Commentary: matched cohorts can be useful

Stephen Evans

London School of Hygiene and Tropical Medicine, London WC1E 7HT

Stephen Evans *professor of pharmacoepidemiology*

Correspondence to: Stephen.Evans@lshtm.ac.uk

Most *BMJ* readers are familiar with matched case-control studies but fewer will be familiar with matched cohort studies. Case-control studies are based on selecting cases of a disease and then finding people who are as similar as possible to the cases. The study by Helms et al is not a case-control study; people were selected not on the basis of having, or not having, the outcome of interest (in this instance mortality) but on the basis of being exposed or not to something that may affect mortality.

Matched cohort studies have been published in the *BMJ* before—for example, a study examining air bags and deaths of car drivers.[1] Helms et al have used similar methods with Danish national data to look at *Salmonella* (reference 19 of their paper). A common feature of these studies is the existence of a large database in which the individuals who are exposed (to bacterial infection or air bags) can be compared with similar unexposed people. Helms et al used record linkage between databases, obtaining data from microbiology laboratories to define exposed patients and using the national Danish civil registration system to obtain unexposed people from the general population. They also used the registration system to obtain outcome data on subsequent mortality for exposed and unexposed people and two further databases to determine possible confounding from hospital admissions for diseases other than bacterial infection.

The main method of analysis for cohort studies is to use the time taken to an event that is the outcome under study, a survival analysis. The outcome is usually death, but it could be another event such as diagnosis of myocardial infarction or cancer. Cohort studies usually have to be very large to obtain a sufficient number of outcome events. This may make their costs prohibitive, but with electronic databases the costs can be greatly reduced. Similarly, the costs of carrying out matching in cohort studies have restricted their use. Matching prevents the possible association between the matching factors and the exposure at the start of the study, although not necessarily associations occurring as an observational study progresses. Matching should be taken into account in the (conditional) analysis, as has been done by Helms et al.[2]

Matching may not increase statistical power (efficiency) but it does not introduce bias (as it does in case-control studies).[3] With large databases any small loss in efficiency may be unimportant, and the convincing power to the reader of the similarity of the exposed and unexposed cohorts at the start is a gain.

What factors should be used for matching? Helms et al used age, sex, and county of residence. They have used a 1:10 exposed:unexposed ratio. They have also adjusted the survival analysis for comorbidity, based on eight different diagnostic groups. It is possible to match for morbidity or other risk factors, but it would make matching difficult and may not offer any gains. An alternative, used particularly in drug safety, is to match on a "propensity" score.[4] This score measures the likelihood of being given the treatment rather than the likelihood of having the outcome. The purpose is to reduce confounding in either the design or the analysis so that comparisons are valid.

Scandinavia has better national databases than elsewhere, but the United Kingdom has good databases based on general practitioners' computer records. The potential of these is considerable, and matched cohort designs could be used more often. Concerns over confidentiality of records may make this difficult, but it is to be hoped that good epidemiology is not going to be stopped because of misguided ethicists and lawyers.[5]

Competing interests: None declared.

1 Cummings P, McKnight B, Rivara FP, Grossman DC. Association of driver air bags with driver fatality: a matched cohort study. *BMJ* 2002;324:1119-22.
2 Rothman KJ, Greenland S. *Modern epidemiology.* 2nd ed. Philadelphia: Lippincott-Raven, 1998.
3 Greenland S, Morgenstern H. Matching and efficiency in cohort studies. *Am J Epidemiol* 1990;131:151-9.
4 Wang J, Donnan PT. Propensity score methods in drug safety studies: practice, strengths, and limitations. *Pharmacoepidemiol Drug Saf* 2001;10:341-4.
5 Walton J, Doll R, Asscher W, Hurley R, Langman M, Gillon R, et al. Consequences for research if use of anonymised patient data breaches confidentiality. *BMJ* 1999;319:1366.