

# ENDGAMES

## STATISTICAL QUESTION

### Nested case-control studies: advantages and disadvantages

Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, London, UK

Researchers investigated whether antipsychotic drugs were associated with venous thromboembolism. A population based nested case-control study design was used. Data were taken from the UK QResearch primary care database consisting of 7 267 673 patients. Cases were adult patients with a first ever record of venous thromboembolism between 1 January 1996 and 1 July 2007. For each case, up to four controls were identified, matched by age, calendar time, sex, and practice. Exposure to antipsychotic drugs was assessed on the basis of prescriptions on, or during the 24 months before, the index date.<sup>1</sup>

There were 25 532 eligible cases (15 975 with deep vein thrombosis and 9557 with pulmonary embolism) and 89 491 matched controls. The primary outcome was the odds ratios for venous thromboembolism associated with antipsychotic drugs adjusted for comorbidity and concomitant drug exposure. When adjusted using logistic regression to control for potential confounding, prescription of antipsychotic drugs in the previous 24 months was significantly associated with an increased occurrence of venous thromboembolism compared with non-use (odds ratio 1.32, 95% confidence interval 1.23 to 1.42). The researchers concluded that prescription of antipsychotic drugs was associated with venous thromboembolism in a large primary care population.

Which of the following statements, if any, are true?

- The nested case-control study is a retrospective design
- The study design minimised selection bias compared with a case-control study
- Recall bias was minimised compared with a case-control study
- Causality could be inferred from the association between prescription of antipsychotic drugs and venous thromboembolism

### Answers

Statements *a*, *b*, and *c* are true, whereas *d* is false.

The aim of the study was to investigate whether prescription of antipsychotic drugs was associated with venous

thromboembolism. A nested case-control study design was used. The study design was an observational one that incorporated the concept of the traditional case-control study within an established cohort. This design overcomes some of the disadvantages associated with case-control studies,<sup>2</sup> while incorporating some of the advantages of cohort studies.<sup>3 4</sup>

Data for the study above were extracted from the UK QResearch primary care database, a computerised register of anonymised longitudinal medical records for patients registered at more than 500 UK general practices. Patient data were recorded prospectively, the database having been updated regularly as patients visited their GP. Cases were all adult patients in the register with a first ever record of venous thromboembolism between 1 January 1996 and 1 July 2007. There were 25 532 cases in total. For each case, up to four controls were identified from the register, matched by age, calendar time, sex, and practice. In total, 89 491 matched controls were obtained. Data relating to prescriptions for antipsychotic drugs on, or during the 24 months before, the index date were extracted for the cases and controls. The index date was the date in the register when venous thromboembolism was recorded for the case. The cases and controls were compared to ascertain whether exposure to prescription of antipsychotic drugs was more common in one group than in the other. Despite the data for the cases and controls being collected prospectively, the nested case-control study is described as retrospective (*a* is true) because it involved looking back at events that had already taken place and been recorded in the register.

Selection bias is of particular concern in the traditional case-control study. Described in a previous question,<sup>5</sup> selection bias is the systematic difference between the study participants and the population they are meant to represent with respect to their characteristics, including demographics and morbidity. Cases and controls are often selected through convenience sampling. Cases are typically recruited from hospitals or general practices because they are convenient and easily accessible to researchers. Controls are often recruited from the same hospital clinics or general practices as the cases. Therefore, the selected

cases may not be representative of the population of all cases. Equally, the controls might not be representative of otherwise healthy members of the population. The above nested case-control study was population based, with the QResearch primary care database incorporating a large proportion of the UK population. The cases and controls were selected from the database and therefore should be more representative of the population than those in a traditional case-control study. Hence, selection bias was minimised by using the nested case-control study design (*b* is true).

The traditional case-control study involves participants recalling information about past exposure to risk factors after identification as a case or control. The study design is prone to recall bias, as described in a previous question.<sup>6</sup> Recall bias is the systematic difference between cases and controls in the accuracy of information recalled. Recall bias will exist if participants have selective preconceptions about the association between the disease and past exposure to the risk factor(s). Cases may, for example, recall information more accurately than controls, possibly because of an association with the disease or outcome. Although in the study above the cases and controls were identified retrospectively, the data for the QResearch primary care database were collected prospectively. Therefore, there was no reason for any systematic differences between groups of study participants in the accuracy of the information collected. Therefore, recall bias was minimised compared with a traditional case-control study (*c* is true).

Not all of the patient records in the UK QResearch primary care database were used to explore the association between prescription of antipsychotic drugs and development of venous thromboembolism. A nested case-control study was used instead, with cases and controls matched on age, calendar time, sex, and practice. This was because it was statistically more efficient to control for the effects of age, calendar time, sex, and practice by matching cases and controls on these variables at the design stage, rather than controlling for their potential confounding effects when the data were analysed. The matching variables were considered to be important factors that could potentially confound the association between prescription of antipsychotic drugs and venous thromboembolism, but they were not of interest as potential risk factors in themselves. Matching in case-control studies has been described in a previous question.<sup>7</sup>

Unlike a traditional case-control study, the data in the example above were recorded prospectively. Therefore, it was possible to determine whether prescription of antipsychotic drugs preceded the occurrence of venous thromboembolism. Nonetheless, only association, and not causation, can be inferred from the results of the above nested case-control study (*d* is false)—that is, those people who were exposed to prescribed antipsychotic drugs were more likely to have developed venous thromboembolism. This is because the observed association between prescribed antipsychotic drugs and occurrence of venous thromboembolism may have been due to confounding. In particular, it was not possible to measure and then control for, through statistical analysis, all factors that may have affected the occurrence of venous thromboembolism.

The example above is typical of a nested case-control study; the health records for a group of patients that have already been collected and stored in an electronic database are used to explore the association between one or more risk factors and a disease or condition. The management of such databases means it is possible for a variety of studies to be undertaken, each investigating the risk factors associated with different diseases or outcomes. Nested case-control studies are therefore relatively inexpensive to perform. However, the major disadvantage of nested case-control studies is that not all pertinent risk factors are likely to have been recorded. Furthermore, because many different healthcare professionals will be involved in patient care, risk factors and outcome(s) will probably not have been measured with the same accuracy and consistency throughout. It may also be problematic if the diagnosis of the disease or outcome changes with time.

Competing interests: None declared.

- 1 Parker C, Coupland C, Hippisley-Cox J. Antipsychotic drugs and risk of venous thromboembolism: nested case-control study. *BMJ* 2010;341:c4245.
- 2 Sedgwick P. Case-control studies: advantages and disadvantages. *BMJ* 2014;348:f7707.
- 3 Sedgwick P. Prospective cohort studies: advantages and disadvantages. *BMJ* 2013;347:f6726.
- 4 Sedgwick P. Retrospective cohort studies: advantages and disadvantages. *BMJ* 2014;348:g1072.
- 5 Sedgwick P. Selection bias versus allocation bias. *BMJ* 2013;346:f3345.
- 6 Sedgwick P. What is recall bias? *BMJ* 2012;344:e3519.
- 7 Sedgwick P. Why match in case-control studies? *BMJ* 2012;344:e691.

Cite this as: *BMJ* 2014;348:g1532

© BMJ Publishing Group Ltd 2014