

ENDGAMES

STATISTICAL QUIZ

Observational study design II

Philip Sedgwick *senior lecturer in medical statistics*

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

Researchers investigated whether the use of oral or transdermal hormone replacement therapy was a risk factor for stroke. Data were taken from the General Practice Research Database, a computerised database of anonymised longitudinal medical records collected prospectively in primary care. A cohort of women was identified, aged 50-79 years with records between 1 January 1987 and 31 October 2006, and without a diagnosis of stroke on the date of registration with their general practice. A woman was identified as a case if she experienced a stroke during follow-up, and up to four controls were randomly selected from the cohort. Controls were matched to cases on age (within one year) at the date of diagnosis of the stroke, general practice where registered, and the year of joining the practice. The risk of stroke was increased with oral hormone replacement therapy of any dose but only with transdermal patches containing high doses of oestrogen.¹

Which of the following best describes the study design used above?

- a) Case-control study
- b) Cohort study
- c) Nested case-control study
- d) Cross sectional study

Answers

Answer *c* best describes the study design used. A nested case-cohort study is an observational design that incorporates a case-control study “nested” within an already established cohort. The cohort comprised all women in the General Practice Research Database aged 50-79 years with records between 1 January 1987 and 31 October 2006, and without a diagnosis of stroke when registered with their general practice. The researchers did not intervene in any way—they simply recorded exposure to hormone replacement therapy and whether a stroke occurred. If a woman experienced a stroke she was identified as a case. For each case, up to four controls—that is, women without a diagnosis of stroke—were randomly selected from the same cohort. Cases and controls were matched for age (within one year) at the date of the diagnosis, the general practice attended, and the year of joining the practice. Cases and controls

were identified retrospectively and after data collection had finished.

Answer *a* is false. A case-control study would have started with two groups of individuals being identified on the basis of their disease status.² The cases would be those women who had a stroke, whereas controls would be those without a diagnosis. In a case-control study, cases and controls are typically identified from easily accessible sources, such as hospital clinics. Information about the risk factor—hormone replacement therapy—would have been obtained retrospectively by examining past records, interviewing each woman, and possibly interviewing their relatives. A case-control study is a retrospective study, and the quality of information collected is typically subject to bias because it depends on medical records and memory recall.

Although a nested case-control study is similar in design to a case-control study, it has several important advantages. The above study was prospective in design, ensuring greater accuracy in the collected data. Cases and controls were identified only when data collection had finished. Furthermore, because cases and controls were selected from the same cohort, unlike in case-control studies, selection bias of controls was not a problem. In case-control studies, controls are often selected from hospital outpatient clinics, which can result in selection bias. The controls may not be representative of the general population, and systematic differences may exist between cases and controls in social background and general health status.

Answer *b* is false. The above study incorporated a prospective cohort. The cohort consisted of women aged 50-79 years, with records between 1 January 1987 and 31 October 2006, without a diagnosis of stroke on the date of registration with their general practice. However, when women were diagnosed with a stroke they were identified as a case, and up to four controls were identified, matched on several variables. The study design is therefore best described as a nested case-control study design. Only a fraction of the records from the cohort were used for subsequent analysis. Because cases and controls were matched on a series of variables, these potential confounders were adjusted for when designing the study, which is more efficient

than analysing the entire cohort and making adjustments during statistical analyses.³

Answer *d* is false. A cross-sectional study usually involves a questionnaire or survey that attempts to record people's behaviour, experiences, or attitudes.⁴ Sometimes cross sectional studies investigate disease status and exposure to potential risk factors. People are contacted only once and all at the same time. A cross sectional study does not involve the retrospective or prospective collection of data, so it would not have been suitable for establishing the association between stroke and hormone replacement therapy. It would also have been difficult to

establish the temporal association between use of hormone replacement therapy and stroke using a questionnaire.

Competing interests: None declared.

- 1 Renoux C, Dell'Aniello S, Garbe E, Suissa S. Transdermal and oral hormone replacement therapy and the risk of stroke: a nested case-control study. *BMJ* 2010;340:c2519.
- 2 Sedgwick P. Endgames: Case-control studies. *BMJ* 2009;339:b4135.
- 3 Sedgwick P. Endgames: Matching. *BMJ* 2009;339:b4581.
- 4 Sedgwick P. Observational study design. I *BMJ* 2011;342:d1742.

Cite this as: [BMJ 2011;342:d1903](#)