised are unique to injury case-control studies nor do any justify the isolation of injury epidemiology as a distinct domain of epidemiology. Instead the future application of epidemiologic methods to the identification of risk factors for injury is likely to emphasize the coherence rather than the fragmentation of epidemiologic principles.

1 Rivara FP, Wolf ME. Injury research: where should we go from here? *Pediatrics* 1989; **84**: 180–1.
2 Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. *Am J Epidemiol* 1992; **135**: 1029–41.
3 Miettinen OS. The 'case control' study: valid selection of subjects. *Journal of Chronic Diseases* 1985; **38**: 543–8.
4 Rothman KJ. *Modern epidemiology.* Boston: Little, Brown

and Company, 1986.

5 Steineck G, Ahlbom A, A definition of bias founded in the concept of the study base. *Epidemiology* 1992; **3:** 477–82.

6 Thomas S, Action C, Nixon J, Battistutta D, Pitt WR, Clark R. Effectiveness of bicycle helmets in preventing head injury in children: case-control study. *BMJ* 1994; **308:** 173–6.

7 Thompson R, Rivara F, Thompson D. A case-control study of the effectiveness of bicycle safety helmets. *N Engl J Med* 1989; **320:** 1361–7.

8 Schieber R, Branche-Dorsey C. Effectiveness of wrist guards in preventing wrist injuries to in-line skaters. *Am J Epidemiol* 1994; **139:** S8 (abstract).

9 Roberts I. Sole parenthood and the risk of child pedestrian injury. *J Paediatr Child Health* (in press).

10 Elley WB, Irving JC. The Elley Irving socioeconomic index 1981 census revision. *New Zealand Journal of Educational Studies* 1985; **20:** 115–28.

11 Roberts I, Norton R. Auckland children's exposure to risk as pedestrians. *NZ Med J* 1994; **107:** 331–3.

12 Roberts I, Keal M, Frith W. Pedestrian exposure and the risk of child pedestrian injury. *J Paediatr Child Health* 1994; **30:** 220–3.

13 Towner EML, Jarvis SN, Walsh SSM, Aynsley-Green A. Measuring exposure to injury risk in schoolchildren aged

11–14. *BMJ* 1994; **308:** 449–52.

14 Walsh SS, Jarvis SN. Measuring the frequency of 'severe' accidental injury in childhood. *J Epidemiol Community Health* 1992; **46:** 26–32.

15 Pless IB, Verreault R, Tenina S. A case-control study of pedestrian and bicyclist injuries in childhood. *Am J Public Health* 1989; **79:** 995–8.

16 Taylor RG. Assessment of childhood injury severity. *Australian Paediatric Journal* 1989; **25:** 6–9.

17 Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977; **105:** 488–95.

18 Graham CJ, Dick R, Rickert VI, Glen R. Left-handedness as a risk factor for unintentional injury in children. *Pediatrics* 1993; **92:** 823–6.

19 Roberts I. Differential recall in a case-control study of child pedestrian injuries. *Epidemiology* 1994; **5:** 473–5.

20 Rothman KJ. Induction and latent periods. *Am J Epidemiol* 1981; **114:** 253–9.

21 Roberts I, Lee-Joe T. Effect of exposure measurement error in a case-control study of child pedestrian injuries. *Epidemiology* 1993; **4:** 477–9.

22 Kemp A, Sibert JR. Drowning and near drowning in children in the United Kingdom: lessons for prevention. *BMJ* 1992; **304:** 1143–6.