

Study design IV. Cohort studies

Kate Ann Levin

Dental Health Services Research Unit, University of Dundee, Dundee, Scotland, UK

Previously in this series I have given an overview of the main types of study design and the techniques used to minimise biased results. In this article I describe more fully cohort studies, their uses, advantages and limitations.

Evidence-Based Dentistry (2003) **7**, 51-52. doi:10.1038/sj.ebd.6400407

A cohort study is one in which a group of subjects, selected to represent the population of interest, is studied over time. Much like a cross-sectional study, information is collected about the outcome of interest and exposure to risk factors, but in cohort studies subjects are followed over time. Subjects are disease-free at the outset of the study and at distinct points in time, data are collected relating to health outcomes and exposure to risk factors. This type of study is observational and used to examine causal factors. Cohort studies may be either fixed, where the study subjects do not vary over time and dropouts are not replaced, or dynamic, where new subjects enter the study in accordance with eligibility criteria.

Prospective versus retrospective cohort studies

Prospective studies, those studies carried out from the present time into the future, can be tailored to collect specific exposure data, but there may be a long wait for events to occur, particularly where the outcome of interest is associated with old age. Studies can therefore be expensive to carry out and are prone to high dropout rates, although these can be overcome by incorporating a dynamic study design.

In contrast (Figure 1), retrospective or historical cohort studies look at medical events from some timepoint in the past up to the present time. The advantage of historical cohort studies is that the information is available immediately. There may be difficulty in tracing subjects for such studies, however. A further disadvantage is the reliance on the memory of subjects and/or the quality of recorded information.

Prospective cohort study design is more commonly used because accurate and complete data, necessary for historical cohort

studies, are rarely available. The rest of this article refers to prospective cohort studies unless otherwise stated.

Selection and follow-up of subjects

The probability of having the outcome of interest will be affected by the selection of subjects into a cohort study.

The 'healthy entrant effect' occurs because of the necessity of a disease-free status on entry to the study. Initially, subjects are seen to have lower levels of disease than might be true of the population in general, with an acceleration of disease rate over time. Following cohorts from birth may overcome healthy entrant effects of this sort, but may result in a lengthy and costly study, depending on the average age at onset of disease.

Follow-up of subjects is carried out to monitor changes in health status over time. It is essential to have a mechanism in place

that achieves the lowest possible dropout rate from the study. Loss to follow-up will increase with the length of study. A greater concern than number of dropouts are any systematic differences related to the outcome or exposure to risk factors, between those who drop out and those who stay in the study. Analysis of data must include a comparison of risk factors between individuals who remain in the study and those who have dropped out. If loss to follow-up is ignored, the reliability of study conclusions may be called into question.

Inferring causality

To infer causality with any degree of certainty, an experimental study design is required. The longitudinal nature of cohort studies, however, enables the assessment of causal hypotheses, as it is known if exposure occurred prior to outcome. Furthermore, measuring changes in levels of exposure over time alongside changes in outcome measure gives an insight into the dose-response relationship between exposure and outcome. Higher levels of exposure, associated with higher levels of outcome provide further argument for causality.

Hypotheses of cohort studies

Analysis of cohort study data takes one of two forms. The first is a straightforward

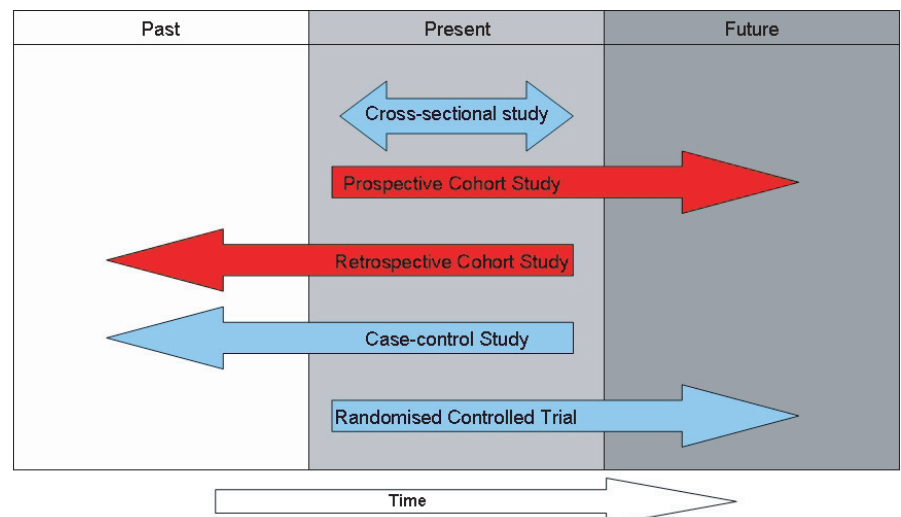


Figure 1. Prospective and retrospective studies

Table 1.			
	Exposed to risk factor:		
Outcome	Yes	No	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	N

comparison of two groups, those with exposure and those without. The hypothesis (H_0) is then: on average, the disease in the exposed group is no different to that of the unexposed group. An example of this might be comparing the dental health of those who smoke with those who do not.

The second type of analysis occurs when there are more than two groups. An example might be where smokers are split into those who do not smoke, those who smoke fewer than 20 cigarettes a day and those who smoke over 20 cigarettes. Where there is this type of grouping in strata, hypotheses fall into three broad categories:

- H_0 : on average, the outcome rates in groups of subjects are no different to that of the population as a whole.
- H_0 : on average, the outcome rates in groups of subjects are no different from each other.
- H_0 : on average, there is a linear trend in outcome rates across groups.

Analysis of data

Using the data collected in a cohort study, the following statistics may be obtained:

Crude rates of outcome: this is the number of individuals with the outcome out of the

total cohort study size. In Table 1 this is given by:

$$\frac{(a+b)}{n}$$

Standardised rates and ratios of the outcome can also be calculated, using demographic information so that rates of the outcome are adjusted for other potential risk factors such as age and sex.

Risk ratio of outcome: the risk of the outcome in exposed subjects relative to those not exposed is given in Table 1 by:

$$\frac{a/(a+c)}{b/(b+d)}$$

Where one or more of the second set of above hypotheses is being examined, more complex data analysis is used. Regression analysis allows investigation of two or more groups of subjects whilst adjusting for characteristics that might act as confounding risk factors.

Advantages of cohort studies

- The temporal dimension, whereby exposure is seen to occur before outcome, gives some indication of causality
- Can be used to study more than one outcome

- Good for the study of rare exposures
- Can measure the change in exposure and outcome over time
- Incidence of outcome can be measured

Disadvantages of cohort studies

- Costly (less so for retrospective) and may take a long time, particularly where onset of the outcome measure can occur both early and late on in life
- Require accurate records for retrospective studies
- When studying rare outcomes, a very large sample size is required
- Prone to dropout
- Changes in aetiology of disease over time may be hard to disentangle from changes observed as age increases
- Selection bias: a difference in incidence of the outcome of interest, between those who participated and those who did not, would give biased results

Recommended reading

1. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume II. The Design and Analysis of Cohort Studies. IARC Scientific Publications no. 82. Town: Oxford University Press; 1987.
2. Grimes DA, Schultz KF. Cohort studies: marching towards outcomes. Lancet 2002; 359:341–345.
3. Squires BP, Elmslie TJ. Cohort studies: what editors want from authors and peer reviewers. CMAJ 1990; 143:179–180.
4. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. Am J Epidemiol 2005; 161:280–288.