8

COHORT STUDIES

Objectives

After reading this chapter, you should be able to:

1. Distinguish between open and closed source populations as they relate to cohort study design.
2. Describe the major design features of risk-based and rate-based cohort studies.
3. Identify hypotheses and population types that are consistent with risk-based cohort studies.
4. Identify hypotheses and population types that are consistent with rate-based cohort studies.
5. Elaborate the principles used to select and measure the exposure in cohort studies.
6. Design and implement a valid cohort study to investigate a specific hypothesis.

8.1 INTRODUCTION

This chapter is concerned with the design, implementation, analysis, and reporting of cohort studies. The word **cohort** denotes a group of study subjects that has a defined characteristic in common, and in epidemiological study design, the characteristic of interest is the exposure status (exposed and non-exposed study subjects). In a cohort study design, we follow subjects from exposure to outcome (Grimes and Schulz, 2002). In this sense, a cohort study closely resembles a controlled trial without the randomisation of exposure. The study subjects are usually individuals, but can be groups, such as families. Most frequently the outcome of interest is the occurrence of a specific disease, although other outcomes such as birth weight, quality of life, obesity/body mass index, or blood pressure might be the main focus of the study.

If the exposure status of potential study subjects is known beforehand, the selection of the study groups from the source population can be based directly on the exposure status (*eg* when we select 2 cohorts, a group of exposed and one of non-exposed study subjects). If the exposure status is not known beforehand, another approach is to select a single group (cohort) of subjects within which there is likely to be a range of the exposure of interest. We denote the first design, with 2 groups defined by exposure status, as a **cohort study** and the latter as a **single cohort** or **longitudinal** study; however, for purposes of study design we need not differentiate between them. In both instances, following their selection, we would ensure that the study subjects meet the inclusion/exclusion criteria for the study, verify the exposure status of each subject, and ensure that they do not have the outcome disease of interest. Then, we would observe the study subjects for a defined follow-up period, and when this period ends we would compare the incidence of the disease in the 2 or more groups defined by exposure status. As mentioned, cohort studies are similar to the structure of clinical trials, but without randomisation of exposure. This similarity is seen as an advantage in terms of making causal inferences (Davis and Al-Alem, 2011); however, as Wilcox and Wacholder (2008) point out, there are often different findings from the 2 approaches and we need to find ways to 'narrow the gap'. A straightforward example of a cohort study is shown in Example 8.1.

In some cohort studies, the outcome is measured on a quantitative scale (*eg* weight and physical measurements of the baby; see Example 8.2). Nonetheless, comparing the weight of exposed and non-exposed subjects, or their offspring, also fits the cohort study design paradigm.

Example 8.1 Retrospective risk-based cohort study investigating readmission rates (risk) of patients discharged against medical advice

This cohort study by Choi *et al* (2011) was based on data recorded at one hospital. The outcome was the patient's readmission to the hospital within 14 days of discharge. The dichotomous exposure was discharge against medical advice (exposed=DAMA) versus being discharged with medical advice (unexposed=DWMA). To reduce potential confounding, each DAMA patient was matched with one DWMA patient of the same 10-year age group, gender, and having been hospitalised with similar clinical characteristics. DAMA was based on a patient not returning to the hospital after 6 hours without medical advice. If a patient had more than one DAMA, the first occurrence only was included in the analysis. The major outcome was being readmitted within 14 days of discharge; this fits the risk-based outcome because all discharged patients were observed for the full 14-day risk period. Twenty-six per cent of DAMA patients were readmitted by 14 days versus 3% of the DWMA. To account for the matching, conditional logistic regression was used for the statistical analysis of the data.

Example 8.2 A retrospective cohort study with outcome measured on a continuous scale

Crane *et al* (2011) conducted a study through interviews of women who had given birth between April 2001 and March 2009 in two provinces in Canada. More than 11,000 women took part; 11% self-declared exposure to environmental tobacco smoke. A number of outcomes were measured including the physical dimensions of the infant, the Apgar score, respiratory distress syndrome, and stillbirths. Using multiple regression to control for a number of potential confounding variables, the study concluded that environmental tobacco smoke was associated with a number of adverse perinatal outcomes including lower birth weight and smaller body size (linear regression), as well as increased stillbirths (logistic regression).

As mentioned, the basis of the cohort study design is to compare the frequency of the outcome (usually disease or a health measure) in 2 or more groups of subjects that are similar in all regards except for exposure; the quality of the study will depend on how closely the real study approaches that ideal. In many cohort studies the outcome is 'time-to-outcome', and hence survival models are used for their analysis. Levin (2006) and Soh and Saw (2010) describe the basic features and analysis of cohort studies. Euser *et al* (2009) describe a number of cohort design features in the context of kidney disease research. Hood (2009) reviews cohort study design in the context of nursing research. Comprehensive and classical reviews of cohort study design and analysis are available (Prentice, 1995; Samet and Munoz, 1998a; 1998b), as well as a more recent series of articles on the critical appraisal of cohort studies, from a clinical perspective (Mamdani *et al*, 2005; Normand *et al*, 2005; Rochon *et al*, 2005). The latter discussions focus on methods to prevent selection bias and confounding. We will discuss the reporting of cohort studies later in this chapter (see Section 8.9).

Each specific study presents its own unique challenges, but the starting point for all studies is to clearly and concisely state the objective. This includes defining the **target** (the population to which inferences will be made) and **source populations** (the population from which the study group will be drawn), the **unit of observation (***eg* individuals or groups**)**, the **exposure**, the **disease**, the **follow-up period**, and the setting (*ie* context or venue) of interest. If sufficient biological facts are known, the hypothesis should indicate the amount, or duration, of exposure that is believed to be needed to 'cause' the disease. Clarifying the study objective often helps us decide whether current or past exposure is relevant, whether lifetime exposure or exposure in a narrower window of time is important, whether repeated measures of exposures are required and if so, how to handle changes in exposure status.

Depending on the availability of suitable records, cohort studies can be performed **prospectively** or **retrospectively** (Euser *et al*, 2009). In a prospective study design, the disease has not occurred at the time the study starts; conversely, in a retrospective study design, the follow-up period has ended, and the disease event has occurred when the study subjects are selected (Hudson *et al*, 2005) (see Examples 8.1–8.5). Prospective studies provide the opportunity for more detailed information-gathering (see Examples 8.6, 8.7, and 8.9) and attention to recording the details of interest than do retrospective studies which require suitable existing databases.

We close this section with an 'interesting' example of creating virtual cohorts. McCartney *et al* (2010) created a virtual cohort based on the consumption of meat products (the exposure factor) available at a suspect meat outlet in Paisley, Scotland. Because many of the actual meat products had been sold by the time the investigation started, the delicatessen's computer records of purchases and sales were used to create the virtual population of families exposed to

Example 8.3 A retrospective cohort study with matching on multiple factors based on the propensity score

Mehta *et al* (2010) reported on a retrospective population-based cohort design to investigate risk factors for falls and fractures in older adults (equal to or greater than 50 years of age). The exposure of interest was the use of atypical versus typical antipsychotic agents. Data were obtained from a health-plan database. The use of antipsychotic drugs was based on prescription-claim data. Patients taking atypical antipsychotic agents were matched with patients not taking atypical antipsychotics.

More than 60 covariates were included in the propensity-score estimation which attempts to identify people with similar covariate backgrounds. Then, the Greedy 5-1 matching technique was used to select exposed and non-exposed subjects with similar propensity scores; this technique reduces the matched pair bias caused by incomplete and inexact matching. Each exposure group contained 5,580 people among which a total of 825 falls/fracture cases occurred. There were 450 fractures in the atypical users and 375 in typical users. The hazard ratio did not differ significantly between the two exposure groups. However, it was noted that the use of any antipsychotic medication for more than 90 days was associated with a significant increased risk (*HR*=1.8) of either a fall or fracture.

the meat products. The distribution of the various types of meat to 'purchasing units' (usually a family or household) formed a number (n=25) of exposure groupings. The outcome was whether members of the purchasing unit had *E. coli* O157 cultured from their stool samples; members of 14 purchasing units were infected, no culture positives were noted in the remaining 967 purchasing units. This allowed the investigators to determine the relative risk of infection by meat product on a purchasing-unit basis. The authors argued that their approach circumvented recall biases that can arise when using a traditional case-control study approach to identify high-risk foods in this setting.

8.2 Selecting the study group

When selecting the exposure groups, it is best if the groups come from one clearly defined source population. This helps ensure that the study subjects have numerous characteristics in common, and can reduce the background 'noise' and/or the risk of unmeasured confounding. As an example, the source population might be administratively or geographically defined, be a member of a specific group, or a patient at a clinic or hospital facility. Other study-subject specific eligibility criteria such as age, race, socioeconomic class *etc* can be used to define the study groups; all eligibility criteria should be specified explicitly. Most often, the study subjects are chosen purposely, not randomly, from the source population. When obtaining the study group using clinic/hospital/laboratory records, we assume that the source population is that population of individuals who would have used these facilities if they needed care. Although this non-random selection increases the risk of selection bias, it is often the only practical way to proceed with a study. Harcombe *et al* (2011) describe four key design features to reduce potential bias when trying to obtain a volunteer study group using a postal survey.

When using secondary data, it can be important to verify that the source population contains valid data on exposure and outcomes; for example, hospital data may not always be valid for assessing the care of patients with specific disease conditions (Jonkman *et al*, 2001). This aspect of study design will be discussed in greater detail in Chapter 12.

We close this section with another interesting example of obtaining study subjects. Swerdlow *et al* (2011) used a 'generations study design' to obtain their study group for an investigation of

Example 8.4 Retrospective risk-based cohort study of surgical complications

Kelz *et al* (2009) compared morbidity and mortality following more than 56,000 general and vascular surgical procedures performed between 2001 and 2004. Data were obtained from the National Surgical Quality Improvement Program database. Time of operation was the exposure variable of interest; this was grouped into 7, 2-hour periods starting at 7:30 am with all procedures starting between 9:30 pm and 7:30 am in a single group. Rates (risk) of morbidity and mortality are given in the report. Following adjustment for patient and procedure characteristics, mortality within 30 days following the procedure had a moderately strong association (*OR*=1.22) with start times after 9:30 pm. Similarly, morbidity within 30 days following the procedure was increased (odds ratio equal 1.32) for those surgical procedures starting after 9:30 pm. However, when the surgery was classified as being an emergency, no odds ratios were significant. Thus, the excess crude risk for late-night, early-morning surgeries was explained by the nature of the clinical cases.

breast cancer. The initial study group was composed of women from a breast cancer charity, and volunteers for the study (generation 1). Each of these women was asked to nominate a family member or friend (generation 2), and if they agreed to participate they were asked to repeat this process (up to 18 generations). Initially, more than 300,000 women >16 years of age were interested in the study. Of these, more than 200,000 agreed to participate and 47% of these returned the questionnaire and became members of the study. Ninety-two per cent of the members returned the blood sample. Of interest was that a mail-out schedule was used to distribute blood vials so that a reasonably constant flow of plasma could be obtained, processed, and analysed. The authors discuss the strengths and weaknesses of their design in moderate detail, and this will be useful reading for those proposing to conduct a large cohort study.

8.2.1 Sample Size

Usually, when developing a cohort design, we assume that an equal number of exposed and non-exposed individuals will be selected. However, if cost or other practicalities dictate different sample sizes across exposure categories, then this can be accounted for. Initial estimates of sample size can be performed assuming the disease is measured by risk (see Section 8.2.2), as shown in Chapter 2. This approach is often sufficient for initial estimates of sample size, even if the population is open and a rate-based study must be used (see Section 8.2.3).

Computer software for sample-size estimation usually allows for unequal sample sizes and repeated measures when planning the study. More recent software is available to estimate sample size when using multivariable regression models, or proportional hazard models (see Section 19.14) (Barthel *et al*, 2006; Latouche *et al*, 2004) for analyses. Latouche and Porcher (2007) describe how to estimate sample size when competing risks are present. Cai and Zeng (2007) discuss sample size estimation in cohort and case-cohort designs. Matsui (2005) discusses sample-size estimation for rate-based designs where the outcome is survival time. Mazumdar *et al* (2006) discuss sample-size estimation in strata-matched designs with survival-time outcomes. Basagana *et al* (2011) describe a sample-size estimator, including the software, that accounts for time-varying exposures. The researcher needs to specify the proposed number of observation times for assessing the outcome, the within-subject variation of exposure, and the exposure covariance between observation times (this can vary from compound symmetry to auto-regressive). Accounting for the time-varying aspects allows the use of smaller sample sizes than if a time invariant structure is assumed. Although the subject is not covered in this

Example 8.5 A retrospective risk-based cohort study of HIV-positive women

This retrospective risk-based cohort study followed approximately 250 HIV-positive women who received care at the Ottawa Hospital General Campus Immunodeficiency Clinic from 2002 to 2005 (Leece *et al*, 2010). Women who had undergone a hysterectomy were excluded from this study. Relevant information was obtained from the patient's medical reports and included the number of clinic visits during the study. Care was taken to validate the accuracy of abstraction from the charts. The outcome of interest was undergoing a cervical screening test, and information on this was obtained from electronic laboratory records. Pap-test results were reported according to the Bethesda system. Overall, 145 of the 250 women underwent the screening test and 48 (33%) of these had at least one abnormal test result. The exposure groups were based on the demographic characteristics of the patients, their HIV status, and whether or not they had primary health-care providers. Analysis was by χ^2 tests (continuous predictors were categorised) and by a general linear (logistic) model. The 12 women without a primary-care provider were less likely to undergo screening (n=8 [67%]) than the 84 of 206 [41%] women with health-care providers; *RR*=1.6). Abnormal screening test results are common in women with HIV but the only significant predictor was a recent low CD4 cell count. The authors included much descriptive data in the report and this helps the reader gain a better understanding of the context, and results, of the study.

text, sample-size formulae for studies based on micro-arrays are available (Matsui *et al*, 2008; Oura *et al*, 2009).

8.2.2 Risk-based (cumulative incidence) designs

This is the simplest form of cohort study, but it requires that a number of assumptions be met in order for the design to be valid. First, the exposure groups must be defined at the start of the study and remain unchanged during the study (*ie* they are **fixed cohorts**). Second, the study groups must be **closed** in that all subjects must be observed for the full risk period (*eg* Examples 8.1, 8.4, and 8.5). In this design, it is best if there are few (or no) losses (losses include study subjects that develop other diseases, or die from other outcomes, as well as withdrawals from the study). If the percentage of study subjects that is lost becomes large (some use >10% as a cutpoint), this will cast doubt on the validity of the study findings. For these reasons, risk-based designs work best for diseases (outcomes) with a relatively short risk period. Since all subjects are observed for the full risk period, this allows the calculation of risk in each of the exposure categories. For chronic diseases such as many cancers, where the risk period is lifelong and often greater than the follow-up period of the study, a rate-based design is often preferred.

In a 2X2 table, the summary algebraic format for classifying the study subjects in a risk-based cohort study is shown below:

	Exposed	Non-exposed	Total
Diseased	a_1	a_0	m_1
Non-diseased	b_1	b_0	m_0
Total	n_1	n_0	n

In this design, we select n_1 exposed and n_0 non-exposed individuals from the N_1 exposed and N_0 non-exposed individuals in the source population. Having ensured that none of the study subjects has the disease at the start of follow-up, we observe all subjects for the defined follow-

up period. During this time, we note that a_1 exposed subjects develop the disease out of the n_1 exposed subjects, and a_0 non-exposed subjects out of the n_0 non-exposed subjects develop the disease. Overall, we observe a total of m_1 diseased and m_0 non-diseased subjects. The 2 risks (R) of interest are:

$$R_1 = a_1/n_1 \qquad \text{and} \qquad R_0 = a_0/n_0 \qquad \textit{Eq 8.1}$$

Note The denominator of interest is the number of subjects in each exposure category (Examples 8.1 and 8.3)

8.2.3 Rate-based (incidence density) designs

In many instances, not every study subject is under observation for their full risk period. This is especially true if the source population is dynamic and the follow-up period is long. Thus, some subjects may be added to the study group part way through the biological risk period for the outcome of interest. In addition, a significant proportion of subjects may withdraw from the study part way through the follow-up period. Also, the exposure status of subjects might change during the study period. In these situations, we cannot just count the number of exposed and non-exposed subjects; rather, we need to accumulate the amount of 'at-risk time' contributed by each study subject in each of the exposed and non-exposed categories. Thus, the denominator of the rate becomes the amount of study-subject-time per exposure category and this requires a rate-based approach to study design and analysis (Examples 8.6–8.7).

In this design, each of the initially selected exposed and non-exposed subjects contributes 'at-risk' time to the denominator of the rates until they develop the disease, or are lost to the study, or their observation ends because the study is terminated. If new individuals are added to the study group during the follow-up period, then the amount of time-at-risk for each study subject is added to the appropriate exposed or non-exposed category.

In a 2X2 table, the summary format for classifying the study subjects in a rate-based cohort study is shown below.

	Exposed	Non-exposed	Total
Diseased	a_1	a_0	m_1
Person-time at risk	t_1	t_0	T

Example 8.6 A prospective longitudinal rate-based cohort study

Chalmers *et al* (2011) conducted this prospective longitudinal rate-based cohort study to identify risk factors for injury in amateur club rugby. The study group consisted of 704 male rugby players, aged 13 years and over. The 'time' component was a 'game' and in total there were 6,263 player-games. The exposures included a variety of factors including age, ethnicity, rugby experience, height, weight, body mass index, physical activity, cigarette smoking, previous injury, playing while injured, position, training, time of season, warm-up, foul-play, weather conditions, ground conditions, and protective equipment. Poisson regression was used to estimate the effect of each factor after adjusting for all other factors. Since it was reasonable to assume constant injury rates over the study period, Poisson regression could be used for the analysis. Some of the findings included Pacific Island versus Maori ethnicity (injury incidence rate ratio (*IR*=1.5); ≥40 hours strenuous physical activity per week (*IR*=1.5); playing while injured (*IR*=1.5); very hard ground condition (*IR*=1.5); foul-play (*IR*=1.9), and use of headgear (*IR*=1.2).

Example 8.7 A prospective rate-based cohort study with numerous examples of coding exposures

This rate-based cohort study was based on women taking part in the Women's Health Initiative Observational Study (Luo *et al*, 2011). The study involved more than 90,000 women aged 50 to 79 years at over 40 clinical centres throughout the United States between 1993 and 1998. A number of exclusion criteria were included and explained in the report. More than 12,000 of the women were excluded because of a history of cancer. All information on exposures and confounders was collected at baseline. Smoking was classified as never, former, or current. Women who were smokers, or had smoked, were asked the age at which smoking started, the number of cigarettes smoked per day, and the duration of smoking in years. Among former smokers, age at quitting was collected. A combined variable, pack-years, was calculated by multiplying the total years of smoking by the number of cigarettes smoked per day divided by 20. Information on exposure to passive smoking was also collected. Invasive breast cancer was the outcome of interest.

After an average of 10.3 years of follow-up, 3,520 incident cases (4.0 per 1,000 person-years in non-smokers; 4.6 per 1,000 person-years in former and current smokers) of invasive breast cancer had been identified. The impact of exposure to passive smoking was investigated among the 41,000 women who had never smoked; 1,692 (4 per 1,000 person-years) of these women had developed invasive breast cancer. A variety of methods were used in order to investigate the cumulative dose of exposure to passive smoking. Given the long period of follow-up, Cox proportional hazards models (instead of Poisson models) were used to estimate the hazard ratios from smoking and/or exposure to environmental smoke. Smoking increased the risk of invasive breast cancer among former smokers (*HR*=1.09), and among current smokers (*HR*=1.16). Among lifetime non-smokers, 88% were exposed to passive smoking. Only those women in the highest level of exposure to passive smoking had an increased risk of breast cancer. The authors noted that they did not observe a significant dose-response trend.

Initially, we might select n_1 exposed and n_0 non-exposed individuals from the source population. All these subjects are followed for the duration of their time-at-risk within the study period. In so doing, we observe a_1 exposed cases of the disease out of t_1 person-time units of exposure and a_0 non-exposed cases of the disease out of t_0 non-exposed person-time units. Here t_1 is the sum of all of the exposed time-at-risk contributed by the study subjects whether they developed the disease or not. Similarly t_0 is the summed time-at-risk study subjects contributed in the non-exposed category. The 2 rates (I) of interest that we wish to estimate are:

$$I_1 = a_1/t_1 \qquad \text{and} \qquad I_0 = a_0/t_0 \qquad \textit{Eq 8.2}$$

If the follow-up time is relatively short, the above rates can be used to measure disease frequency, and analysed with Poisson models. If the follow-up time is sufficiently long that the assumption of a constant rate over the entire follow-up period is highly suspect, survival analysis methods can be used to overcome this difficulty (see Chapter 19). Examples 8.3, 8.6 and 8.7 describe 3 rate-based cohort studies.

8.3 THE EXPOSURE

In cohort studies, our objective is to identify the consequences of a specific exposure factor. The exposure refers to any potential cause of disease and, for example, these might range from characteristics of the study subject, to infectious or noxious agents, to environmental, housing, or food-related factors. Exposures that can be manipulated are of special interest as these lead more directly to control of the disease than factors that we cannot manipulate. Although

measuring exposure might seem simple at first glance, careful thought should be given to the manner in which it is measured and expressed (see Example 8.9). Each study design should include the details of what constitutes exposure, and whenever possible, we should specify how long after an **exposure threshold** is reached before one might reasonably expect to see the disease arise from that exposure (if indeed it was a cause of the disease (*ie* the **induction period**)). The more complex the exposure, the more important it becomes to validate that the assessment/measurement of exposure in the proposed study is valid. Heinavaara *et al* (2011) give an example of validating exposure to mobile phone use. Duffy *et al* (1994) discuss design and analysis when both exposure and outcome are measured multiple times during the study period.

Exposure status can be measured on a dichotomous scale (*eg* exposed or non-exposed) (see Examples 8.1 and 8.3), an ordinal scale (*eg* low, medium, or high dose) (Example 8.10), or a continuous scale (*eg* organisms per gm of feces, ppm of a toxin in air or water *etc*) (Example 8.8). Exposure can be expressed separately in terms of dosage and duration or as a combination of the two (*eg* pack-years for quantifying cigarette exposure). Often it is necessary to decide whether lifetime exposure, historical exposure, or current exposure (relative to the time of disease occurrence) are the best measures of 'exposure'. Because both exposure status and 'time exposed' are crucial components of a valid cohort study, it is vital to reduce the measurement error for exposure. To achieve this often requires a thorough understanding of the 'exposure agent' and of the etiologically high-risk period for disease causation.

8.3.1 Permanent exposures

The exposure might be a permanent factor (*ie* time-invariant) or a factor that can change over time (*ie* time varying). Permanent exposures include factors such as sex or race, and one-time exposures such as vaccination. Permanent and 'one-time' exposures are relatively easy to measure, but even here a moment's thought might suggest that defining 'sufficient' with respect

Example 8.8 Cohort study with primary risk factor measured on a continuous scale

Warensjo *et al* (2011) conducted a cohort study that investigated risk factors for fractures and osteoporosis in women who were members of the Swedish Mammography Cohort (this was established between 1987 and 1990 and involved more than 90,000 women). Seventy-four per cent of these women completed a questionnaire covering diet and lifestyle. Fracture events were identified through linkage to the Swedish National Patient Registry. Details on the diet questionnaire and its validation are included in the paper. Calcium intake was the major exposure of interest. The lifetime use of dietary supplements and multivitamins was reported. One calcium dose was considered to be 500 mg if it was from calcium supplements, and 120 mg if it was from multivitamins. Total calcium intake included supplemental calcium.

Both dietary and total calcium intake in the 1997 food-frequency questionnaire correlated well with estimates from 14 repeated, 24-hour recalls over one year (r=0.77). The authors examined the relationship between quintiles of cumulative dietary calcium intake and incidence of fracture and osteoporosis. The end of the follow-up was December 31, 2008. Both Cox proportional hazard models and logistic models were used for analysis of the data. The findings of the study indicated an association between a chronic, low-dietary intake of calcium and an increased incidence of fractures and osteoporosis. Above the base level of calcium intake there were only minor differences in incidence of fracture or osteoporosis. In people with the highest dietary calcium intakes, the rate of hip fracture was somewhat increased.

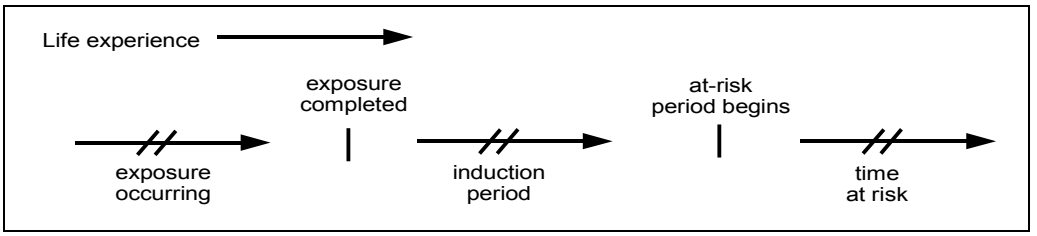

Fig. 8.1 Life experience with exposure, induction period and time at risk

to vaccination or smoking might be more complex than it first appears. Age at vaccination, or at initiation of smoking could be important features of exposure. In any event, for factors where the exposure is based on a threshold or dosage, the amount of exposure necessary to deem an individual as being 'exposed' needs to be clearly stated (sometimes this is one of the objectives of the study itself (Rohan *et al*, 2007)). In early studies of a health problem, the objective might be to assess if there is an exposure threshold, and if so, at what exposure level? When a detailed definition of what constitutes exposure is available, if the disease event occurs before exposure is completed in an individual, it should not be included as an outcome event in the analysis because exposure has not been completed (so it could not have caused the disease). These issues are shown in Fig. 8.1.

When exposure is measured on a continuous scale but, for purposes of the analysis, categorisation of exposure is desirable, the criteria for classification of the exposure should be clearly explained.

8.3.2 Non-permanent exposures

Non-permanent exposures can include factors such as food intake, lifestyle, or environmental exposures that can change over time, or study-subject-specific factors such as vaccination where the timing (age, stage of pregnancy, *etc*) of the procedure can be important. For these exposures, both the timing and the extent of the exposure might be important to measure and analyse. This adds complexity to the measurement of the exposure factor (*eg* the timing of lifestyle changes and exposure to smoking in Example 8.9). Sometimes a simple summary measure of exposure will suffice (*eg* years smoked), whereas in other studies more complex measures of exposure are needed (*eg* cumulative pack-years smoked). The more information that can be collected on exposure the better, such as the exposure level(s) when exposure started, and when (if) exposure stopped, because it adds credibility to conclusions about causal

Example 8.9 Prospective cohort study with numerous examples of coding exposures

Rohan *et al* (2007) published their design of a prospective cohort study based largely on alumni from 3 universities in Ontario, Canada who were recruited over a 3-year period, 1995-1998. The major outcome was new cases of cancer. Participants completed baseline lifestyle and food-frequency questionnaires, measured their waist and hip circumferences and provided hair and toenail specimens (for trace element and DNA analysis). More than 73,000 individuals were recruited and 97% of them provided biological specimens. The exposure factors of interest include exercise lifestyle factors, molecular markers, and the characteristics of the subject's diet. There is a good discussion on creating compound nutritional variables in the context of measuring nutritional intake using food frequency questionnaires and subsequently, verifying the data.

relationships, is more useful for preventive action or intervention, and enhances our biological understanding of the problem. Examples 8.7 and 8.8 are cohort studies with exposures that changed over time.

To obtain the exposure time, for each study subject, the time-at-risk in each exposure category accumulates from the moment exposure is completed until the event of interest occurs, the subject becomes lost to follow-up, or the study ends. With losses to follow-up, time-at-risk accumulates until the last date exposure status is known (if the precise time of loss is unknown, use the midpoint of the last known exposure period). If there is a known induction period following completion of exposure, then, until that period is over, the time at risk of 'exposed' individuals should be added to that of the non-exposed group. Some researchers prefer to discard the disease experience during the induction period for exposed individuals because of uncertainties about the duration of the induction period. In the face of uncertainty about these effects, this is likely the best choice providing there is sufficient time-at-risk in the exposed group to maintain precision.

Note If the exposure status changes, an individual study subject can accumulate time-at-risk in both exposed and non-exposed categories. Previously non-exposed subjects will contribute time-at-risk to the exposed category after the exposure threshold is reached. Similarly, if previously exposed individuals become non-exposed, we would add the non-exposure time (of previously exposed individuals) to the non-exposed cohort only after the time period when any lag effects that could be present had ended. Provided lag effects are minimal, when different exposure categories exist for the same study subject, the exposure category assigned to subjects who develop the disease is that level of exposure the subject was in at the time the outcome event occurred.

As noted above, exposure status of study subjects can be classified as exposed or non-exposed (*ie* a dichotomous exposure) (Examples 8.1 and 8.3) or perhaps on an ordinal level of exposure category (*eg* Example 8.10—never, former, or lifetime alcohol consumption). However, in many studies, exposure is measured on a continuous scale (cigarettes or packs per day) and the

Example 8.10 Estimating the population attributable fraction in a cohort study

Schutze *et al* (2011) reported on a multicentre prospective cohort study that, from 1992 to 2000, recruited about 520,000 randomly selected men and women between 35 to 70 years of age from 10 European countries. The main exposure factor of interest was alcohol consumption, and the outcome of interest was the new occurrence of a cancer. At recruitment, exposure was measured as alcohol in grams per day. Prior consumption of alcohol was based on self-reported consumption of beer, wine, and spirits. Based on the consumption in the past (and at recruitment) 3 categories of alcohol exposure were created—never, former, and lifetime consumers. Information on the incidence of cancer was obtained through record linkage with regional cancer registries. The follow-up period ended between 2002 and 2005.

Given the size and complexity of this study, the details on the analytical procedures were equally complex. Separate statistical models were used for men and women, and most models contained a large number of potential confounders. Essentially, the impact of alcohol consumption, as measured by the regression coefficient, was combined with information from the populations on the prevalence of alcohol consumption above recommended levels. The authors also examined for interactions between smoking and alcohol consumption. If causality is assumed, consuming alcohol beyond the recommended levels could be responsible for 10% of all cancers in men and 3% of all cancers in women.

threshold to complete exposure may either be unknown, or it is deemed more appropriate to model the exposure-outcome association in a dose-response manner (*eg* Example 8.8). As in other instances, maintaining the continuous scale has advantages because the categorisation of a continuous exposure variable usually results in loss of information. In this instance, one might relate the disease frequency (*ie* risk or rate) to exposure on a continuous exposure scale using an appropriate regression model.

Hernan *et al* (2009) have published an extensive discussion of the issues involved with observing the study subjects over the follow-up period when there are time varying exposure factors of interest.

8.4 Disease as exposure

Disease can serve as an exposure for other outcomes such as other diseases, mortality or quality of life measures. Lazo *et al* (2011) reported on a study based on following more than 11,000 adults, for 18 years, who participated in the Third National Health and Nutrition Examination Survey. The exposure of interest was the presence of non-alcoholic, fatty-liver disease (NAFLD). The prevalence of NAFLD with elevated levels of liver enzymes in the source population was 3.1% and 16.4% had NAFLD with no increased levels of liver enzymes. Those with NAFLD but with normal liver enzyme levels had a similar risk (hazard ratio) of mortality and cause specific mortality as those without NAFLD. This was also true for those who had fatty-liver disease and increased levels of liver enzymes.

8.5 Ensuring exposed and non-exposed groups are comparable

If the study subjects in the different exposure groups are not comparable (*ie* not exchangeable) with respect to factors related to both the outcome and exposure, a biased (*ie* confounded, see Chapter 13 for a discussion of confounding) assessment of the exposure-outcome association can result (Klein-Geltink *et al*, 2007). As Hernan (2012) notes, this is a key reason to prefer randomised experiments over observational studies because exchangeability is expected in the former. Investigators conducting observational studies need to use their expert knowledge to identify and measure all potential confounders in the hope they achieve exchangeability conditional on the measured covariates. Unfortunately, exchangeability cannot be empirically tested in observational studies, so we never know if we have succeeded.

In general, one or more of the following 3 approaches can be used to help ensure that the exposed and non-exposed groups are comparable except for their exposure status. The first of these approaches is applied prior to subject selection and involves the use of exclusion (also called restricted sampling) of study subjects. In this approach, we identify variables likely to be confounders and then we restrict the selection of study subjects to those that have only one level of these variables (*eg* include only one age, one race, or one sex of person in the study). This prevents confounding by the specified factors, serves to reduce the background variability among study subjects, and might help reduce confounding from other unknown factors (see Section 13.2.1 for further discussion on restriction).

A second approach can be used at the time of study-subject selection and involves matching the level of confounders in study subjects across the exposure categories. To accomplish this, we identify major (*ie* strong) confounding variables, and then select the non-exposed subjects so that they have the same level of the confounder as the exposed subjects (the exact criteria for

matching should be specified and reported) (see Example 8.1). Matching can help achieve greater study efficacy as well as prevent confounding in cohort studies. When there are numerous confounders to control, the process can be generalised using propensity scores (Example 8.3) (see Section 13.3 for a discussion of matching and Section 13.8 for propensity scores).

The third method to prevent confounding is to use analytic control (see Example 8.2). In this approach, we identify and measure the important confounders and then use statistical control (*eg* ranging from Mantel-Haenszel-type stratification to multivariable regression approaches) during the analysis to adjust for these confounders (see all of Chapter 13 for a more detailed discussion of confounding). Hernan (2012) discusses 'exchangeability' and other features needed for obtaining unbiased measures of association and extending these with inferences about causation.

8.6 Follow-up period

To enhance the validity of a cohort study, the follow-up process must be as complete as possible and unbiased with respect to exposure status. Achieving unbiased follow-up may require some form of observer-blinding process as to exposure status. 'Blinding' can be implemented in both prospective and retrospective studies (although the latter has more limited options). For example, in a prospective study, one set of researchers that is unaware of the exposure status can be assigned the task of follow-up. In a retrospective study, the researchers reviewing records for the outcome should be kept unaware of the exposure status whenever possible. In either situation, the fact and date of outcome occurrence should be as accurate as possible to reduce the possibility of bias and measurement errors. If passive surveillance for cases is used, cases occur when identified (*eg* this might be the date of first symptoms, or physician examination). With active surveillance and regular evaluation of study subjects, it is feasible to get more accurate data on time of outcome occurrence.

Unless the follow-up period is short, it is helpful to enumerate and characterise the population at risk at specified times during the study period, as noted by Tooth *et al* (2005). Collecting ancillary information is useful to help manage issues such as loss to follow-up because of competing risks including death or relocation of study subjects, and to assess if censorship is unrelated to exposure. Chang *et al* (2009) described an analytic method (shared parameter models) to account for losses to follow-up and reduce bias, since in many studies, one cannot assume that the losses are randomly distributed.

8.7 Measuring the outcome

Although the most frequent outcome in cohort studies is the occurrence of a specific disease, measured as either a risk or a rate, the disease outcome could be measured on an ordinal scale (*eg* none, mild, moderate, severe), or the outcome could be measured on a continuous scale (*eg* weight, birth weight, quality of life index *etc* (see Example 8.2). For example, Harley *et al* (2011) investigated the association between exposure to polybrominated diphenyl ethers (PBDEs—a flame retardant) and infant birth weight, length, head circumference, and length of gestation. Exposure was indicated by serum PBDE levels during week 26 of pregnancy. Here, both the exposure and the outcome were measured on a continuous scale.

Each study will need explicit protocols for determining the occurrence and timing of outcome

events. Clear definition(s) of **diagnostic criteria** are useful to ensure as few diagnostic errors as possible. This can prove difficult in retrospective studies when only the summary diagnostic information is available. The specific diagnostic criteria should be included in the study plan for prospective studies. When possible, in prospective studies, ensuring **blinding** of the diagnosticians is helpful to equalise, but not necessarily reduce, diagnostic errors.

When the outcome (*eg* disease) is measured as incidence, strictly speaking, this requires at least 2 examinations: the first at the start of the follow-up period to ensure that the study subjects did not have the disease of interest, and the second to investigate whether or not (and when) the disease developed during the observation period. Including only new disease events circumvents the reverse-causation problem from measuring prevalence as well as ensuring that the associations are not biased by duration-of-disease effects and survival bias (see Chapter 12). In retrospective studies, one often has to assume freedom from the disease at the start of the follow-up period; whereas in prospective studies, it is desirable to formally ensure that the study subjects are free of the disease at the start of follow-up.

If clinical diagnostic data are used to indicate the occurrence and timing of the disease event, the incident date usually will be based on time of diagnosis not on time of occurrence of disease. For diseases that might remain in the subclinical state for extended periods, ignoring this difference could lead to inferential errors about causal associations. If the study group is screened for the disease event at regular intervals, the time of occurrence of the disease should be placed at the midpoint between examinations.

One of the advantages of a cohort study is that we can assess multiple outcomes from a given exposure factor. In terms of causal inferences, Kunzli *et al* (2001) have indicated that following a defined group of study subjects over time allows the researcher to capture all deaths in the study group regardless of whether the effects of exposure are short or long term. However, if multiple outcomes are assessed (Crane *et al*, 2011; Harley *et al*, 2011), some might be significantly associated with the exposure by chance alone. In this instance, it might be best to consider the study as hypothesis-generating not hypothesis-testing, unless the outcomes were specified *a priori,* or a penalty is applied to the P-value that is used to declare a test as 'statistically significant'.

8.8 ANALYSIS

8.8.1 Risk-based cohort analysis

If the source population is closed, we can measure the average risk of disease and survival times during the follow-up period. Bivariable, risk-based analyses are shown in Chapter 6, and stratified analyses (to control confounding) in Chapter 13. Traditionally, multivariable models for risk-based outcomes have been built using logistic-regression models (Chapter 16), which use odds ratios as the base measure of association. See Section 18.4.1 for a discussion of methods for multivariable estimation of risk ratios (the natural measure of association for a risk-based cohort study). Cheung (2007) has described the use of linear regression if risk difference, rather than risk ratio, is the association measure of interest.

Both Cox (2006) and Greenland (2004) discuss the estimation of population attributable fractions (AF_p) in cohort studies. Cox suggests using a log-linear model approach for adjusted AF_p when the prevalence of exposure is known (*ie* a single cohort or longitudinal study

sampling) or estimable (available in some cohort study situations). Greenland demonstrates how to obtain a variety of association measures, including AF_p, using one or more of logistic, log-linear, and Poisson models (see Example 8.10).

8.8.2 Rate-based cohort analyses

If the source population is open, rates are used to measure disease frequency (see Examples 8.6 and 8.7) and if the rate is deemed to be constant over the follow-up period, the adjusted incidence rate ratios (*IR*s) can be obtained using Poisson models. Because of the long risk period for the outcome, most formal analyses of rate data in the medical literature have used survival models (see Examples 8.3 and 8.7). Callas *et al* (1998) compared a proportional hazard, Poisson and logistic model for the analysis of cohort data, and concluded that either one of the former 2 approaches would be preferable to logistic analysis. This has been confirmed more recently by others (Greenland, 2004). For multivariable analyses of grouped data, we can use a Poisson regression model (see Chapter 18) that includes the study-subject time-at-risk in each exposure category as the offset; the coefficients from this model provide direct estimates of the incidence rate ratio. As noted earlier, the incidence of disease is expressed relative to the time at risk in each level of exposure, not to the number of exposed (or non-exposed) individuals. If the time of disease (outcome) occurrence is of more interest than the fact of its occurrence, survival models are the method of choice (Case *et al*, 2002).

In the majority of rate-based cohort studies, a proportional hazards model is the basis of the analysis (Hernan, 2010; see Chapter 19). As Hernan points out, the hazard ratio (*HR*) can be thought of as equivalent to an *IR*. However, Hernan goes on to discuss what he perceives as 2 drawbacks to *HR*s; first, the average *HR* (which is usually what is reported) can change over time, and second, the period-specific *HR*s have a built in bias. The latter, Hernan argues, is due to the conditional probability which forms the basis of the *HR*. The *HR* at time '*t*' is conditional on not developing the outcome prior to time '*t*'. Thus, as the duration of follow-up increases, it becomes increasingly likely that the number of people who are susceptible to the exposure effects becomes smaller as the study period gets longer, so the risk in the exposed group decreases over time relative to that in the non-exposed group. If true, this suggests that the value of the average *HR* is dependent on the duration of follow-up. Hernan describes how to circumvent this problem through the use of covariate adjusted survival curves—see his paper for details. Earlier, Hernan *et al* (2008) proposed an analytic strategy for cohort studies that more closely approximates the analysis of a randomised trial using data on postmenopausal hormone therapy and coronary heart disease as their example. The main features of this approach are to subdivide the follow-up period into shorter time intervals, and treat each of these as a 'trial.' In each 'trial', subjects are censored when they change their baseline exposure. Then, the uncensored subject time is weighted by the inverse of their estimated probability of remaining uncensored until that point (these were obtained using a logistic model). They fit separate models for initiators (exposed) and non-initiators (non-exposed), and in each 'trial', each subject contributed as many observations to the model as the amount of time that subject was on the baseline therapy. Danaei *et al* (2011) elaborated and expanded on this approach.

Gran *et al* (2010) describe a sequential technique based on the Cox hazard model to analyse data with time-dependent confounders. The latter occurs when a covariate is affected by past exposure, and is a predictor of future exposure and outcome (*eg* the CD4 count when assessing the impact of treatment on HIV).

Laaksonen *et al* (2010) describe how to estimate the population attributable fraction (PAF) for the incidence of a disease during lifetime and a PAF for the prevalence of disease at a point in time. They propose a method to adjust for deaths during the follow-up period, rather than just censoring the data at the point of death.

Xue *et al* (2010) explain how to apply marginal and mixed-effect models to the analysis of cohort data when repeated measurements of exposure and covariates are used. The authors compare and contrast these approaches to logistic and proportional hazard models, with respect to their use and interpretation.

8.9 Reporting of cohort studies

As mentioned in Chapter 7, there has been a widespread initiative to improve the reporting of observational studies (STROBE; von Elm *et al* (2007)). We elaborated on these in this chapter as they should be used to help plan and report the study, as well as help you assess the validity of published cohort studies. See Table 8.1 for design aspects specific to cohort studies (Tooth *et al*, 2005).

We have noted an increased frequency of reporting on the design and implementation of proposed or early stage studies. We view this as a positive trend because it allows everyone to assess the potential strengths and weaknesses of a particular study without being biased by outcome data, as mentioned in Section 7.2. See Gern *et al* (2009); Hermsen *et al* (2011); Origasa *et al* (2011); Poulos *et al* (2011); Schuz *et al* (2011)as examples.

Table 8.1 Criteria used to assess published cohort studies

Criteria
1. Are the objectives or hypotheses of the study stated?
2. Is the target population defined?
3. Is the sampling frame defined?
4. Is the study population defined?
5. Are the study setting (venue) and/or geographic location stated?
6. Are the dates between which the study was conducted stated or implicit?
7. Are eligibility criteria stated?
8. Are issues of ‘selection in’ to the study mentioned?
9. Is the number of participants justified?
10. Are numbers meeting and not meeting the eligibility criteria stated?
11. For those not eligible, are the reasons why stated?
12. Are the numbers of people who did/did not consent to participate stated?
13. Are the reasons that people refused to consent stated?
14. Were responders compared with non-responders?
15. Was the number of participants at the beginning of the study stated?
16. Were methods of data collection stated?

17. Was the reliability (repeatability) of measurement methods mentioned?
18. Was the validity (against a 'gold standard') of measurement methods mentioned?
19. Were any confounders mentioned?
20. Was the number of participants at each stage specified?
21. Were reasons for loss to follow-up quantified?
22. Was the 'missingness' of data items at each wave mentioned?
23. Was the type of analysis conducted stated?
24. Were 'longitudinal' analysis methods stated?
25. Were absolute effect sizes reported?
26. Were relative effect sizes reported?
27. Was loss to follow-up taken into account in the analysis?
28. Were confounders accounted for in the analyses?
29. Were missing data accounted for in the analyses?
30. Was the impact of biases assessed qualitatively?
31. Was the impact of biases estimated quantitatively?
32. Did authors relate results back to a target population?
33. Was there any other discussion of 'generalisability'?

References

Barthel FMS, Babiker A, Royston P, Parmar MKB. Evaluation of sample size and power for multi-arm survival trials allowing for non-uniform accrual, non-proportional hazards, loss to follow-up and cross-over. Stat Med. 2006;25(15):2521-42.

Basagana X, Xiaomei L, Spiegelman D. Power and sample size calculations for longitudinal studies estimating a main effect of a time-varying exposure. Stat Meth Med Res. 2011;20(5):471-87.

Cai J, Zeng D. Power calculation for case-cohort studies with nonrare events. Biometrics. 2007;63(4):1288-95.

Callas PW, Pastides H, Hosmer DW. Empirical comparisons of proportional hazards, Poisson, and logistic regression modeling of occupational cohort data. Am J Ind Med. 1998 Jan;33(1):33-47.

Case LD, Kimmick G, Paskett ED, Lohman K, Tucker R. Interpreting measures of treatment effect in cancer clinical trials. The Oncologist. 2002;7(3):181-7.

Chang CC, Yang HC, Tang G, Ganguli M. Minimizing attrition bias: a longitudinal study of depressive symptoms in an elderly cohort. Int Psychogeriatrics / IPA. 2009;21(5):869-78.

Cheung YB. A modified least-squares regression approach to the estimation of risk difference. Am J Epidemiol. 2007;166(11):1337-44.

Choi M, Kim H, Qian H, Palepu A. Readmission rates of patients discharged against medical advice: a matched cohort study. PLoS ONE. 2011;6(9):e24459.

Cox C. Model-based estimation of the attributable risk in case-control and cohort studies. Stat Methods Med Res. 2006;15(6):611-25.

Crane JM, Keough M, Murphy P, Burrage L, Hutchens D. Effects of environmental tobacco smoke on perinatal outcomes: a retrospective cohort study. B J Obstet Gynaec. 2011;118(7):865-71.

Danaei G, Garcia Rodriguez LA, Cantero OF, Logan R, Hernan MA. Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. Stat Meth Med Res. 2011.

Davis FG, Al-Alem U. Allergies and adult gliomas: cohort results strengthen evidence for a causal association. J Nat Cancer Res. 2011;103(21):1562-3.

Duffy SW, Rohan TE, McLaughlin JR. Design and analysis considerations in a cohort study involving repeated measurement of both exposure and outcome: the association between genital papillomavirus infection and risk of cervical intraepithelial neoplasia. Stat Med. 1994;13(4):379-90.

Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. NephronClini Pract. 2009;113(3):c214-7.

Gern JE, Visness CM, Gergen PJ, Wood RA, Bloomberg GR, O'Connor GT, et al. The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. BMC pulmonary medicine. 2009;9:17.

Gran JM, Roysland K, Wolbers M, Didelez V, Sterne JA, Ledergerber B, et al. A sequential

Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. Stat Med. 2010;29(26):2757-68.

Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. Am J Epidemiol. 2004;160(4):301-5.

Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet. 2002;359(9303):341-5.

Harcombe H, Derrett S, Herbison P, McBride D. "Do I really want to do this?" Longitudinal cohort study participants' perspectives on postal survey design: a qualitative study. BMC Med Res Meth. 2011;11(1):8.

Harley KG, Chevrier J, Schall RA, Sjodin A, Bradman A, Eskenazi B. Association of prenatal exposure to polybrominated diphenyl ethers and infant birth weight. Am J Epidemiol. 2011;174(8):885-92.

Heinavaara S, Tokola K, Kurttio P, Auvinen A. Validation of exposure assessment and assessment of recruitment methods for a prospective cohort study of mobile phone users (COSMOS) in Finland: a pilot study. Environ Health. 2011;10:14.

Hermsen LA, Leone SS, van der Windt DA, Smalbrugge M, Dekker J, van der Horst HE. Functional outcome in older adults with joint pain and comorbidity: design of a prospective cohort study. BMC Musculoskel Disorders. 2011;12:241.

Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiol. 2008;19(6):766-79.

Hernan MA, McAdams M, McGrath N, Lanoy E, Costagliola D. Observation plans in longitudinal studies with time-varying treatments. Stat Meth Med Res. 2009;18(1):27-52.

Hernan MA. The hazards of hazard ratios. Epidemiol. 2010;21(1):13-5.

Hernan MA. Beyond exchangeability: The other conditions for causal inference in medical research. Stat Meth Med Res. 2012;21(1):3-5.

Hood MN. A review of cohort study design for cardiovascular nursing research. J Cardiovasc Nurs. 2009;24(6):E1-9.

Hudson JI, Pope HG, Jr., Glynn RJ. The cross-sectional cohort study: an underutilized design. Epidemiology. 2005;16(3):355-9.

Jonkman JN, Normand SL, Wolf R, Borbas C, Guadagnoli E. Identifying a cohort of patients with early-stage breast cancer: a comparison of hospital discharge and primary data. Medical care. 2001;39(10):1105-17.

Kelz RR, Tran TT, Hosokawa P, Henderson W, Paulson EC, Spitz F, et al. Time-of-day effects on surgical outcomes in the private sector: a retrospective cohort study. J Am Coll Surgeons. 2009;209(4):434-45.e2.

Klein-Geltink JE, Rochon PA, Dyer S, Laxer M, Anderson GM. Readers should systematically assess methods used to identify, measure and analyze confounding in observational cohort studies. J Clin Epidemiol. 2007;60(8):766-72.

Kunzli N, Medina S, Kaiser R, Quenel P, Horak F, Jr., Studnicka M. Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or on cohort studies? Am J Epidemiol. 2001;153(11):1050-5.

Laaksonen MA, Harkanen T, Knekt P, Virtala E, Oja H. Estimation of population attributable fraction (PAF) for disease occurrence in a cohort study design. Stat Med. 2010;29(7-8):860-74.

Latouche A, Porcher R, Chevret S. Sample size formula for proportional hazards modelling of competing risks. Stat Med. 2004;23(21):3263-74.

Latouche A, Porcher R. Sample size calculations in the presence of competing risks. Stat Med. 2007;26(30):5370-80.

Lazo M, Hernaez R, Bonekamp S, Kamel IR, Brancati FL, Guallar E, et al. Non-alcoholic fatty liver disease and mortality among US adults: prospective cohort study. BMJ (Clin Res). 2011;343:d6891.

Leece P, Kendall C, Touchie C, Pottie K, Angel JB, Jaffey J. Cervical cancer screening among HIV-positive women. Retrospective cohort study from a tertiary care HIV clinic. Can Fam Physician. 2010;56(12):e425-31.

Levin KA. Study design IV. Cohort studies. Evidence-based Dent. 2006;7(2):51-2.

Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, et al. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. BMJ (Clin Res). 2005;330(7497):960-2.

Matsui S. Sample size calculations for comparative clinical trials with over-dispersed Poisson process data. Stat Med. 2005;24(9):1339-56.

Matsui S, Zeng S, Yamanaka T, Shaughnessy J. Sample size calculations based on ranking and selection in microarray experiments. Biometrics. 2008;64(1):217-26.

Mazumdar M, Tu D, Zhou XK. Some design issues of strata-matched non-randomized studies with survival outcomes. Stat Med. 2006;25(23):3949-59.

McCartney G, Cowden J, Murray S, Ahmed S. The use of a new virtual cohort study design to investigate an outbreak of E. coli O157 linked to a supermarket delicatessen. Epidemiol and Inf. 2010;138(10):1439-42.

Mehta S, Chen H, Johnson ML, Aparasu RR. Risk of falls and fractures in older adults using antipsychotic agents: a propensity-matched retrospective cohort study. Drugs & Aging. 2010;27(10):815-29.

Normand SL, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. BMJ (Clin Res). 2005;330(7498):1021-3.

Origasa H, Goto S, Shimada K, Uchiyama S, Okada Y, Sugano K, et al. Prospective cohort study of gastrointestinal complications and vascular diseases in patients taking aspirin: rationale and design of the MAGIC Study. Cardiovasc Drugs and Therapy. 2011;25(6):551-60.

Oura T, Matsui S, Kawakami K. Sample size calculations for controlling the distribution of false discovery proportion in microarray experiments. Biostatistics. 2009;10(4):694-705.

Poulos RG, Hatfield J, Rissel C, Grzebieta R, McIntosh AS. Exposure-based cycling crash, near miss and injury rates: The Safer Cycling Prospective Cohort Study protocol. Injury Prevention. 2011.

Prentice RL. Design issues in cohort studies. Stat Meth Med Res. 1995;4(4):273-92.

Rochon PA, Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. BMJ (Clin Res). 2005;330(7496):895-7.

Rohan TE, Soskolne CL, Carroll KK, Kreiger N. The Canadian Study of Diet, Lifestyle, and Health: design and characteristics of a new cohort study of cancer risk. Cancer Detect Prevent. 2007;31(1):12-7.

Samet JM, Munoz A. Evolution of the cohort study. Epidemiol Rev. 1998a;20(1):1-14.

Samet JM, Munoz A. Perspective: cohort studies. Epidemiol Rev. 1998b;20(1):135-6.

Schutze M, Boeing H, Pischon T, Rehm J, Kehoe T, Gmel G, et al. Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. BMJ (Clin Res). 2011;342:d1584.

Schuz J, Elliott P, Auvinen A, Kromhout H, Poulsen AH, Johansen C, et al. An international prospective cohort study of mobile phone users and health (Cosmos): design considerations and enrolment. Cancer Epidemiol. 2011;35(1):37-43.

Soh SE, Saw SM. Cohort studies: design and pitfalls. Am J Ophthalmology. 2010;150(1):3-5.

Swerdlow AJ, Jones ME, Schoemaker MJ, Hemming J, Thomas D, Williamson J, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. Brit J Cancer. 2011;105(7):911-7.

Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. Am J Epidemiol. 2005;161(3):280-8.

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Epidemiol. 2007;18(6):800-4.

Warensjo E, Byberg L, Melhus H, Gedeborg R, Mallmin H, Wolk A, et al. Dietary calcium intake and risk of fracture and osteoporosis: prospective longitudinal cohort study. BMJ (Clin Res). 2011;342:d1473.

Wilcox A, Wacholder S. Observational data and clinical trials: narrowing the gap? Epidemiol. 2008;19(6):765.

Xue X, Gange SJ, Zhong Y, Burk RD, Minkoff H, Massad LS, et al. Marginal and mixed-effects models in the analysis of human papillomavirus natural history data. Cancer Epidemiol Biomarkers & Prev. 2010;19(1):159-69.