
Chapter 13

Interpretation of epidemiological studies

In epidemiology, studies are carried out to identify exposures that may affect the risk of developing a certain disease or other health-related outcome and to estimate quantitatively their effect. Unfortunately, errors are inevitable in almost any epidemiological study, even in the best conducted randomized trial. Thus, when interpreting findings from an epidemiological study, it is essential to consider how much the observed association between an exposure and an outcome may have been affected by errors in the design, conduct and analysis. Even if errors do not seem to be an obvious explanation for the observed effect, it is still necessary to assess the likelihood that the observed association is a causal one. The following questions should be addressed before it is concluded that the observed association between exposure and outcome is a true cause–effect relationship:

- (1) Could the observed association be due to systematic errors (*bias*) in the way subjects were selected and followed up or in the way information was obtained from them?
- (2) Could it be due to differences between the groups in the distribution of another variable (*confounder*) that was not measured or taken into account in the analyses?
- (3) Could it be due to *chance*?
- (4) Finally, is the observed association likely to be *causal*?

Most of these issues were already raised in Chapters 7–11 in relation to each specific study design. In this chapter, we will consider them in a more structured way.

13.1 Could the observed effect be due to bias?

Bias tends to lead to an incorrect estimate of the effect of an exposure on the development of a disease or other outcome of interest. The observed estimate may be either above or below the true value, depending on the nature of the error.

Many types of bias in epidemiology have been identified (Sackett, 1979) but, for simplicity, they can be grouped into two major types: *selection bias* and *measurement bias*.

13.1.1 Selection bias

Selection bias occurs when there is a difference between the characteristics of the people selected for the study and the characteristics of those

who were not. In all instances where selection bias occurs, the result is a difference in the relation between exposure and outcome between those who entered the study and those who would have been eligible but did not participate. For instance, selection bias will occur with volunteers (self-selection bias). People who volunteer to participate in a study tend to be different from the rest of the population in a number of demographic and lifestyle variables (usually being more health-conscious, better educated, etc.), some of which may also be risk factors for the outcome of interest.

Selection bias can be a major problem in case-control studies, although it can also affect cross-sectional studies and, to a lesser extent, cohort studies and randomized trials.

The selection of an appropriate sample for a *cross-sectional survey* does not necessarily guarantee that the participants are representative of the target population, because some of the selected subjects may fail to participate. This can introduce selection bias if non-participants differ from participants in relation to the factors under study.

In [Example 13.1](#), the prevalence of alcohol-related problems rose with increasing effort to recruit subjects, suggesting that those who completed the interview only after a large number of contact attempts were different from those who required less recruitment effort. Constraints of time and money usually limit the recruitment efforts to relatively few contact attempts. This may bias the prevalence estimates derived from a cross-sectional study.

In *case-control studies*, controls should represent the source population from which the cases were drawn, i.e. they should provide an estimate of the exposure prevalence in the general population from which the cases come. This is relatively straightforward to accomplish in a *nested case-control study*, in which the cases and the controls arise from a clearly defined population—the cohort. In a *population-based case-control study*, a source population can also be defined from which all cases (or a random sample) are obtained; controls will be randomly selected from the disease-free members of the same population.

The sampling method used to select the controls should ensure that they are a representative sample of the population from which the cases originated. If they are not, selection bias will be introduced. For instance, the method used to select controls in [Example 13.2](#) excluded women who were part of the study population but did not have a telephone. Thus, control selection bias might have been introduced if women with and without a telephone differed with respect to the exposure(s) of interest. This bias could be overcome by excluding cases who did not have a telephone, that is, by redefining the study population as women living in households with a telephone, aged 20–54 years, who resided in the eight selected areas. Moreover, the ultimate objective of the random-digit dialling method was not merely to provide a random sample of households with telephone but a random sample of *all* women aged 20–54 years living in these households during the study period. This depended on the extent to which people

Example 13.1. A cross-sectional study was conducted in St Louis, Missouri (USA) to assess the prevalence of psychiatric disorders. This study is quite unusual in that great efforts were made to recruit as many eligible subjects as possible. Subjects were chosen according to a five-stage probability sampling plan which gave each household a known probability of being selected. Once the household was selected, the residents were enumerated by age and sex, and one resident over age 18 years was randomly chosen to enter the study. Replacement was not allowed. Enumeration of residents was completed in 91% of the eligible households. Of the 3778 selected subjects, 3004 (80%) were successfully interviewed. Figure 13.1 shows that 32% of the respondents were interviewed after two contact attempts and 66% after five. Not until the 14th attempt were 95% of the interviews completed. The maximum number of attempts that resulted in an interview was 57. The mean for the 3004 responders was 5.3; the median, 4. Being young, male, black, a non-rural resident, well educated, and full-time employed were the demographic characteristics associated with increased contact efforts (Cottler et al., 1987). Table 13.1 shows prevalence estimates for alcohol-related problems by number of attempts made to obtain an interview.

Interview completed within this number of contact attempts	Cumulative sample (n)	Estimated prevalence of current alcohol abuse and dependence disorder (%)
5	1943	3.89
7	2285	3.98
8	2415	4.22
9	2511	4.26
57	2928 ^b	4.61

^a Data from Cottler *et al.* (1987).

^b This number is slightly lower than the number of subjects for whom the interview was completed (3004) because questionnaires with missing data were excluded from this analysis.

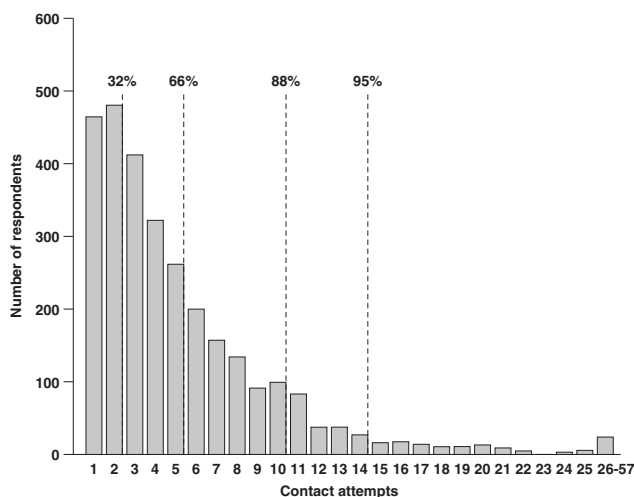


Table 13.1.

Prevalence estimates for current alcohol abuse and dependence disorder by number of contact attempts necessary to complete interview, St Louis Epidemiologic Catchment Area project, 1981–82.^a

Figure 13.1.

Number of contact attempts necessary to complete interview for the St Louis Epidemiologic Catchment Area project, 1981–82 (reproduced with permission from Cottler *et al.*, 1987).

Example 13.2. The effect of oral contraceptive use on the risk of breast, endometrial and ovarian cancers was investigated in the Cancer and Steroid Hormone Study. The study population was women aged 20–54 years who resided in eight selected areas in the USA during the study period. Attempts were made to identify all incident cases of breast, ovarian and endometrial cancer that occurred in the study population during the study period through local population-based cancer registries. Controls were selected by random-digit dialling of households in the eight locations. A random sample of household telephone numbers were called; information on the age and sex of all household members was requested and controls were selected among female members aged 20–54 years according to strict rules (Stadel et al., 1985).

answering the telephone numbers selected were willing to provide accurate information on the age and sex of all individuals living in the household.

Sometimes, it is not possible to define the population from which the cases arise. In these circumstances, *hospital-based controls* may be used, because the source population can then be re-defined as ‘hospital users’. Hospital controls may also be preferable for logistic reasons (easier and cheaper to identify and recruit) and because of lower potential for recall bias (see below). But selection bias may be introduced if admission to hospital for other conditions is related to exposure status.

Example 13.3. A classic case–control study was conducted in England in 1948–52 to examine the relationship between cigarette smoking and lung cancer. The cases were 1488 patients admitted for lung cancer to the participating hospitals, 70% of which were located in London. A similar number of controls were selected from patients who were admitted to the same hospitals with other conditions (except diseases thought at that time to be related to smoking). The smoking habits of the hospital controls were compared with those of a random sample of all residents in London (Table 13.2). These comparisons showed that, among men of similar ages, smoking was more common in the hospital controls than in the population sample (Doll & Hill, 1952).

Table 13.2. Age-adjusted distribution of male hospital controls without lung cancer and of a random sample of men from the general population from which the lung cancer cases originated according to their smoking habits.^a

Subjects	Percentage of subjects					Number (%) interviewed
	Non-smokers	Most recent number of cigarettes smoked per day				
		1–4	5–14	15–24	25+	
Hospital controls	7.0	4.2	43.3	32.1	13.4	1390 (100)
Sample of general population	12.1	7.0	44.2	28.1	8.5	199 (100)

^a Data from Doll & Hill (1952).

Table 13.2 shows that among men of similar age, the percentage of non-smokers was 7.0 in the hospital controls and 12.1 in the population sam-

ple. The percentage smoking at least 25 cigarettes per day was 13.4 in the hospital controls but only 8.5 in the population sample. The investigators stated that this difference in smoking habits between the hospital controls and the population random sample might be explained by previously unknown associations between smoking and several diseases. Thus, the strength of the association between lung cancer and cigarette smoking was *underestimated* in the case-control study, because the prevalence of smoking in the hospital controls was higher than in the general population from which the cases of lung cancer were drawn.

A hospital control series may fail to reflect the population at risk because it includes people admitted to the hospital for conditions caused (or prevented) by the exposures of interest. Individuals hospitalized for diseases related to the exposure under investigation should be excluded in order to eliminate this type of selection bias. However, this exclusion should not be extended to hospital patients with a history of exposure-related diseases, since no such restriction is imposed on the cases. Thus, patients admitted to hospital because of a smoking-related disorder (e.g., chronic bronchitis) should be excluded from the control series in a case-control study looking at the relationship between smoking and lung cancer, whereas those admitted for other conditions (e.g., accidents) but with a history of chronic bronchitis should be included.

Selection bias is less of a problem in *cohort studies*, because the enrolment of exposed and unexposed individuals is done before the development of any outcome of interest. This is also true in *historical cohort studies*, because the ascertainment of the exposure status was made some time in the past, before the outcome was known. However, bias may still be introduced in the selection of the 'unexposed' group. For instance, in occupational cohort studies where the general population is used as the comparison group, it is usual to find that the overall morbidity and mortality of the workers is lower than that of the general population. This is because only relatively healthy people are able to remain in employment, while the general population comprises a wider range of people including those who are too ill to be employed. This type of selection bias is called

Example 13.4. Suppose that a particular cohort was defined as workers who were in active employment at a particular point in time and that no new members were allowed to join the cohort later. If all the cohort members are followed up, including those who leave or retire for health reasons, the healthy worker effect will decline with time, because the initial healthy selection process at the time of recruitment into the workforce will become weaker with the passage of time. Note that employed people are not only generally healthier but also, if unhealthy, less likely to suffer from long-term and easy-to-detect diseases (e.g., cardiovascular conditions) than from diseases which remain asymptomatic for long periods (e.g., many cancers) (McMichael, 1976).

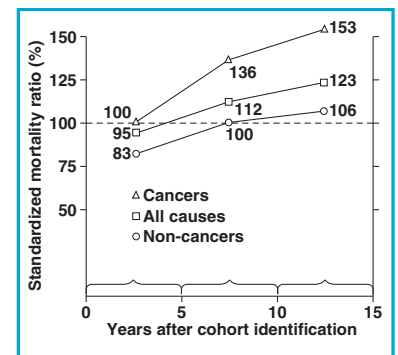


Figure 13.2.

Decline in the healthy worker effect with passage of time after initial identification of a cohort of active workers. Graph based on mortality data among a cohort of asbestos product workers (Enterline, 1965) (reproduced with permission from McMichael, 1976). © American College of Occupational and Environmental Medicine, 1976).

the *healthy worker effect*. It may be minimized by restricting the analysis to people from the same factory who went through the same healthy selection process but have a different job (see Section 8.2.2).

Example 13.4 shows that the healthy selection bias varies with type of disease, being less marked for cancer than for non-cancer conditions (mainly cardiovascular disorders), and it tends to decline with time since recruitment into the workforce.

Incompleteness of follow-up due to non-response, refusal to participate and withdrawals may also be a major source of selection bias in cohort studies in which people have to be followed up for long periods of time. However, this will introduce bias only if the degree of incompleteness is different for different exposure categories. For example, subjects may be more inclined to return for a follow-up examination if they have developed symptoms of the disease. This tendency may be different in the exposed and unexposed, resulting in an over- or under-estimation of the effect. Definitions of individual follow-up periods may conceal this source of bias. For example, if subjects with a certain occupation tend to leave their job when they develop symptoms of disease, exposed cases may not be identified if follow-up terminates at the time subjects change to another job. A similar selection bias may occur in migrant studies if people who become ill return to their countries of origin before their condition is properly diagnosed in the host country.

Randomized intervention trials are less likely to be affected by selection bias since subjects are randomized to the exposure groups to be compared. However, refusals to participate after randomization and withdrawals from the study may affect the results if their occurrence is related to exposure status. To minimize selection bias, allocation to the various study groups should be conducted only after having assessed that subjects are both eligible and willing to participate (see Section 7.9). The data should also be analysed according to ‘intention to treat’ regardless of whether or not the subjects complied with their allocated intervention (see Section 7.12).

13.1.2 Measurement bias

Measurement (or information) bias occurs when measurements or classifications of disease or exposure are not valid (i.e., when they do not measure correctly what they are supposed to measure). Errors in measurement may be introduced by the observer (*observer bias*), by the study individual (*responder bias*), or by the instruments (e.g., questionnaire or sphygmomanometer or laboratory assays) used to make the measurements (see [Chapter 2](#)).

Misclassification of a single attribute (exposure or outcome) and observed prevalence

Suppose that a cross-sectional survey was conducted to assess the prevalence of a particular attribute in a certain study population. A questionnaire was administered to all eligible participants; there were no refusals. Let us denote the observed proportion in the population classified as having the

attribute by p^* and the true prevalence by p . p^* has two components. One component comes from those individuals with the attribute that are correctly classified by the questionnaire as having it (true positives). The other component comes from those individuals that actually do not have the attribute but erroneously have been classified (that is, misclassified) as having it (false positives). The proportion of individuals that is classified by the questionnaire as having the attribute (p^*) is then:

$$p^* = p \times \text{sensitivity} + (1 - p) \times (1 - \text{specificity})$$

and thus depends on the true prevalence of the attribute (p) and on the sensitivity and specificity of the measurement method. For example, if $p = 0.1\%$, sensitivity = 90% and specificity = 90%, then $p^* = 10.1\%$. This means that if the prevalence were estimated by the proportion that is classified as having the attribute by the questionnaire, the estimate would be 10.1%, compared with a true prevalence of only 0.1%. This misclassification corresponds to a 100-fold overestimation.

It is possible to obtain a corrected estimate of the true prevalence in situations where the sensitivity and the specificity of the measurement procedure are both known, or may be estimated, by re-arranging the above formula as follows:

$$p = \frac{p^* + \text{specificity} - 1}{\text{sensitivity} + \text{specificity} - 1}$$

True prevalence (%)	Specificity (%)	Sensitivity (%)		
		99	90	80
0.1	99	1.1	1.1	1.1
	90	10.1	10.1	10.1
	80	20.1	20.1	20.1
0.5	99	1.5	1.4	1.4
	90	10.4	10.4	10.4
	80	20.4	20.4	20.3
2.5	99	3.5	3.2	3.0
	90	12.2	12.0	11.8
	80	22.0	21.8	21.5
12.5	99	13.3	12.1	10.9
	90	21.1	20.0	18.8
	80	29.9	28.8	27.5
50.0	99	50.0	45.5	40.5
	90	54.5	50.0	45.0
	80	59.5	55.0	50.0
62.5	99	62.3	56.6	50.4
	90	65.6	60.0	53.8
	80	69.4	63.8	57.5

Table 13.3.

Observed prevalence (p^*) (%) of an attribute in a population for different levels of the true prevalence (p), and different values of sensitivity and specificity of the measurement procedure.

Table 13.3 shows the effects of different levels of sensitivity and specificity of a measurement procedure on the observed prevalence of a particular attribute. In most cases, the observed prevalence is a gross overestimation of the true prevalence. For instance, the observed prevalence at 90% sensitivity and 90% specificity assuming a true prevalence of 0.1% is only about one half of that which would be obtainable if the true prevalence were 12.5% (i.e., 10% versus 20%). The bias in overestimation is severely influenced by losses in specificity (particularly when the true prevalence is less than 50%). In contrast, losses in sensitivity have, at most, moderate effects on the observed prevalence.

Misclassification and exposure–outcome relationships

Two main types of misclassification may affect the interpretation of exposure–outcome relationships: nondifferential and differential. *Nondifferential misclassification* occurs when an exposure or outcome classification is incorrect for equal proportions of subjects in the compared groups. In other words, nondifferential misclassification refers to errors in classification of outcome that are unrelated to exposure status, or misclassification of exposure unrelated to the individual's outcome status. In these circumstances, all individuals (regardless of their exposure/outcome status) have the same probability of being misclassified. In contrast, *differential misclassification* occurs when errors in classification of outcome status are dependent upon exposure status or when errors in classification of exposure depend on outcome status.

These two types of misclassification affect results from epidemiological studies in different ways. Their implications also depend on whether the misclassification relates to the exposure or the outcome status.

Nondifferential exposure misclassification

Nondifferential exposure misclassification occurs when all individuals (regardless of their current or future outcome status) have the same probability of being misclassified in relation to their exposure status. Usually, this type of misclassification gives rise to an underestimation of the strength of the association between exposure and outcome, that is, it '*dilutes*' the effect of the exposure.

In the case–control study illustrated in Example 13.5, nondifferential exposure misclassification introduced a bias towards an underestimation of the true exposure effect.

Nondifferential misclassification of exposure will also affect the results from studies of other types. For instance, historical occupational cohort studies rely upon records of exposure of individuals from the past. However, there often were no specific environmental exposure measurements that would allow accurate classification of individuals. This means that individuals were classified as 'exposed' or 'unexposed' by the job they did or by membership of a union. These proxy variables are very crude markers of the true exposure levels and their validity is limited. It is, however, unlikely that the validity of

Example 13.5. A case-control study was conducted to assess whether coffee intake increased the risk of pancreatic cancer. Subjects were classified as “ever” or “never” drinkers. The true results of this study in the absence of exposure misclassification (i.e., exposure measurement with sensitivity = 100%; specificity = 100%) are shown in Table 13.4.

	Coffee intake		Total
	Ever	Never	
Pancreatic cancer cases	200	100	300
Controls	150	150	300
Total	350	250	600
True odds ratio = $(200/100) / (150/150) = 2.00$.			

Suppose that the information on coffee intake was obtained through a self-administered questionnaire and that only 80% of study subjects who usually drank coffee (regardless of whether or not they had pancreatic cancer) reported this in the questionnaire (sensitivity = 80%). Similarly, only 90% of those who never drank coffee correctly mentioned this in the questionnaire (specificity = 90%).

	Coffee intake		Total
	Ever	Never	
Pancreatic cancer cases	$(200 \times 0.8) + (100 \times 0.1) = 170$	$(100 \times 0.9) + (200 \times 0.2) = 130$	300
Controls	$(150 \times 0.8) + (150 \times 0.1) = 135$	$(150 \times 0.9) + (150 \times 0.2) = 165$	300
Total	305	295	600
Observed odds ratio = $(170/130) / (135/165) = 1.60$			

Table 13.4.

Results from a hypothetical case-control study: true exposure status (sensitivity = 100% and specificity = 100% for both cases and controls).

Table 13.5.

Bias due to nondifferential exposure misclassification: observed exposure status (sensitivity = 80% and specificity = 90% for both cases and controls).

these records would be different for those who later developed the outcome of interest and those who did not. As long as the misclassification is nondifferential, it will generally dilute any true association between the exposure and the outcome.

The implications of nondifferential exposure misclassification depend heavily on whether the study is perceived as ‘positive’ or ‘negative’. This bias is a greater concern in interpreting studies that seem to indicate the absence of an effect. In these circumstances, it is crucial that the researchers consider the problem of nondifferential exposure misclassification in order to determine the probability that a real effect was missed. On the other hand, it is incorrect to dismiss a study reporting an effect simply because there is substantial nondifferential misclassification, since an estimate of effect without the misclassification would generally be even greater.

It should be noted that nondifferential exposure misclassification leads to an underestimation of the relative risk (bias towards relative effect of 1) only if the exposure is classified as a binary variable (e.g., ‘exposed’ versus ‘unex-

posed’). When there are three or more exposure categories, nondifferential exposure misclassification can result in either over- or under-estimation of the effect (Flegal *et al.*, 1986). For example, in a study of the effect of different levels of coffee intake (classified as ‘never’, ‘low’ and ‘high’) on the risk of pancreatic cancer, the relative risk associated with low levels of coffee intake will be overestimated if there is a general tendency for subjects with a true high intake to underreport it (but not so extreme as to report themselves as non-drinkers). On the other hand, if some people with a true high intake are classified as never-drinkers, the relative risk associated with low levels of coffee intake would be underestimated or even reversed (as in [Example 13.6](#)).

Example 13.6. Table 13.6 shows results from a hypothetical case–control study with a positive exposure–response trend. Suppose that both cases and controls were classified correctly in relation to exposure, except that 60% of subjects who were truly in the ‘none’ exposure group were misclassified into the ‘high’ exposure group, and 60% of those truly in the ‘high’ group were misclassified into the ‘none’ exposure group. While the odds ratios in the original data were 2.00 and 8.00 for the low and high exposure categories, respectively, they were 0.90 and 0.73 in the misclassified data. This misclassification led to the creation of an inverse exposure–response trend.

Table 13.6. Nondifferential exposure misclassification involving more than two exposure categories.^a

Disease status	True exposure status		
	None ^b	Low	High
Cases	53	40	60
Controls	424	160	60
True odds ratio	1.00	2.00	8.00
Misclassified exposure status (60% misclassification)			
	None ^b	Low	High
Cases	53–(53×0.6)+(60×0.6)=57	40	60–(60×0.6)+(53×0.6)=56
Controls	424–(424×0.6)+(60×0.6)=206	160	60–(60×0.6)+(424×0.6)=278
Observed odds ratio	1.00	0.90	0.73

^a Data from Dosemeci *et al.* (1990).
^b Taken as the baseline category

In this discussion, we have assumed that the misclassification occurs between ‘exposure’ and ‘no exposure’ or between different levels of exposure. Obviously, one exposure may be misclassified as another. For example, when studying the effect of oral contraceptive pills on the risk of breast cancer, women may confuse them with other drugs that they might have taken in the past.

Differential exposure misclassification

This type of misclassification can *bias the estimates of the association in either direction* and, hence, it can be responsible for associations which prove to be spurious.

Example 13.7. Consider the example of a case-control study on oral contraceptive use and breast cancer. The results of this study in the absence of misclassification are shown in Table 13.7.

	Oral contraceptive use		Total
	Ever	Never	
Breast cancer cases	150	200	350
Controls	150	200	350
Total	300	400	700
True odds ratio = $(150/200) / (150/200) = 1.00$			

Suppose now that 20% of the cases who had never used oral contraceptives incorrectly reported having done so, whereas all case users correctly reported their habits (sensitivity = 100%; specificity = 80%). The controls correctly reported their use (sensitivity = 100%; specificity = 100%).

	Oral contraceptive use		Total
	Ever	Never	
Breast cancer cases	$150 + (200 \times 0.20) = 190$	$200 - (200 \times 0.20) = 160$	350
Controls	150	200	350
Total	340	360	700
Observed odds ratio = $(190/160) / (150/200) = 1.58$			

Table 13.7.

Data from a hypothetical case-control study: true exposure status (sensitivity = 100% and specificity = 100% for both cases and controls).

Table 13.8.

Bias due to differential exposure misclassification: observed exposure status (sensitivity = 100% and specificity = 80% among cases; sensitivity = 100% and specificity = 100% among controls).

In [Example 13.7](#), patients with breast cancer were more likely to incorrectly report having ever used oral contraceptives than healthy controls, resulting in a spurious association between oral contraceptives and breast cancer. This is a particular type of differential misclassification called *recall bias*. One way of minimizing recall bias is to use hospital controls, because then the controls would generally have the same incentive as the cases to remember events in the past. Subjects should also be unaware of the specific study hypothesis.

In [Example 13.8](#), there is no biological evidence to suggest that the effect of vasectomy would be different between Catholic and Protestant men. Thus, the most likely explanation for the observed findings is that Catholic controls were less likely to report that they had had a vasectomy than Catholic men with testicular cancer. This differential exposure misclassification biased the exposure effect upwards among Catholic men.

Sometimes, it is possible to obtain direct evidence of the presence and magnitude of differential misclassification. In [Example 13.9](#), cases, but not controls, considerably overreported their inability to tan after being diagnosed with skin melanoma: the proportion of cases reporting inability to tan was 26% (=9/34) in 1982 before skin cancer was diagnosed but 44% (=15/34) after the diagnosis.

Example 13.8. A case-control study was conducted in western Washington state (USA) to assess the effect of vasectomy (surgical sterilization) on the risk of testicular cancer. Exposure information was obtained by telephone interview. The results showed a 50% excess risk associated with vasectomy: odds ratio (OR) = 1.5 (95% CI: 1.0–2.2). However, further analyses showed that the effect of vasectomy was considerably different for Catholic and Protestant men: OR = 8.7 for Catholic, and OR = 1.0 for Protestant background. Whereas a history of vasectomy was reported with approximately equal frequency by Catholic and non-Catholic cases, only 6.2% of Catholic controls reported such a history in contrast to 19.7% of other controls (Strader et al., 1988).

Table 13.9. History of vasectomy in testicular cancer cases and controls by religious background.^a

Religious background	Cases		Controls		Age-adjusted odds ratio (95% CI) ^c
	No.	% with vasectomy ^b	No.	% with vasectomy	
Protestant	129	18.0	295	19.0	1.0 (0.6–1.7)
Catholic	42	24.4	96	6.2	8.7 (2.8–27.1)
Other	57	32.3	122	21.0	1.3 (0.6–3.0)

^a Data from Strader *et al.* (1988).
^b Adjusted to age distribution of all controls.
^c CI = confidence interval.

Example 13.9. The Nurses' Health Study is an on-going cohort study of 121 700 female nurses who were aged 30–55 years when the cohort was assembled in 1976. These women have been sent postal questionnaires biennially since then. The 1982 questionnaire included questions on risk factors for skin melanoma. A nested case-control study was conducted in 1984, which included 34 skin melanoma cases diagnosed after the return of the 1982 questionnaire and 234 controls randomly selected from cohort members without a history of cancer who responded to the same mailed questionnaire. Two questions from the 1982 questionnaire were asked again in the case-control questionnaire. These related to hair colour and ability to tan (Weinstock et al., 1991). The responses given in the two questionnaires are shown in Table 13.10.

Table 13.10. Self-reported hair colour and tanning ability before and after the diagnosis of skin melanoma in a nested case-control study.^a

	1982 questionnaire		Case-control questionnaire	
	Cases	Controls	Cases	Controls
Hair colour				
Red or blond	11	37	11	41
Brown or black ^b	23	197	23	193
	OR=2.5 (95% CI=1.1–5.7)		OR=2.3 (95% CI=1.0–5.0)	
Tanning ability				
No tan or light tan	9	79	15	77
Medium, deep or dark ^b	25	155	19	157
	OR=0.7 (95% CI=0.3–1.5)		OR=1.6 (95% CI=0.8–3.5)	

^a Adapted from Weinstock *et al.* (1991).
^b Baseline category.

Similarly, observers who know the outcome status of an individual may be consciously or unconsciously predisposed to assess exposure variables according to the hypothesis under study. This type of bias is known as *observer bias*. One way of minimizing this type of bias is to keep the observers *blind* to the outcome status of the study subjects. For many diseases/conditions this is clearly impossible.

Example 13.10. Consider a hypothetical trial in which 20 000 men were randomly allocated to receive a new intervention or placebo. They were then followed up for five years to assess whether the intervention could prevent lung cancer. There were no losses to follow-up. A total of 75 lung cancer cases occurred during the follow-up period. With no outcome misclassification (i.e., outcome measurement method with sensitivity = 100% and specificity = 100%), the results would be as shown in Table 13.11.

	Intervention		Total
	New	Placebo	
Cases	25	50	75
Non-cases	9 975	9 950	19 925
Total	10 000	10 000	20 000
True risk in the new intervention group (r_1) = 25/10 000 = 25 per 10 000			
True risk in the placebo group (r_0) = 50/10 000 = 50 per 10 000			
True risk ratio = r_1/r_0 = 0.50			
True risk difference = $r_0 - r_1$ = 25 per 10 000			
True prevented fraction = $(r_0 - r_1)/r_0$ = 25 per 10 000 / 50 per 10 000 = 0.50 = 50%			

Suppose that only 80% of the cases of lung cancer were correctly identified (sensitivity = 0.80) among both those who received and those who did not receive the new intervention. This would give the results presented in Table 13.12.

	Intervention		Total
	New	Placebo	
Cases	$25 \times 0.80 = 20$	$50 \times 0.80 = 40$	60
Non-cases	$9\,975 + (25 \times 0.20) = 9\,980$	$9\,950 + (50 \times 0.20) = 9\,960$	19 940
Total	10 000	10 000	20 000
Observed r_1 = 20 per 10 000			
Observed r_0 = 40 per 10 000			
Observed risk ratio = r_1/r_0 = 0.50			
Observed risk difference = $r_0 - r_1$ = 20 per 10 000			
Observed prevented fraction = $(r_0 - r_1)/r_0$ = 20 per 10 000 / 40 per 10 000 = 0.50 = 50%			

Table 13.11.

Results from a hypothetical intervention trial: true outcome status (sensitivity = 100% and specificity = 100% for both the new intervention and placebo groups).

Table 13.12.

Bias due to nondifferential outcome misclassification: observed outcome status (sensitivity = 80% and specificity = 100% for both the new intervention and placebo groups).

Nondifferential outcome misclassification

Misclassification of the outcome is nondifferential if it is similar for those exposed and those unexposed. This type of misclassification either does not affect the estimates of relative effect (if specificity = 100% and sensitivity < 100%), or it introduces a bias towards a relative effect of 1 (if specificity < 100% and sensitivity ≤ 100%).

When specificity = 100%, as in [Example 13.10](#), a lack of sensitivity does not result in a bias in the estimate of the risk ratio or prevented fraction. The estimate of the risk difference is, however, biased.

Example 13.11. Consider again the hypothetical trial described in [Table 13.11](#). Suppose now that all cases of lung cancer were correctly identified (sensitivity = 100%), but that 10% of the subjects without lung cancer in each study group were misclassified as being lung cancer cases (specificity = 90%). The results would now be as shown in [Table 13.13](#).

Table 13.13.
Bias due to nondifferential outcome misclassification: observed outcome status (sensitivity = 100%; specificity = 90% for both the new intervention and placebo groups).

	Intervention		Total
	New	Placebo	
Cases	25 + (9 975×0.10) = 1 022	50 + (9 950×0.10) = 1 045	2 067
Non-cases	9 975 × 0.90 = 8 978	9 950 × 0.90 = 8 955	17 933
Total	10 000	10 000	20 000
Observed r_1 = 1022 per 10 000			
Observed r_0 = 1045 per 10 000			
Observed risk ratio = r_1 / r_0 = 0.98			
Observed risk difference = $r_0 - r_1$ = 23 per 10 000			
Observed prevented fraction = $(r_0 - r_1) / r_0$ = 23 per 10 000 / 1045 per 10 000 = 0.022 = 2.2%			

In [Example 13.11](#), a decline in specificity resulted in a risk ratio close to unity and in marked underestimation of the prevented fraction.

Differential outcome misclassification

Outcome misclassification is *differential* if the misclassification differs between the exposed and unexposed subjects. Differential disease misclassification introduces a bias towards an under- or over-estimation of the true exposure effect.

Differential disease misclassification is introduced if the exposure influences the follow-up and identification of cases. For example, exposed subjects may be more (or less) likely than unexposed subjects to report symptoms of disease or to visit a doctor. Similarly, the staff involved in the follow-up and diagnosis of disease may be influenced by awareness of the exposure status of subjects (*observer bias*). One way of minimizing this type of bias is to keep the observers blind to the exposure status of the individuals. But for certain exposures (e.g. surgical procedures, interventions with typical secondary effects), this is clearly not feasible.

Example 13.12. *A historical cohort study was conducted in England and Wales to examine the relationship between occupational exposure to asbestos and mortality from lung cancer. Information on cause of death was obtained from death certificates supplemented by autopsy reports. A total of 11 lung cancer deaths were observed among asbestos workers compared with only 0.8 expected on the basis of the mortality of men in the general population of England and Wales (SMR (O/E) = 14). However, autopsies among asbestos workers were likely to have been much more frequent than in the general population since asbestosis is an occupational disease for which people are entitled to compensation (Doll, 1955).*

In [Example 13.12](#), the risk of lung cancer among asbestos workers may have been slightly overestimated because the cause of death is likely to have been ascertained more carefully among asbestos workers than among the general population.

13.1.3 How can we identify bias in epidemiological studies?

Bias is a consequence of defects in the design or execution of an epidemiological study. Bias cannot be controlled for in the analysis of a study, and it cannot be eliminated by increasing the sample size, except for those introduced by non-differential misclassification of exposure or outcome. Ways of minimizing bias in different types of epidemiological study were discussed in previous chapters (see Chapters 7–11). [Box 13.1](#) highlights some of the questions that help to identify bias in epidemiological studies.

In addition to identifying potential sources of bias in a particular study, we also need to estimate their most likely direction and magnitude. Some procedures can be introduced deliberately into the study to assess the effect of a potential bias. For instance, in a mortality study, the vital status of people who were lost to follow-up may be ascertained from routine vital statistics registries and their mortality compared with that of the people who did participate in the study. Again, it is essential that the same sort of procedures will be applied to any subject irrespective of his/her exposure or disease status.

13.2 Could the observed effect be due to confounding?

Confounding occurs when an estimate of the association between an exposure and an outcome is mixed up with the real effect of another exposure on the same outcome, the two exposures being correlated.

For example, tobacco smoking may confound estimates of the association between work in a particular occupation and the risk of lung cancer. If death rates for lung cancer in the occupational group are compared with those for the general population (of a similar sex and age composition), it might appear that the occupational group has an

Box 13.1. How to check for bias in epidemiological studies

- Selection bias
 - Was the study population clearly defined?
 - What were the inclusion and exclusion criteria?
 - Were refusals, losses to follow-up, etc. kept to a minimum?

In cohort and intervention studies:

 - Were the groups similar except for the exposure/intervention status?
 - Was the follow-up adequate? Was it similar for all groups?

In case-control studies:

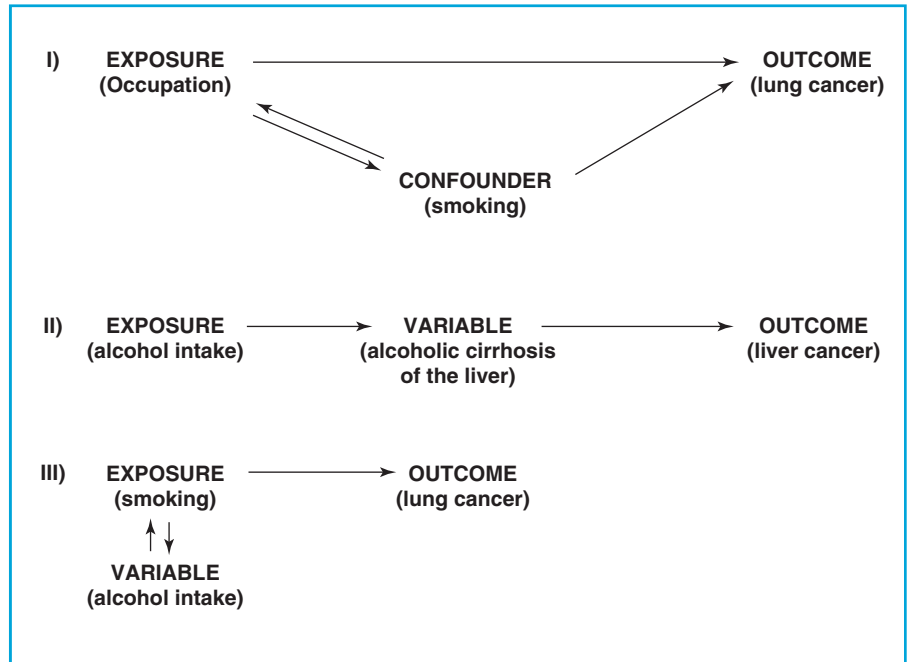
 - Did the controls represent the population from which the cases arose?
 - Was the identification and selection of cases and controls influenced by their exposure status?
- Measurement bias
 - Were the exposures and outcomes of interest clearly defined using standard criteria?
 - Were the measurements as objective as possible?
 - Were the subjects and observers blind?
 - Were the observers and interviewers rigorously trained?
 - Were clearly written protocols used to standardize procedures in data collection?
 - Were the study subjects randomized to observers or interviewers?
 - Was information provided by the patient validated against any existing records?
 - Were the methods used for measuring the exposure(s) and outcome(s) of interest (e.g.; questionnaires, laboratory assays) validated?
- Were strategies built into the study design to allow assessment of the likely direction and magnitude of the bias?

increased risk of lung cancer. This might lead to the inference that the occupation is a direct cause of lung cancer. However, without further analysis, this inference would be invalid if those people employed in the occupation smoked more heavily than members of the general population.

For a variable to be a *confounder*, it must be associated with the exposure under study and it must also be an independent risk factor for the disease. In [Figure 13.3](#), confounding occurs only in example I—smoking is associated with the particular occupation under study and it is on its own a risk factor for lung cancer. In example II, alcoholic cirrhosis of the liver is an intermediate factor in the causal path between the exposure (alcohol intake) and the disease (liver cancer). In example III, alcohol intake is associated with the exposure under study (smoking) but is not a risk factor for the disease (lung cancer).

A potential confounder is any factor which is believed to have a real effect on the risk of the disease under investigation. This includes both factors that have a direct causal link with the disease (e.g., smoking and lung cancer), and factors that are good proxy measures of more direct unknown causes (e.g., age and social class).

Confounding can be dealt with at the study design level or, provided the relevant data have been collected, in the analysis. Three approaches may be used to control for confounding in the design of an epidemiological study:



Randomization—this is the ideal method of controlling for confounders because it ensures that the distribution of known and, more importantly, unknown confounding variables will be similar in the groups to be compared, provided that the sample size is relatively large. But this method can be used only in experimental studies.

Restriction—a procedure that limits participation in the study to people who are similar in relation to the confounder. For instance, if participation in a study is restricted to non-smokers, any potential confounding effect of smoking will be removed.

Matching—a procedure whereby controls are selected in such a way that the distribution of potential confounders (e.g., age, sex, race or place of residence) among them will be identical to those of the cases. This can be accomplished by selecting for each case one or more controls with similar characteristics (e.g., of the same age, sex) (individual matching) or by ensuring that as a group the controls have similar characteristics to the cases (frequency matching). In practice, matching is mainly used in case-control studies (see Section 9.3.4.), because it would be too costly to match subjects in large cohort studies and unnecessary to do so in large intervention studies. However, in small intervention studies, particularly in community trials, matching is frequently used (see Section 7.9.2.)

Figure 13.3. Diagram illustrating the relationship between exposure, outcome and potential confounding variables.

Confounding can also be controlled for in the analysis by using:

Stratification—a technique in which the strength of the association is measured separately within each well defined and homogeneous category (*stratum*) of the confounding variable. For instance, if age is a confounder, the association is estimated separately in each age-group; the results can then be pooled using a suitable weighting to obtain an overall summary measure of the association, which is *adjusted* or *controlled* for the effects of the confounder (that is, that takes into account differences between the groups in the distribution of confounders; see Chapter 14). It should be noted that standardization, a technique mentioned in Chapter 4, is an example of stratification.

Statistical modelling—sophisticated statistical methods, such as regression modelling, are available to control for confounding. They are particularly useful when it is necessary to adjust simultaneously for various confounders. These techniques are briefly introduced in Chapter 14.

It is possible to control for confounders in the analysis only if data on them were collected; the extent to which confounding can be controlled for will depend on the accuracy of these data. For instance, nondifferential (random) misclassification of exposure to a confounder will lead to underestimation of the effect of the confounder and consequently, will attenuate the degree to which confounding can be controlled. The association will persist even after the adjustment because of *residual confounding*. But in contrast to *non-differential misclassification of exposure or outcome*, *non-differential misclassification of a confounder will cause a bias in either direction*, depending on the direction of the confounding. For example, in a study of risk factors for cervical cancer, the association between cigarette smoking and cervical cancer persisted even after controlling in the analysis for the number of sexual partners. This might be due, at least in part, to residual confounding because ‘number of sexual partners’ is an inaccurate measure of sexual behaviour, including infection by human papillomavirus.

13.3 Could the observed effect be due to chance?

The assessment of the role of chance in the interpretation of results from epidemiological studies was discussed in Chapter 6. In summary, the role of chance can be assessed by performing appropriate statistical significance tests and by calculating confidence intervals.

A statistical significance test for the effect of an exposure on an outcome yields the probability (P or P -value) that a result as extreme as or more extreme than the one observed could have occurred by chance alone, i.e., if there were no true relationship between the exposure and the outcome.

If this probability is very small, it is usual to declare that chance is an unlikely explanation for the observed association and consequently, that there is a '*statistically significant*' association between exposure and outcome. If the P -value is large (usually greater than 0.05), it is conventionally thought that chance cannot be excluded as an explanation for the observed association.

A number of issues should be kept in mind when interpreting results from statistical tests. Firstly, the value of P depends both on the magnitude of the association and on the study size. It is therefore possible with small studies to obtain a P -value which is not statistically significant, despite the fact that the true population exposure effect is large. Conversely, with a large sample, small effects, which may be clinically and epidemiologically irrelevant, may easily achieve statistical significance. Confidence intervals are more informative than P -values because they provide a range of values for the exposure–outcome association, which is likely to include the true population effect (usually with a probability of 0.95, i.e. 95%). They also indicate whether a non-significant result is or is not compatible with a true effect that was not detected because the sample size was too small (see Chapter 6). Sample size calculations should always be done at the design stage to ensure that the study will have enough power (i.e., its sample size will be large enough) to detect the hypothetical effect or enough precision to quantify it (see Chapter 15).

Secondly, investigators tend to collect data on many potential risk factors and to perform multiple statistical tests to determine whether any of these variables is significantly associated with the disease or outcome of interest. However, the probability of obtaining a statistically significant P -value just by chance increases with the number of tests performed. It is particularly important to keep this in mind when the data are interpreted, particularly if an unexpected association is found that was not hypothesized at the beginning of the study. Such an unexpected association would have to be reproduced in studies specifically designed to test it before it could be accepted as real.

Finally, statistical methods assess only the effects of sampling variation and cannot control for non-sampling errors such as confounding or bias in the design, conduct or analysis of a study.

13.4 Could the observed effect be causal?

If bias, confounding and chance do not seem to explain an observed association, can we conclude that the association is likely to be causal? Since we can never be sure that the study was not affected by any of these factors, we need to consider the findings in the context of all available evidence. In a paper published in 1965, Bradford Hill mentioned aspects that need to be considered when assessing whether an association is likely to be causal (Hill, 1965). We will consider each of these aspects below. As an illustration, we show how these aspects were applied to evaluate the association between smoking and lung cancer.

Temporal relationship

This is an essential aspect. For an exposure to be the cause of a disease, it has to precede its biological onset. This is generally easier to establish from intervention or cohort studies than from cross-sectional or case-control studies in which measurements of the exposure and the outcome are made at the same time.

“The criterion of temporality requires that cigarette smoking antedate the onset of cancer. Support for this criterion is provided by all the major prospective [cohort] studies in which an enormous number of initially disease-free subjects were followed over varying time intervals.” (US Surgeon General, 1982)

Biological plausibility

The association is more likely to be causal if consistent with other biological knowledge (e.g., animal experiments, biological mechanisms). However, this aspect should not be taken too seriously, because lack of plausibility may simply reflect lack of scientific knowledge or the fact that human beings are biologically different from animals.

“Benign and malignant tumours have been induced in the larynx of hamsters by long-term exposure to diluted cigarette smoke. Attempts to induce significant numbers of bronchogenic carcinoma [lung cancer] in laboratory animals were negative in spite of major efforts with several species and strains. Neither rats nor hamsters nor baboons inhale cigarette smoke as deeply and as intensely as the cigarette smokers who have provided the data with the consequences of their ‘experiment’ in the form of clinical evidence gathered by epidemiologists.” (US Surgeon General, 1982)

Consistency

If similar results have been found in different populations using different study designs, the association is more likely to be causal, since it is unlikely that all studies were subject to the same type of bias and/or confounding. However, a lack of consistency does not exclude a causal association, since different intensity levels and other conditions may reduce the impact of the exposure in certain studies.

“More than 50 retrospective [case-control] studies have reported smoking patterns (by type and quantity of tobacco smoked, duration of smoking, and inhalational practice) in a variety of subjects with lung cancer (e.g., males and females, different occupational groups, hospitalized patients, autopsy cases, all individuals who died from lung cancer in an area, nationwide sample of individuals who died from lung cancer, and different races and ethnic groups).

Many of these subjects have been compared with matched controls also drawn from a variety of groups (e.g., healthy individuals, patients hospitalized for cancer or other diseases, deaths from cancers of other sites, and samplings of the general population). Regardless of the method, these studies have consistently found an association between smoking and lung cancer. Relative risks for smokers are consistently greater than for nonsmokers. (...)

Eight major prospective [cohort] studies have examined the relationship between smoking and lung cancer mortality in a large number of subjects, in different countries, and in different time periods. The results of these studies (...) are consistent with each other as well as with the retrospective [case-control] studies." (US Surgeon General, 1982)

Strength

The strength of an association is measured by the magnitude of the relative measure of effect. A strong association is not easily explained by potential sources of bias or confounding and hence it is more likely to be causal than is a weak association, which could more easily be the result of confounding or bias. For example, a relative risk of magnitude 10 is more likely to reflect a true causal association than one of 1.5.

"Prospective [cohort] studies have shown that the death rate from lung cancer among cigarette smokers is approximately 10 times the rate in non-smokers (...). To account for such a high relative risk in terms of an indirect association would require that an unknown causal factor be present at least 10 times more frequently among smokers (...) than among non-smokers. Such a confounding factor should be easily detectable, and if it cannot be detected or reasonably inferred, the finding of such a strong association makes a conclusion concerning causality more probable." (US Surgeon General, 1982)

Exposure–response relationship

Further evidence of a causal relationship is provided if increasing levels of exposure are associated with increasing incidence of disease.

"The strongest exposure-relationship measured in most epidemiological studies was for the number of cigarettes smoked per day at the time of entry into the study. However, other important measures of dosage include the age at which smoking began, the duration of smoking, and inhalation practice. (...) The data (...) indicate that as the number of cigarettes smoked per day increases there is a gradient of risk for lung cancer mortality. (...) Male smokers who smoked more than 20 cigarettes daily had lung cancer mortal-

ity ratios 15 to 25 times greater than nonsmokers. Similar findings were observed among female smokers (...).

Four prospective [cohort] studies which examined lung cancer mortality by age began smoking (...) show a strong inverse relationship (...), i.e., the younger the age one began smoking, the greater the lung cancer mortality rate.” (US Surgeon General, 1982)

Specificity

If a particular exposure increases the risk of a certain disease but not the risk of other diseases, this provides evidence in favour of a cause–effect relationship. However, one-to-one relationships between exposure and disease are rare and lack of specificity should not be used to refute a causal relationship.

“Tobacco smoke is a complex mixture consisting of several thousand chemical substances. These diverse substances are capable of producing more than a single biological response. The specificity of the association between smoking and lung cancer is evidenced by comparison of the magnitude of lung cancer mortality ratios to those of other cancers, as has been done in most of the prospective [cohort] studies. The mortality ratios for lung cancer are very high when compared with those of other cancers.” (US Surgeon General, 1982)

Reversibility

When the removal of a possible cause results in a reduced incidence of the disease, the likelihood of the association being causal is strengthened. Ideally, this should be assessed by conducting a randomized intervention trial, but for many exposures–disease associations, such randomized trials are impossible in practice.

“Since cigarette smoking is significantly associated with lung cancer, it is logical to expect that cessation of smoking would lead to a decrease in mortality rates from lung cancer among quitters compared to persons who continue to smoke cigarettes. In fact, all of the major studies which examined cessation showed this decrease in lung cancer risk. (...) After 15 to 20 years, the ex-smoker’s risk of dying from lung cancer gradually decreases to a point where it more closely approximates the risk of the non-smoker (...). The magnitude of the residual risk that ex-smokers experience is largely determined by the cumulative exposure to tobacco prior to smoking cessation (i.e., total amount the individual smoked, age when smoking began, and degree of inhalation), and varies with number of years since quitting smoking, as well as with the reasons for quitting smoking (e.g., quitting due to symptoms of disease).” (US Surgeon General, 1982)

Coherence

The putative cause–effect relationship should not seriously conflict with the natural history and biology of the disease.

“The final criterion is the coherence of the association between smoking and lung cancer with known facts in the biology and natural history of lung cancer. Coherence of the association has been noted in the following facts: 1) Sex differences in lung cancer mortality correlate well with sex differences in smoking patterns; 2) Population differences in lung cancer mortality correlate well with differences in tobacco consumption; 3) Time trends in lung cancer mortality correlate with time trends in prevalence of cigarette smoking.” (US Surgeon General, 1982)

Although these aspects are often referred to as ‘*criteria*’ (as in the report of the US Surgeon General (1982)), They should not be regarded as necessary conditions to established causality. The only exception is temporality—for an exposure to be a cause of a disease it clearly has to precede its biological onset.

13.5 General comments

In epidemiology, it is rare that one study alone will provide enough ‘proof’ that a certain exposure affects the risk of a particular disease. However, our degree of belief in the association will depend on the type of study design. Ecological studies *per se* can show associations, but because of their great potential for confounding they can never be used to establish causation at an individual level. Well conducted randomized trials are the best tools to assess causality, but their findings should always be interpreted in the context of all other available evidence, including evidence from other areas of research. However, for some exposure–disease relationships, such randomized interventions are not possible or are unethical.

Probability of:	Ecological	Cross-sectional	Case–control	Cohort	Randomized trial
Selection bias					
Selection of subjects	N/A	medium	high	low	low
Loss to follow-up	N/A	N/A	low	high	medium
Recall bias	N/A	high	high	low	low
Confounding	high	medium	medium	low	very low
^a Modified from Beaglehole <i>et al.</i> (1993) N/A = Not applicable.					

Table 13.14.

Probability of selection bias, recall bias and confounding for the different study designs.^a

Table 13.14 gives an idea of how much the results from different types of study may be affected by bias and confounding.

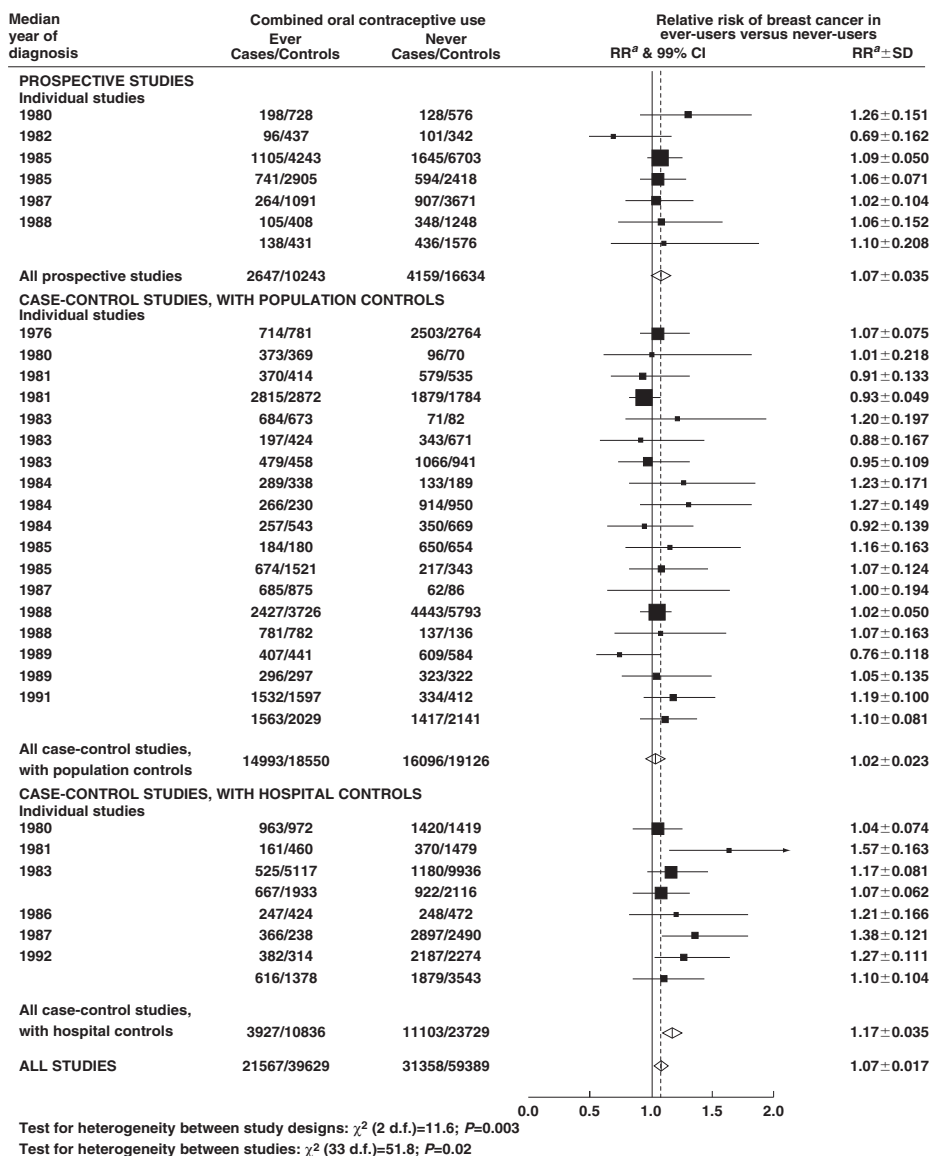
Data from various studies that have addressed the same question may be pooled together and *re-analysed* by using special techniques that take

into account the fact that the original studies may differ in a large number of respects (e.g., characteristics of the study subjects, methods of data collection). This approach is particularly useful when the effect of the exposure is likely to be small, so that a very large sample size will be necessary to detect and quantify it precisely (as in [Example 13.13](#)).

Meta-analysis is another technique used to combine results from various studies. It differs from a *re-analysis* in that there is no access to the raw data from each individual study. Only the published estimates of exposure effect from each study are available and can be used to generate a pooled overall result.

It is essential to ensure that practically all studies (published and unpublished) are included in any re-analysis or meta-analysis. This is because published studies tend to be a biased sample with an overrepresentation of those which showed 'positive' findings.

Example 13.13. *A large number of cohort and case-control studies were conducted in recent decades to examine the relationship of oral contraceptive use and breast cancer. Their results were largely inconsistent. To clarify this issue, data from most of these studies were brought together and re-analysed. The 54 studies included in this re-analysis were conducted in 25 countries in Europe, North America, Asia, Australasia, Africa and Latin America. Together they included 53 297 women with invasive breast cancer and 100 239 without breast cancer. Figure 13.4 shows relative risks of breast cancer in ever-users versus never-users of combined oral contraceptives (the most widely used type). The risk estimates varied substantially from study to study, but most confidence intervals were wide and included the null hypothesis value of 1. The pooled estimate was slightly above 1 (1.07) and its confidence interval did not include the null hypothesis value. Current users of combined oral contraceptives or those who stopped less than 10 years previously appeared to have a small increase in the risk of having breast cancer diagnosed relative to never-users, but this excess in risk disappeared 10 or more years after stopping use (Figure 13.5d) (Collaborative Group on Hormonal Factors in Breast Cancer, 1996).*



^a Relative risk (given with 99% CI) relative to never-users, stratified by study, age at diagnosis, parity and, where appropriate, the age of a woman at the birth of her first child and the age when her risk of conception ceased. Separate results are given for individual studies. Each relative risk and its 99% CI is plotted as a black square and a line. The area of the square is proportional to the amount of statistical information (i.e., to the inverse of the variance of the logarithm of the relative risk). Diamonds indicate 99% CI for totals. The solid vertical line represents a relative risk of 1.0 and the broken vertical line indicates the overall relative risk for all studies combined.

Figure 13.4.

Relative risk of breast cancer in ever-users compared with never-users of combined oral contraceptives (reproduced with permission from Collaborative Group on Hormonal Factors in Breast Cancer, 1996, © by The Lancet Ltd, 1996; see this paper for full discussion of the methods and full reference to individual studies).

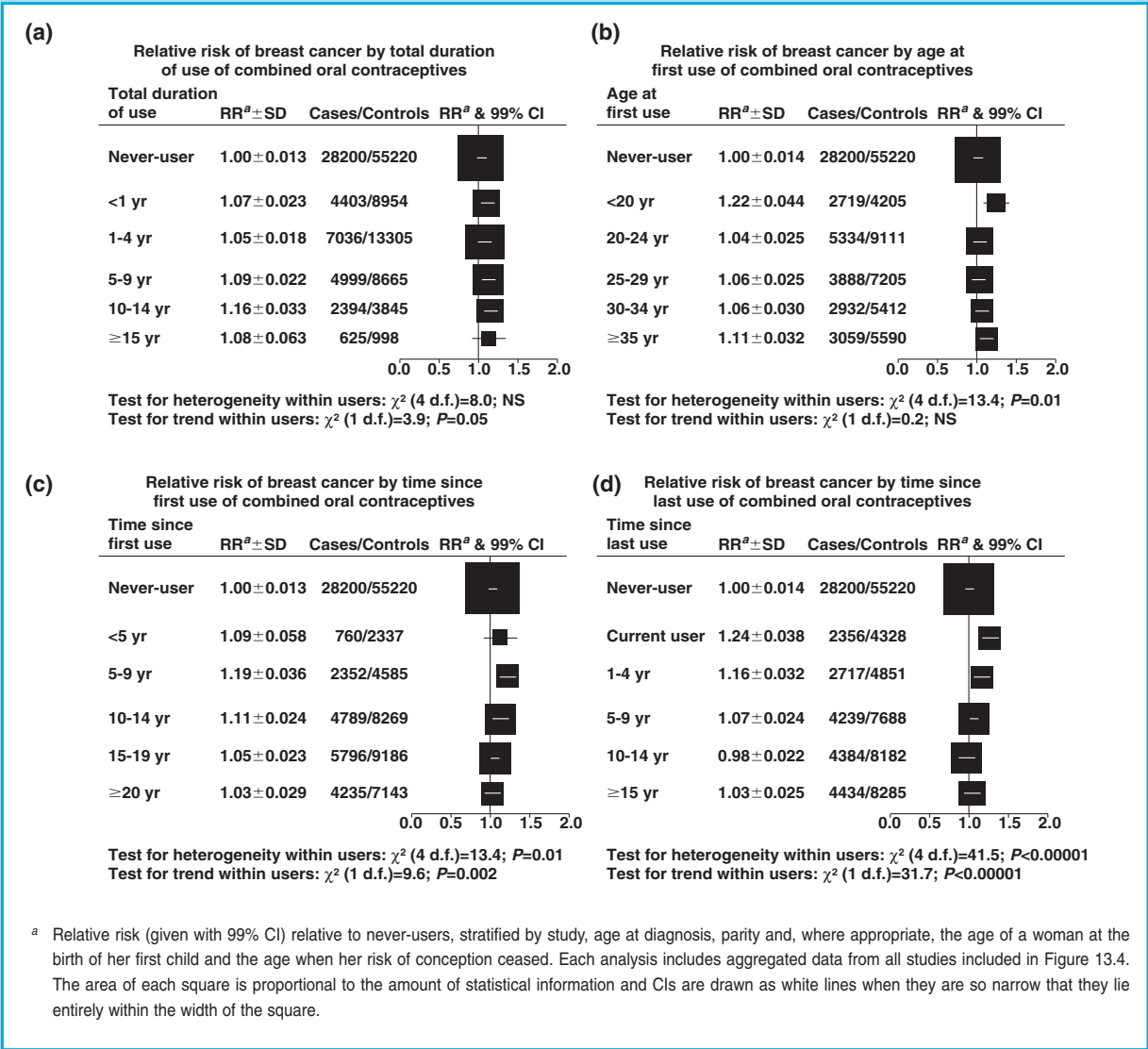


Figure 13.5. Relative risk of breast cancer for various indices of timing of use of combined oral contraceptives (reproduced with permission from Collaborative Group on Hormonal Factors in Breast Cancer, 1996, © by The Lancet Ltd, 1996; see this paper for full discussion of the methods and full reference to individual studies).

Box 13.2. Key issues

- The interpretation of findings from any epidemiological study requires consideration of bias, confounding and chance.
- *Bias* tends to lead to an incorrect estimate of the effect of an exposure on the outcome of interest. There are two main types of bias:

Selection bias occurs when there is a difference between the characteristics of the people who participated in the study and the characteristics of those who did not. Selection bias can be a major problem in case-control studies, but can also affect cross-sectional studies and, to a lesser extent, cohort studies and randomized trials.

Measurement bias occurs when measurements or classifications of outcome or exposure are not valid. The consequences of measurement bias depend on the type of measurement error (differential or nondifferential) and on whether it affects the exposure or the outcome.

- *Confounding* occurs when an estimate of the association between an exposure and an outcome is mixed up with the real effect of another exposure on the same outcome, the two exposures being correlated. There are various ways of dealing with confounding at the design stage (randomization, restriction, matching) and in the analysis (stratification and statistical modelling).
- The role of *chance* is assessed by performing statistical tests and, more importantly, calculating confidence intervals.
- Even if bias, confounding and chance do not seem to explain the observed association, we should assess the likelihood of an association being causal by using as guidelines the aspects suggested by Hill (1965).

Further reading

* The aspects that need to be considered when assessing whether a particular exposure–outcome relationship is likely to be causal were presented by Bradford Hill in a paper published in 1965.