Soz Präventivmed 1992; 37: 22–26

0303-8408/92/010022-05 $1.50 + 0.20/0

Social pediatrics

Case-control studies in pediatric epidemiology: Parent surrogates and potential pitfalls of inaccurate and selective recall

Harris Pastides [1], Robert J. Goldberg [2]

[1] Department of Biostatistics and Epidemiology, University of Massachusetts School of Public Health, Amherst
[2] Department of Medicine, University of Massachusetts Medical School, Worcester

The reasons for the widespread adoption of the case-control study design in research of pediatric diseases are numerous. First, most medical outcomes of interest appear after a latency period which separates them from their first, or cumulative, exposure by a substantial time period. We are usually unwilling to wait for the time which would be required by the cohort study where groups of healthy children would be followed for some prescribed length of time or until enough of them develop the disease. Even when the latency period is reasonably short, such as with diseases resulting from exposures during pregnancy, many adverse outcomes are too uncommon to provide a basis for a detailed and efficient analysis using the prospective cohort approach. Additionally, case-control studies are typically less expensive than cohort studies since expenditures in research are usually proportional to the duration of a study. Finally, a multiplicity of exposures can be systematically examined using this cost-effective design approach.

While there has been, historically, much criticism of the case-control study [1], this has been mainly a result of the mistaken opinion that it has inherent weaknesses which increase the likelihood that bias will occur. On the contrary, and as has been pointed out, biased case-control studies are due to incorrect subject selection, data collection or epidemiolgic analysis rather than to any intrinsic deficiencies of the study design [2]. Nevertheless, case-control studies in pediatric research present numerous opportunities for errors of interpretation; among the most potent of these is the ascertainment of accurate exposure histories. This paper proposes techniques which can be adopted to assess and improve the quality of information reported by parents or other surrogates in epidemiologic studies of childhood illness.

Parents as surrogates

In pediatric epidemiology, opportunities for inaccurate, incomplete, and selective recall are numerous. Since the cases and controls are often unable to recognize or communicate the majority of risk factors of interest, parent (or guardian) surrogates are the usual choice for the ascertainment of exposure information [3–5]. Even when studying older children and adolescents, many investigators choose to use parents as the primary source of information collection, ostensibly, in order to improve reporting quality. Another reason for parental interviews is for the study of risk factors in deceased children. In these studies, childrens' exposure may be estimated by evaluating the exposure status of the cases' parents, as in one study of Sudden Infant Death Syndrome [6].

Reasons why parents may inaccurately report childhood exposures are numerous. Most central is the fact that, due to the large proportion of a day which children spend in school, in day-care, or with other relatives, parents only partially share and recognize the physical and social environment of their children. This results in a substantial cumulative risk period for children which is, at best, only partially known to a parent. Specific exposures which might therefore be inaccurately reported by parents include: exposures in a school or day-care's physical environment such as the air, water, soil, play sand and paint, as well as those in the social environment including transmission of infectious agents between children or between adults and children; occurrence of minor injury or trauma; and level of emotional well being during times away from the respondent.

Risk factor misclassification which would result from this lack of complete knowledge might sometimes be expected to be random or non-differential; that is, parents of cases and controls would be expected to misclassify exposures to similar degrees. This would result in risk estimates which underestimate the underlying association and are biased toward the null hypothesis of no association between exposure and disease [2]. We should be especially concerned with this error when a study's results show a small association, or one that is absent, since the true association could only be stronger. When a substantial association is observed in a case-control study, the effect of random misclassification is usually not of primary concern.

More insidious is the bias on a risk estimate of having parents of cases and controls misclassify exposure information in different directions or to widely differing degrees. Such differential recall can bias risk estimates and obscure our understanding of a relationship in unforeseen but destructive ways. In pediatric epidemiology, there are situations where the parents of cases may be expected to systematically over-report, or under-report, childhood exposures. Over-reporting may be expected when studying exposures not directly attributable to direct parental actions or responsibility; these could include childhood exposures to x-rays, ambient environmental pollutants or physical contaminants. Since parents of cases are generally very concerned about identifying a cause of their child's illness, they may be more prone to associate and report characteristics of the environment as potential reasons for their child's illness than would parents of healthy controls. Parents of children with leukemia may, for example, be more likely to report that former residences or schools were proximal to high-tension power lines or hazardous waste sites, or to report a history of frequent viral infections, than parents of controls. Similarly, parents of cases may be more likely to be aware of relatives or neighbors who have had a similar disease diagnosis while parents of controls may not. In these instances the result could be a spurious positive association between the illness and the putative risk factor.

On the other hand, when investigating childhood exposures to factors for which parents have direct responsibility, such as exposure to parents' cigarette smoke, children's nutritional intake, or the quality of various types of parent-child interactions, parents of cases might provide responses which are perceived to reflect favorable parental behavior. For example, in a case-control study examining the relationship between adolescent suicide attempts (the outcome) and the amount and types of discipline in the home (the risk factor), parents of adolescents who have attempted or committed suicide might tend to report less extreme disciplinary practices than the actual experience. Alternatively, if parents believed that more or earlier disciplinary practices might have served to prevent the suicide their reponses might be influenced by this.

If both parents are questioned about former disciplinary actions with their offspring, differences in the perception and extent of agreement of these practices might occur between the mother and father thereby clouding data classification and interpretation. This example highlights the fact that trying to predict the direction and magnitude of parental surrogate bias is not straightforward, making the problem more difficult to identify as well as to correct.

To minimize and counteract the potential effects of differential recall by parents of cases and controls, several strategies exist which can be employed during the stages of study design, data collection, and data analysis of a case-control study.

Control group selection

In the design of case-control studies in pediatric epidemiology, consideration is most frequently given to control groups which are thought to be representative of the underlying "healthy" population which also gave rise to the cases [7]. In instances where biased recall may be a concern, however, a control group should be considered which includes parents of children with diseases or medical conditions about which a similar level of etiologic knowledge is available. If studying Reye's Syndrome and aspirin use, for example, the assessment of reporting bias would be facilitated by interviewing parents of controls who have conditions which have a similar level of etiologic uncertainty. If aspirin use is plausibly related to the medical condition of their children, then various factors which may influence the parents of cases to provide inaccurate exposure histories may similarly affect the parents of controls. As in any case-control study there should, of course, be no known or theorized link between the risk factor being studied and the control's diagnosis; rather, it is important only that such a link seem plausible. If such a control group is used in addition to a control group of healthy children, the investigator will have an opportunity to evaluate the possibility that reporting bias was present by comparing information provided by the parents of the healthy and sick control children.

Collection of exposure data from both parents

Depending on the risk factor being studied, another technique which might help mitigate reporting inaccuracy in pediatric studies is to collect information from both the mother and father, where possible. If conducted as a joint interview this strategy should help to prevent the willful misrepresentation of information by a parent, help to resolve discrepancies or points of confusion, and therefore serve to keep missing information about risk factors or control variables to a minimum. Separate interviews of mothers and fathers, on the other hand, can help assess the reliability of the information collected; unfortunately one cannot feel more confident about the validity of the data even if reasonable agreement between parents does appear to exist. Of course, interviewing both parents separately will require more time and money for data collection. It is important to remember that when interviewing both parents, whether interviews are joint or separate, coding decisions for subsequent

data analysis will eventually be required for responses for which parental disagreement exists.

Collection of exposure data from children and parents

When the children who are the subjects of epidemiologic studies are thought to be mature enough to understand and respond to carefully worded questions, consideration should be given to inviting them to be the primary information source. For example, in one study of the relationship of self image and perceptions of family cohesion to functional somatic complaints [8], responses obtained from adolescents (age range 11–18 years) served as the primary data for analysis. Standardized instruments were administered to adolescents with and without functional somatic complaints and no additional information was elicited from other sources (e.g., parents, teachers, school counselors). Of course, reasons why children might not provide factual information would also have to be considered in light of the specific research questions of a particular study.

A limited number of studies have examined the extent of agreement between information elicited from parents with that of their offspring, concordancy rates of collected information between parents, and differences in perceived attitudes and/or behaviours in the child with that of surrogate respondents. In a study of 151 children of ages 6–12 years attending school in upstate rural New York, 24 hour dietary recalls completed by the children and their mothers were compared; the accuracy of the child's recall of his/her school lunch eaten was also compared with that of actual tray observation. The results of this study showed, in general, reasonably good agreement between the mothers and their children in terms of the different food groups consumed irrespective of the child's age. The children were also able to recall on average approximately two-thirds of the food actually eaten during a typical school lunch with this proportion increasing with age from approximately 60% of first graders to approximately 80% of 4th graders[9].

A second study examined teachers' and parents' perceived changes in children's behaviours over two-week blocks of time in children between the ages of 6–17 years who were referred for neurologic consultation because of attention deficit disorders. These assessments were carried out as part of a controlled, double-blind placebo, crossover trial. While a number of children were lost to follow-up, thereby raising concerns as to the generalizability of the study findings, parental and teacher summary scores and comments were in agreement as to changes in the children's behaviors in the majority of children assessed [10]. However, considerable variability was observed in both parents' and teachers' ratings of the responses of children to the active intervention.

Finally, in a study designed to assess practices of keeping sick children home from day care centers, the opinions of day-care staff, working mothers and pediatricians were compared as to the perceived guidelines about when to send sick children home. Fifty-two licensed day-care centers in three North Carolina counties comprised the study sample [11]. Questionnaire responses were elicited from 302 staff members, 134 mothers and 69 pediatricians. Significant differences were reported in the exclusionary practices for children ill at day care. The temperature level considered to be reflective of fever was perceived to be different for the pediatricians, mothers, and day-care staff. Staff of the various day-care centers were significantly more likely than the other two comparison groups to request immediate school pick-up of young children for each level of temperature considered to be associated with illness. These results focus attention on either attitudinal differences between mothers, physicians and day care center staff as to when children should be sent home with illnesses, or to differences in staff policies and/or knowledge of pediatric illnesses. These examples reinforce the advantage of soliciting responses from the child, parent and/or other surrogate respondent to accurately ascertain the observed risk factor and disease experience.

Review of medical records

Another way to attempt confirmation of surrogate reported exposure information is by reviewing medical records. This would be considered, of course, only when the historical exposure information being sought is routinely recorded in the medical record. It should be kept in mind that the unstructured interview between physician and parent or child which forms the foundation of the child's medical record is perhaps more susceptible to recall error than the research interview because the information given was not intended to be part of a standardized study. In any event, this strategy offers the advantage of documenting exposure information provided by the parents which was elicited at an earlier stage of a child's illness (or before the illness) when the influences for inaccurate reporting discussed earlier might have been fewer or less intense. Another advantage of using medical records is that the elapsed time between putative exposure and the reporting of it to the pediatrician would be shorter than the time elapsed between the exposure and the reporting of it to a study interviewer, thereby possibly improving accuracy. Finally, one would only utilize medical records as a potential source of validation if exposure information for cases and controls would be equally likely to be contained in them.

Validity scales

Yet another method for the control of recall bias in case-control research is the validity scale approach as suggested by Raphael[12]. With this technique, an investigator includes a number of "fake" but plausible risk factors in the interview or questionnaire; these factors constitute the validity scale. When respondents assert an unusually large number of exposures for their child, the possibility of overreporting is hypothesized. An overall validity scale is calculated for parents of cases and controls and these scores are then compared. This quantitative estimate of validity is later entered as an independent term for each subject in multivariate analyses as a way to control for its effect. From a practical standpoint, this method results in a longer interview than would otherwise be necessary; more critically, one must consider that some children, be they cases or controls, will in truth have a larger number of exposures to a variety of plausible risk factors.

Interviewer training

Through careful training and sensitization, interviewers employed by a research study can help to create a relaxed yet motivating environment for the communication of factual information. Since epidemiologic studies frequently solicit very personal or "private" information, special care in the recruitment, training, and supervision of the data collection staff can be fruitful in encouraging sincere and thoughtful responses from participants. Since it is often impractical to "blind" interviewers as to the case/control status of respondents, training efforts should stress the investigator's insistence on impartiality in the data collection process. Occassionally, interviewers may not be told the specific study hypotheses but, in general, factors included in an interview might be thought of as potential risk factors by study personnel. The use of postal questionnaires as a means of collecting epidemiologic data may be considered as a way of preventing interviewers bias and is, of course, cheaper than conducting personal or telephone interviews. Other strengths and weaknesses of this approach have been addressed in detail[13].

Conclusion

Since case-control studies will undoubtedly continue to be a mainstay methodology in pediatric epidemiology, strict attention to, and control of, a variety of obvious as well as subtle potential biases is required. To this end, the methods presented in this paper can be utilized to minimize susceptibility to certain biases or inaccuracies in the reporting of exposures by parents of cases and controls. Each technique presented, of course, has its own advantages and disadvantages as well as applicability to particular research questions. The method chosen should depend on the particular situation as well as the resources of the study. In all cases, however, the evaluation and control of inaccurate or incomplete recall as a result of the use of parent surrogates in pediatric case-control studies should be a carefully planned and monitored activity.

Summary

The case-control study is quite popular as a study design for exploring associations between risk factors and disease in pediatric epidemiology. Since data concerning exposures to the child are often collected through interviews with parents or other surrogates, researchers should be aware of the opportunities for bias due to inaccurate or incomplete recall. Methods which exist for the control of this problem are presented. These include: the selection of control groups with childhood conditions of similar etiologic uncertainty as the disease being studied; collecting exposure data from both parents; collection of data from children where possible; diligent interviewer training; reviewing clinical records; and use of validity scales. Strengths and weaknesses of these strategies are discussed.

Résumé

Etudes cas-contrôle en épidémiologie pédiatrique: problèmes médologiques liés au rôle des parents comme substitut d'un souvenir sélectif

L'étude cas-contrôle est fréquemment employée pour explorer de possibles associations entre les facteurs de risque et les maladies en épidémiologie pédiatrique. La plupart des données concernant l'exposition des enfants étant obtenues auprès des parents ou d'autres proches, les chercheurs doivent être conscients des erreurs systématiques potentielles dues aux souvenirs imprécis ou incomplets. Les méthodes permettant de contrôler ces erreurs incluent: la sélection de témoins avec des pathologies dont l'étiologie est aussi incertaine que celle de la maladie étudiée, la collecte de données d'exposition chez les deux parents, la collecte de données auprès des enfants lorsque c'est possible, la formation appropriée des intervieweurs, l'analyse des dossiers cliniques, et l'utilisation de chefs de validation. Les forces et les faiblesses de ces stratégies sont discutées.

Zusammenfassung

Fall-Kontroll-Studien mit Kindern: Die Befragung der Eltern als mögliche Fehlerquelle

In epidemiologischen Studien, die Erkrankungen von Kindern betreffen, werden oft Fall-Kontroll-

Studien angewendet, um die Beziehung zwischen Risikofaktoren und Erkrankung zu studieren. Da die Angaben zur Exposition des Kindes in der Regel von den Eltern oder anderen Erwachsenen stammen, muss der Untersucher sich der Gefahr bewusst sein, dass unvollständige oder ungenaue Angaben dieser Stellvertreter das Resultat verfälschen könnten. Es werden Methoden vorgestellt, die es ermöglichen, diese Gefahr zu kontrollieren: als Kontrollpersonen Patienten mit einer Erkrankung wählen, deren Ursache so unklar ist wie jene der untersuchten Krankheit; die Angaben von beiden Elternteilen erheben; das Kind selber befragen, wann immer dies möglich ist; die Interviewer sorgfältig schulen; Krankengeschichten beiziehen; mit der Frage nach Scheinursachen einen Zuverlässigkeits-Index konstruieren. Stärken und Schwächen dieser Strategien werden diskutiert.

References

1 *Feinstein AR.* Scientific standards in epidemiologic studies of the menance of daily life. Science 1988; *242*: 1251–1263.
2 *Rothman J.* Modern Epidemiology. Boston: Little Brown and Company, 1986.
3 *Starko LM, Ray G, Dominquez LG, Stranberg WL, Woodall DR.* Reyes syndrome and saheylate use. Pediatrics 1980; *66*: 859–864.
4 *Miller DS, Lin EHB.* Children in sheltered homeless families: reported health status and use of health servies. Pediatrics 1988; *81*: 668–673.
5 *Zittelli BJ, Miller JW, Gartner JC et al.* Changes in lifestyle after liver transplantation. Pediatrics 1988; *82*: 173–180.
6 *Lewis NC, McBride JT, Brooks JG.* Ventilatory chemosensitivity in parents of infants with sudden infant death syndrome. J Pediatr 1988; *113*: 307–311.
7 *Goldberg RJ, Pastides H, Ellison RC, et al.* Uses of the case-control and cohort epidemiological approaches in pediatric practice and research. Ped Res 1985; *19*: 787–790.
8 *Robinson DP, Green JW, Walker LS.* Functional somatic complaints in adolescents: relationship to negative life events, self-concept and family characteristics. J Pediatr 1988; *113*: 588–593.
9 *Emmons L, Hayes M.* Accuracy of 24-hour recalls of young children. J Am Dietet Assoc 1973; *62*: 409–415.
10 *McBride MC.* An individual double-blind crossover trial for assessing methylphenidate response in children with attention deficit disorder. J Pediatr 1988; *113*: 137–145.
11 *Landis DE, Earp JL, Sharp M.* Day-care center exclusion of sick children: comparison of opinions of day-care staff, working mothers, and pediatricians. Pediatrics 1988; *81*: 662–667.
12 *Raphael K.* Recall bias: a proposal for assessment and control. Int J Epidemiol 1987; *16*: 167–169.
13 *Kelsey JL, Thompson WD, Evans AF.* Observational Epidemiology. New York: Oxford University Press, 1986.

Address for correspondence:
Harris Pastides, MD, PhD, Associate Professor
at the University of Massachusetts School of Public Health,
407 Arnold House, Amherst, MA 01003/USA