BMJ

BMJ 2012;344:e4209 doi: 10.1136/bmj.e4209 (Published 21 June 2012)

ENDGAMES

STATISTICAL QUESTION

Observational study designs

Philip Sedgwick *senior lecturer in medical statistics*

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

Researchers investigated whether pioglitazone was associated with an increased risk of bladder cancer in people with type 2 diabetes.[1] Use of pioglitazone, an oral antidiabetic agent in the thiazolidinedione class, is controversial.[2] The researchers used the general practice research database to extract data from individual patients' primary care records between 1988 and 2009. A cohort of 115 727 patients with type 2 diabetes was established, with patients entering the cohort if they had been newly treated with oral hypoglycaemic agents. Patients were considered to have been exposed to pioglitazone if they had ever taken it, and measures of duration of use and cumulative dosage were recorded.

In the cohort 376 cases of bladder cancer were diagnosed. Patients were considered to be a case if their cancer was diagnosed at least one year after entry to the cohort, to account for latency. Each case was matched to as many as 20 controls on year of birth, year of cohort entry, sex, and duration of follow-up. A total of 6699 controls were identified. The researchers reported that the use of pioglitazone was associated with an increased risk of bladder cancer among people with type 2 diabetes.

Which one of the following study designs best describes that used above?

a) Case-control study

b) Cohort study

c) Cross sectional study

d) Nested case-control study

Answers

Nested case-control study (answer *d*) best describes the study design used above.

Case-control, cohort, cross sectional, and nested case-control studies are all observational studies by design. Researchers in such studies do not intervene in any way but simply observe the behaviour and risk factors of the study participants and record whether a disease or condition develops.

The nested case-control study (answer *d*) incorporates a case-control study nested within a larger study. Typically the case-control study is nested within a prospective cohort study, as in the example above. Case-control and cohort studies have been described in previous questions.[3 4] In the example above, data in the general practice research database were collected prospectively. The cohort of patients with type 2 diabetes who had been newly treated with oral hypoglycaemic agents was identified after the medical record data between 1988 and 2009 had been collated. Patients within the cohort who developed bladder cancer were identified as cases. Controls were patients in the same cohort without a diagnosis of bladder cancer. Each case was matched with up to 20 controls on year of birth, year of cohort entry, sex, and duration of follow-up. The cases and controls were then compared for past exposure to pioglitazone.

In a case-control study (answer *a*), two groups of people are chosen on the basis of their disease status: those with the condition or disease (the cases) and those without (the controls). Case-control studies are retrospective in design: individuals are asked about past exposure to proposed risk factors. The aim is to provide insight into which factors may raise or lessen the risk of the disease. Cases are typically obtained from hospital lists or disease registries. Controls are often selected from hospitals or the community. The choice of controls is important: they should be representative of the population at risk of developing the condition or disease. The cases and controls are not part of a larger study, as in a nested case-control study.

The nested case-control study design in the example above overcomes some disadvantages associated with case-control studies while incorporating advantages of the cohort study. It is the nested nature of the case-control study within the cohort study that provides the design's strength. The choice of controls does not pose the same concerns as in a case-control study. All members of the prospective cohort were representative of the population at risk and therefore at risk of developing bladder cancer. Furthermore, as the general practice research data were collected prospectively, it was possible to ascertain whether exposure to pioglitazone preceded development of bladder cancer, permitting assessment of causality. As data on exposure to pioglitazone and other risk factors for bladder cancer were collected prospectively, recall bias will have been minimised. Recall bias, described in a previous question,[5] is the systematic

p.sedgwick@sgul.ac.uk

difference between cases and controls in the accuracy of recalled information regarding exposure to risk factors.

The matching of cases and controls meant that not everyone in the cohort of 115 727 patients with type 2 diabetes and who had been newly treated with oral hypoglycaemic agents was included in any subsequent analysis. However, this would have resulted in relatively minor losses in statistical accuracy. Matching of cases and controls minimised confounding by reducing systematic differences between the two patient groups. Any differences between cases and controls—that is, diagnoses of bladder cancer—would therefore not have resulted from differences in year of birth, year of cohort entry, sex, and duration of follow-up but rather from differences in prescriptions of pioglitazone and other risk factors. Matching on the basis of birth year, cohort entry, sex, and follow-up meant that these potential confounders were adjusted for in the study design, which was more efficient than adjusting during statistical analyses**.**

In a cohort study (answer *b*) all members of the cohort would be included in the analysis. In the example above, the cohort was all patients in the general practice research database whose data were collected prospectively. However, only those patients with type 2 diabetes that had been newly treated with oral hypoglycaemic agents were initially selected. Then from this subgroup the cases of bladder cancer and a total of 6699 matched controls were chosen for analysis.

Cross sectional studies (answer *c*) are carried out at a single point in time. They may be used to record people's perceptions, behaviours, and attitudes concerning, for example, screening for breast cancer. Cross sectional studies are also suitable for estimating the prevalence of a medical condition, such as depression, in the population.

Competing interests: None declared.

1 Azoulay L, Yin H, Filion KB, Assayag J, Majdan A, Pollak MN, et al. The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case-control study. *BMJ* 2012;344:e3645.
2 Moynihan R. European drug agency extends review of safety of pioglitazone. *BMJ* 2011;342:d4105.
3 Sedgwick P. Case-control studies. *BMJ* 2009;339:b4135.
4 Sedgwick P. Cohort studies. *BMJ* 2010;340:c1002.
5 Sedgwick P. What is recall bias? *BMJ* 2012;344:e3519.

Cite this as: *BMJ* 2012;344:e4209