12

VALIDITY IN OBSERVATIONAL STUDIES

OBJECTIVES

After reading this chapter, you should be able to:

1. Identify the different types of selection bias and assess whether or not a particular study is likely to suffer from excess selection bias.

2. Determine the likely direction and magnitude of a selection bias through the use of estimates of sampling fractions or sampling odds.

3. Apply the principles of bias prevention in the design of a study; for example, how to avoid detection bias in secondary-base studies.

4. Explain the differences between non-differential and differential misclassification bias in terms of sensitivity and specificity.

5. Evaluate misclassification of exposure, disease or both in 2X2 tables.

6. Explain why one cannot use the population sensitivity and specificity estimates to correct for disease status misclassification in case-control studies.

7. Evaluate the likely impact of misclassification on observed associations using sensitivity analysis.

8. Know how to apply validation studies and adjust observed data using techniques such as regression calibration.

9. Modify sample size estimates to account for misclassification.

12.1 INTRODUCTION

An awareness of the key features of study design implementation and analysis should help ensure that we obtain valid results from our research efforts. In this regard, the term **validity** relates to the absence of a systematic bias in results; that is, a valid measure of association in the study group will have the same value as the true measure in the source population (except for variation due to sampling error). To the extent that the study group and the study population measures differ, systematically, the result is said to be biased. There are 3 major types of bias:

1. selection bias: due to factors affecting the selection of study subjects, or to other factors that relate to the willingness of potential study subjects to participate in a research project
2. information bias: due to factors relating to obtaining accurate information on the exposure, outcome and covariates of interest, and
3. confounding bias: due to the effects of factors other than the exposure of interest on the observed measure of association.

In this chapter, we discuss the nature, impact and prevention of selection and information bias; confounding is discussed in Chapter 13.

Most analytic studies are conducted on non-randomly sampled study subjects, so there is always some uncertainty about how well the attributes and the associations in the study group reflect the attributes and associations in the larger source population from which the study group is drawn. In addition, once the study groups are selected, we must be able to accurately measure the exposure, extraneous factors and outcome of interest, and control confounding, if we want to make valid conclusions about the exposure-outcome association. In this context, an **internally valid** study will allow us, based on the study group data, to make unbiased inferences about the association(s) of interest in the **source population**. **External validity** relates to the ability to make correct inferences to populations beyond the source population (the first of these being the **target population**). In this regard, while it is certainly desirable that the study and source populations be 'representative' of the larger target population, one should not sacrifice internal validity in order to gain external validity (see Alonso *et al*, 2007). In the extreme, there is no value in being able to extrapolate incorrect results. **Generalisability** is an inferential step beyond external validity and refers to the ability to develop and extend valid scientific theories to broadly defined populations (*eg* associations that are valid across populations and/or species).

12.2 SELECTION BIAS

Selection bias results from the fact that the composition of the study group(s) differs from that in the source population and this biases the association observed between the exposure(s) and the outcome(s) of interest. This bias can have large effects on study results. Hence, the criteria used to select study subjects and maintain them in the study are important to describe (Grimes and Schulz, 2002; Sandler, 2002; Beck, 2009, and pertinent sections in Chapters 7-10). From a sampling perspective, each study will have an objective that relates to a **target population** (*eg* the impact of a disease of dairy cows in a defined area such as a province in Canada). For practical purposes, it is often necessary to obtain the study subjects from a subset of the target population (*eg* all dairy herds in the province with computerised records), and this constitutes the **source population**. The actual group of subjects (*ie* participating herds) in which the study

is conducted is called the **study group** (see Section 2.1.3). In the ideal, the source population will completely reflect the target population, and the study group will completely reflect the source population.

As noted in Chapters 7-10, we investigate associations by contrasting outcomes in 2 or more groups of subjects. As described in Chapter 1, the ideal comparison group for causal inferences is the counterfactual group. For example, in a cohort study, the ideal counterfactual group for the exposed study group would be the exact same subjects if they had not been exposed. However, as this ideal group is non-existent, we strive to select the non-exposed study group in a manner that ensures that the 2 groups are totally comparable with respect to all factors that might bias the measure of association. Our intent is to have the association that is under investigation be the same in the study group as in the source population. From a selection bias point of view, this means that the 2 groups under study should be comparable at the initiation of the study and any decrease in this 'comparability' throughout the study period should not be a result of the study process. We also would note that clinical trials (Chapter 11) are not immune to selection bias. Although randomisation helps ensure that the groups receiving the treatment(s) are comparable (*ie* exchangeable), since the study subjects usually are volunteers, they may differ from the source population in a manner that leads to biased results (*eg* if the treatment interacts with the characteristics of the study groups that differ from the source population) (Beck, 2009).

If it occurs, **selection bias** happens either before the study begins, or during the study implementation, and it results from the procedures used to obtain study subjects and factors that influence participation and participant behaviour (*eg* their management of the herd). These factors influence participation in the study in such a way that the composition of the study group(s) differs from that in the source population and this biases the observed association. The basic conditions for selection bias can be shown pictorially using the techniques of directed acyclic graphs and the concept of statistical conditional dependence (Hernan *et al*, 2004; Sjolander *et al*, 2008). In the left column of Fig. 12.1, we indicate that both exposure (*E*) and the outcome disease (*D*) directly affect the selection (*S*) of study subjects. In this depiction, *E* and *D* are independent of each other (*ie* not associated) in the source population; however, when we conduct the study using only the responders (*ie* condition on selection *S*), assuming that there is some non-response, *E* and *D* become associated. Alternatively, had *E* and *D* been associated in the source population, the observed association in the study group would differ from that in the source population; *ie* selection bias would occur. On the right, disease directly affects selection in the source population. Exposure does not directly affect selection, but unless the **bias variable** (*eg* behaviour or attitude) which is directly related to (or correlated with) exposure and with selection is controlled or 'adjusted for', exposure will be statistically related to selection and a biased association between exposure and disease will result in the study group. As a third example (not shown), the bias variable could be related to the disease, not the exposure, and the exposure be directly related to selection. In summary, as Hernan *et al*, 2004, demonstrate, using directed acyclic graphs (*aka* causal diagrams), that selection bias is a result of conditioning on the common effects of exposure and disease, or on the effects of variables related to exposure and disease. Shahar (2009) has elaborated on this with application to information bias.

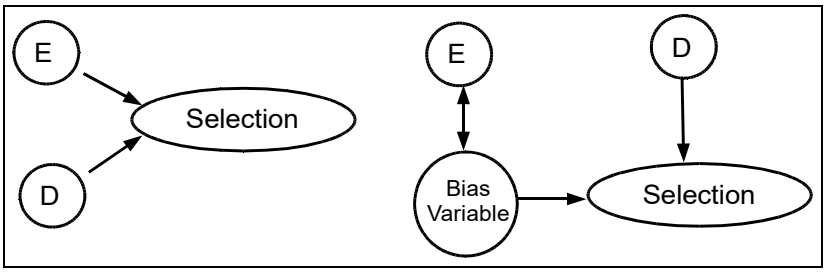

Fig. 12.1 A diagram depicting basic conditions for selection bias
Note $OR_{ED}=1$, but $OR_{ED|S}\neq 1$

We also can gain some understanding of selection bias using sampling fractions. Assume the source population and study group have the structure shown in Table 12.1 (upper-case letters represent the number of subjects in the source population, lower case letters the study group).

Table 12.1 A representation of the structure of the source population and study group

Source population structure				Study group structure			
	E+	E-			E+	E-	
D+	A_1	A_0	M_1	D+	a_1	a_0	m_1
D-	B_1	B_0	M_0	D-	b_1	b_0	m_0
	N_1	N_0	N		n_1	n_0	n

12.2.1 Sampling fractions and sampling odds

Our intent is to select the study group in a manner that avoids selection bias. Formally, the study group is a sample of the source population. We can visualise the sampling fractions (*sf*) in each of the 4 categories of exposure and disease as:

$$\begin{aligned} sf_{11} &= a_1 / A_1 \\ sf_{12} &= a_0 / A_0 \\ sf_{21} &= b_1 / B_1 \\ sf_{22} &= b_0 / B_0 \end{aligned} \qquad \textit{Eq 12.1}$$

where the subscripts refer to the row-cell combination in the 2X2 table structure (row 1, column 1 is the upper left cell: exposed and diseased *etc*) of Table 12.1. If the study subjects were obtained by random selection, of '*n*' from '*N*' subjects, the 4 sampling fractions would be equal, except for random variation. Under this selection method, it is reasonable, and correct, to assume that if all 4 sampling fractions are equal, there is no selection bias (Morabia, 1997). Furthermore, if the sampling fractions are equal, the odds ratio (*OR*) of the sampling fractions (OR_{sf}) equals 1. It is noteworthy that the 4 sampling fractions can be unequal and not produce bias in the observed *OR* provided the OR_{sf} equals 1. Under this latter condition, there is also no bias to the risk ratio (*RR*) if disease is infrequent. In reality, we rarely know the values of the *sf* so this limits the practical utility of this approach. However, understanding the role of sampling fractions provides a theoretical basis for understanding the conditions under which bias will or will not occur. See Example 12.1 for an application of using the sampling fraction odds ratio to investigate selection bias arising from non-response.

In practise, sampling odds might be easier to conceptualise than the individual sampling fractions. For example, in a risk-based cohort, or longitudinal study, one could express the sampling odds of disease ($so_{D+|E}$) among exposed subjects versus the sampling odds of disease

Example 12.1 Selection bias due to non-response

In order to demonstrate that non-response can bias an association measure, we first give an hypothesised example where the non-response is related only to exposure and not to the outcome. In this situation, one would not expect the non-response to bias the measure of association. For this example, we will initially assume the following scenario:

- 10% of the subjects in the source population are exposed
- In the exposed subjects in the source population, 30% are non-responders (nr) and that the risk of the outcome in the non-responders is the same as that in the responders (r) at 25%.
- In the non-exposed subjects in the source population, 10% are non-responders and these subjects have the same risk of the outcome as the responders at 12%.

Based on these assumptions, the source population structure is:

	Exposed_r	Exposed_{nr}	Non-exposed_r	Non-exposed_{nr}
D+	175	75	972	108
D-	525	225	7128	792
	700	300	8100	900
Risk	0.25	0.25	0.12	0.12

If we initially contact 100 exposed and 100 non-exposed individuals, in the source population, the overall response 'rate' is 80% and the study group will have the following structure:

	Exposed_r	Non-exposed_r
D+	18	11
D-	52	79
	70	90

Apart from rounding error, the ratio of risks (*RR*) in the study group (*RR*=2.04) matches the risk ratio in the source population (*RR*=2.08), as does the *OR* (2.49 vs 2.44). There is no bias.

Now, given exactly the same response risks, we will assume that non-response is related to both exposure and outcome, and the risk of the outcome is twice as high in non-responders as in responders in both the exposed and non-exposed groups.

Under this scenario (and ignoring rounding errors), the population structure would be:

	Exposed_r	Exposed_{nr}	Non-exposed_r	Non-exposed_{nr}
D+	133	114	891	198
D-	567	186	7209	702
	700	300	8100	900
Risk	0.19	0.38	0.11	0.22

The ratio of the risks in the source population is 0.247/0.121=2.04 and the odds ratio is 2.38.

(*continued on next page*)

Example 12.1 (*continued*)

As before, if we initially contact 100 exposed and 100 non-exposed individuals, the study group will have the following structure (apart from sampling error):

	Exposed$_r$	Non-exposed$_r$
D+	13	10
D-	57	80
	70	90

Now the study group *RR* is 0.19/0.11=1.73, and the *OR* is 1.90; both are biased estimates of the true associations.

Note that in this scenario the *sf*s are:

$$sf_{11} = 13/247 = 0.053$$
$$sf_{21} = 57/753 = 0.075$$
$$sf_{12} = 10/1089 = 0.009$$
$$sf_{22} = 80/7911 = 0.010$$

and the odds ratio of the *sf*s is:

$$OR_{sf} = \frac{0.052 * 0.01}{0.075 * 0.009} = 0.8$$

Thus, based on the *OR* of the *sf*s, the bias would be expected to be toward the null, and we note that the true *OR* multiplied by the sampling fraction odds ratio gives the observed *OR* (*ie* 2.38*0.8=1.90).

If we doubled the non-response risk in both exposed and non-exposed groups, the sampling fraction of the odds ratio would be 0.66. Thus, it is conceivable to produce considerable bias from this form of selection bias.

in the non-exposed subjects as:

$$so_{D+|E+} = sf_{11} / sf_{21}$$
$$so_{D+|E-} = sf_{12} / sf_{22} \qquad \textit{Eq 12.2}$$

If these selection odds are equal, there is no bias, and this becomes the goal of study-subject selection strategies in observational studies. If the ratio of the sampling odds is greater than 1, then the bias is away from the null; if the ratio of the sampling odds is less than 1, the bias is toward the null. Thus, from a practical perspective in designing a cohort study we need to ask ourselves, over and above the associations between exposure and disease in the exposure cohorts, am I more (or less) likely to select for disease in the exposed than in the non-exposed cohort? As noted, in Example 12.1, because of the non-response, the sampling odds for disease among the exposed is 5.89 (*ie* 0.053/0.009), and among the non-exposed, it is 7.5 (*ie* 0.075/0.010) giving a ratio of sampling odds equal to 0.8. In relative terms, we have over-sampled disease in the non-exposed, and biased the observed association toward the null. Similarly, in designing a case-control study, we wish to avoid a differential selection for exposure that is over and above any associations between exposure and disease in the case and control groups in the source population.

12.3 EXAMPLES OF SELECTION BIAS

12.3.1 Choice of comparison groups

A general principle is that the study groups should be selected from the same source population. In cohort studies, it is important that the subjects in the non-exposed group be comparable with those in the exposed group with respect to other risk factors for the outcome that are related to the exposure. This is more of an issue with the usual 2-group (*ie* exposed and non-exposed) cohort design, than with a single-cohort study design. Similarly, in a case-control study, it is important that the control group reflects either prevalence of exposure in the 'non-case' members of the population from which the cases arose (risk-based study) or the proportion of exposed animal-time at risk for the non-case group in the source population (rate-based study). Since the members of the study groups are rarely obtained by random sampling, decisions about how to select the study subjects must include knowledge about the context and the biology of the problem being investigated as well as the structure and dynamics of the source population.

As an example, a recent study has documented how the design of a surveillance system can bias the risk of disease by breed type (Ducrot *et al*, 2003). In France, BSE cases were identified through a mandatory reporting system (MRS) based on clinical cases, as well as targeted surveillance (TS) using a diagnostic test on cattle found dead on the farm, subjected to euthanasia, or emergency slaughter. The MRS detected "34 BSE cases, all from dairy herds, while the TS program discovered 49 cases, with only 36 (73.5%) from dairy herds". Investigations subsequently demonstrated that the MRS system was biased toward detecting cases in dairy herds (perhaps because dairy cattle in England had been reported to have an excess of BSE cases, and the system depended on recognising and reporting clinical signs, hence subclinical or atypical cases were missed). The sampling of the TS system was not biased by production system but the obvious clinical cases were excluded. Thus, data from both systems were needed to get an unbiased measure of association between breed and BSE risk.

Tongue *et al* (2006) describe another instance where the selection of the study subjects can bias prion protein genotype in studies of scrapie. They posited that different genotypes may be related to different ages when scrapie develops—culling of early cases and their relatives would alter the subsequent distribution of genotypes in the flock or population of interest. Also, given knowledge of the relationship of genotype to scrapie risk, the willingness of owners to test their flock could vary from genotype to genotype. Information from different databases are very helpful in helping to interpret associations under these conditions (as was found in the previous example of BSE).

12.3.2 Non-response

Non-response bias can be a major problem in both descriptive and analytic studies and its level and effects are often understated (Stang, 2003; Morton *et al*, 2006; Mezei and Kheifets, 2006). Non-response leads to bias if the association between exposure and the outcome in the responders differs from that in non-responders; hence, the association in the study group (*ie* only the responders) differs from that in the source population. Although non-response produces its effects through a process similar to a confounding variable, it may not be directly controlled in the same manner. The stronger the association between exposure and disease and

the greater the proportion of non-responders, the greater the potential bias. In veterinary research, non-response on behalf of the owner could be a surrogate indicator for management, housing, or feeding differences of the owner's animals that could relate to both the outcome and the exposure factor. In studies where humans are the units of concern, willingness to enrol in a study might be related to both the exposure and the outcome, hence the study group produces a biased response. One factor that appears to influence participation in surveys and observational studies is the socio-economic status (SES) of the potential participant; subjects from a higher SES are more likely to participate than those from a lower SES. In veterinary studies, if SES is associated with both the animal's or the herd's exposure and the disease of interest, then a bias could result.

One way to assess the possible effects of response bias is to ascertain if the extent of non-response within each group (*ie* the exposure cohorts, or the case and control groups) is approximately equal. If they are equal, or approximately equal, there will be little to no selection bias. Low overall response rates do not necessarily result in selection bias and high response rates do not guarantee a lack of bias (Nohr *et al*, 2006; Bjertnaes *et al*, 2008). Achieving an equal response in the groups based on exposure or outcome should be a major consideration when designing and implementing observational studies. A second approach to assess possible bias is to compare responders and non-responders using whatever information you have on exposure, disease, or other features, recognising that because the owner/participant won't respond (or collaborate), the data might be limited. For example, if some information on potential participants in a study of risk factors for foot problems in dairy cows is available (*eg* milk production of dairy herds), then we can compare this between responders and non-responders) to see if they differ. If the differences are negligible, or if the variable(s) the groups differ on is not related to the outcome of interest in the study subjects, then selection bias will likely have only a small impact on the study results. Example 12.1 demonstrates how non-response produces bias.

12.3.3 Selective entry or survival bias

Sometimes the groups we study are highly selected in that only subjects that possess certain desirable attributes are selected for membership. The analogous problem in studies of humans is called the 'healthy worker' effect, and is a major issue especially in occupational-health studies. In veterinary research, adult food-producing animals (sows, cows *etc*) are highly selected for herd membership on entry (*eg* they might need to meet specific growth rate and fertility criteria) and, once admitted to the herd, these animals must maintain certain production standards (*eg* number of piglets produced per year) to remain in the herd. Similarly, horses that are currently racing are a biased subset of all horses that started to enter the race circuit, and they are very likely to be healthier than all horses that have raced at least once. As one example of this potential bias, if we wanted to assess the impact of post-partum disease on fertility in dairy cows, and we chose the calving-to-conception interval as an outcome measure, we could get a biased view of the disease effects. Animals that did not get pregnant would be excluded from the outcome measure. Hence, because these cows did not pass the entry criterion of becoming pregnant, they are excluded from the study and yet this failing is a crucial component of assessing the fertility status of the herd. Braback *et al (*2006) demonstrated that selection bias likely contributed to the lower prevalence of asthma and allergic keratoconjunctivitis in farmers —farm children with these condition are less likely to pursue farming because of the physical demands and the known risks of other respiratory conditions (*eg* chronic bronchitis) than

children without these conditions. Knowledge of this helped them interpret the results of earlier studies investigating the association between occupation and respiratory disease (the presence of early disease was affecting exposure (occupation) and also, later-life disease status.

With respect to selective survival, if both the exposure and disease being studied affect whether or not a food animal remains in the herd (or whether or not a horse is still racing), then a study group drawn from only 'existing' (*eg* racing) animals might give a biased measure of association between the exposure and disease. The premature removal of animals from the original group might be highly correlated with the exposure factor and the outcome, thus leaving the study group as a biased subgroup from the source population. Whenever selective survival is likely to be an issue, the study group(s) should be drawn from all animals that entered the herd (or ever raced) during a specified time period, not just from animals that are in the herd (or are racing) at the start of the current study period. In many of these instances, implementing longitudinal studies commencing at the birth, or herd entry, of the study subjects would provide the best evidence of the impact of specific exposures throughout the life of the study subjects (Brabeck *et al*, 2006).

Survival bias can also result from the use of prevalent cases of disease (*eg* in case-control or cross-sectional studies). If the duration of survival after disease onset differs by exposure status, then bias could result. Cross-sectional studies are problematic in this regard, and partly for this reason, it is recommended that case-control studies usually should include only incident cases.

Unintentional selection bias might be at play in many studies of antimicrobial resistance patterns (Miller and Tang, 2004). Often the data are based on isolates obtained from clinically ill subjects, or from subjects prophylactically exposed to antimicrobials. Hence, many of the isolates would have been exposed to antimicrobials prior to culturing of tissue specimens. Thus, the number and type of bacterial isolates, and their level of antimicrobial resistance (or minimal inhibitory concentrations) might be more a function of what antimicrobials had been used and how effective they were at reaching and killing susceptible organisms in the tissue samples that get cultured than of the prevalence of pathogenic organisms or their level of antimicrobial resistance in the source population. If the objective of the study is to describe the extent of antimicrobial resistance in the source population, samples should be obtained from randomly selected subjects (some of which may have been exposed to antimicrobials). The impact of exposure to antimicrobials on the level of resistance could be assessed in a valid manner.

12.3.4 Detection bias

Detection bias results when the probability of identifying the disease (or outcome) differs by exposure status. In cohort studies, detection bias is best viewed as a misclassification. It can arise if those assessing the outcome know the exposure status of the study subject, and if they alter their assessment of the outcome because of that knowledge.

In case-control studies, the central issue in detection bias is one of selection, in that animals that have the disease of interest might be misclassified as not having that disease because they were less likely (or never) to be examined for the disease (see Section 12.6). This potential bias is of concern when a large percentage of the cases would be found (and therefore identified as potential study subjects) as a result of undergoing examination in a screening or diagnostic process where participation is influenced by exposure status (*ie* the act of being assessed is directly or indirectly influenced by the exposure status). Given this scenario, the issue is how

best to select controls. A frequently suggested guideline is that the controls should be non-cases that have undergone the same screening, but the nature of the exposure, disease and the context of diagnostic testing need to be considered (Harris *et al*, 2005). Here the concern was misclassifying untested subjects as non-cases when, in reality, they were mildly diseased undetected cases. In many instances, severity of illness would be the 'bias' variable (Fig. 12.1) that leads to testing (or not).

Detection bias was at the root of protracted discussions about the appropriate control group for a series of uterine cancer cases in a study of the potential impacts of hormone-replacement therapy in women (Greenland and Neutra, 1981). Women on estrogen tended to evidence vaginal bleeding and therefore would be examined (in a manner that could lead to the detection of uterine cancer) more frequently than women not on estrogen. Hence, the possibility of detection bias was raised. Some researchers argued that the controls should be restricted to those women who had been examined because of vaginal bleeding and found negative for cancer. However, it was subsequently determined that because all cases of uterine cancer were detected, ultimately (regardless of screening), it was not necessary to enforce the general principle that controls should undergo the same testing regime as the cases. Another method to evaluate detection bias is described in Section 12.4.1.

Another example of potential detection bias was related to misclassification of the outcome (Singer *et al*, 2001). These workers were selecting birds with avian cellulitis in the slaughter plant using the presence of certain gross lesions as indicators that the birds had the disease, and then culturing these birds for specific strains of *Escherichia coli*. Their concern was that, if certain strains of *E. coli* only produced lesion(s) that were not being detected visually, then these birds would not be selected. Hence, only a biased subset of the *E. coli*-caused lesions would be detected. These researchers developed a method to assess possible selection bias based on comparing the findings in the birds that they detected with findings in birds detected using independent criteria by the USDA inspectors. In general, it is desirable to have a sensitive and specific set of inclusion criteria when selecting study subjects.

12.3.5 Admission risk bias

Admission risk bias has provoked much debate over the validity of secondary-base case-control studies, and is the basis of **Berkson's fallacy** (Schwartzbaum *et al*, 2003; Sadetski *et al*, 2003). In this instance, the probability of admission to the registry or hospital (*ie* the secondary-study base) is related to both the disease and the exposure. That the exposure of interest has an independent risk of admission to the hospital or registry (*ie* $p(H|E+)>0$) is a prerequisite for a bias to occur. In practise, this effect is through the production of exposure-related diseases other than the case disease of interest. A differential admission risk between the cases (p(H|D+)) and the average admission risk of the control disease categories (p(H|D-)) is also needed to produce the bias, but this is a very common situation in most case-control studies. Thus, the controls drawn from the hospitalised population might not reflect the actual exposure status of the population from which the cases arose.

In terms of the direction of bias, provided exposure leads to 'being in the registry', if the risk of hospitalisation (*ie* being in the registry) is greater for the disease of interest than the average risk for the potential controls, the sample (*ie* study group) *OR* will be less than the source population *OR*. Thus, if the study data leads to a statistically significant *OR*, the true association in the source population would be even stronger. Conversely, if the risk of hospitalisation (*ie*

being in the registry) is lower for the disease of interest than the average risk for the potential controls, the sample *OR* will be more than the source population *OR*.

A frequently cited example involves investigations of the association between smoking and lung cancer using hospital-based case-control studies. Smoking can lead to hospitalisation for many diseases and thus, it is suspected that the prevalence of smoking is higher in many of the disease categories from which controls might be selected (Sadetzki *et al*, 2003). In a similar manner, since the frequency of many animal diseases increases with age, then the average age of animals in veterinary hospitals would be older than the age of the general population and this could bias the association between age and specific diseases when the non-diseased study subjects are selected from veterinary hospitals. Thus, it is important to try and obtain quantitative estimates of the likely degree of bias that different potential control groups might produce (see Example 12.2). Since it is nearly impossible to assess the degree of selection bias in any given secondary-base study, this constrains the inferences that should be drawn from a single secondary-base study. In Chapter 9 and Section 12.4, we develop guidelines for selecting cases and controls in a manner to prevent or minimise the magnitude of bias.

Because of the difficulties in selecting appropriate controls in secondary-base studies, some researchers obtain controls directly from the putative source population. Tam *et al* (2003) have documented that disease severity and societal factors influence the inclusion of subjects in registries for infectious intestinal disease. Their research suggests that we need to be careful when using population controls as they may not be representative of the actual population that gave rise to the cases. The same authors also supported the use of case-case studies to avoid this potential bias (see Section 10.3).

12.3.6 Loss to follow-up and follow-up biases

Similar to non-response bias, if there is a differential loss to follow-up that is related to the exposure status and the outcome, then bias in the measure of association will result. Thus, in the design and implementation of the study protocol, we should try to follow-up study subjects as completely as possible and minimise losses. Failing that, we should try to ensure that both groups are followed with equal rigour (this tends to equalise, but does not reduce, the losses). Unfortunately, the larger the losses, the more difficult it becomes to ensure equality of losses across the study groups. Robinson *et al* (2007), have surveyed the literature for strategies to reduce follow-up losses and they provide a list of the 12 most frequently stated strategies. Among the more frequent strategies were: obtaining community involvement; creating study identity (*eg* study name and consistent letterhead); having study personnel with excellent communication skills; clearly explaining the benefits of the research, having regular scheduled contacts with participants; regular reminders; minimising participant burden, and providing participant specific benefits (perhaps a free consultation or a specific information package).

Bias also can result from differential management of exposed and non-exposed subjects during the study. More generally, behavioural changes in study subjects as a result of being studied are referred to as the **Hawthorne effect**. In an observational study, the role of the researcher is to observe, not alter, the normal (*ie* usual) events experienced by the study subject or his/her animals. However, it is often difficult to 'hide' the reason for the study and the act of enquiring into specific lifestyle/management/housing factors could lead the participant to modify his/her protocols in ways that are not obvious to the researcher. This could lead to differential

management by exposure status, or at the very least, it could lead to exposure status changes during the study period. Ducrot *et al* (1998), describe the interrelations between the observer (*ie* the researcher) and the observed (*ie* the participants) in the context of on-farm studies. Being aware of this effect and implementing the study in a manner designed to minimise any bias, through complete and equivalent follow-up of the groups, is the best prevention.

12.3.7 Bias due to missing data

If the missing data are distributed randomly, they will reduce precision and power, but not lead to biased associations. However, missing data can create a bias similar to non-response, because the researcher must adjust the analysis (*eg* impute the missing value) (Cole *et al*, 2006; Fraser and Yan, 2007; Fraser and Yan, 2009), drop the variable(s) with missing values (and possibly leave a confounding bias), or drop the observation (and hence, effectively produce a non-response). Thus, minimising missing data and assessing whether the level of missing data is equivalent in the groups being compared (*eg* cases and controls) are recommended features of study design. We discuss the problem of missing values further in Section 15.5.

12.4 REDUCING SELECTION BIAS

Most of the specific recommendations for preventing selection bias are contained in Section 12.3 or in the study design chapters (Chapters 7-10) and will not be repeated here. However, being aware of the potential pitfalls in selecting study subjects, and conceptualising how these pitfalls might apply to selection of study subjects from the proposed source population is the first step in prevention. In cohort studies where explicit exposed and non-exposed groups are selected, care needs to be taken when selecting the comparison group, and due consideration should be given to minimising non-response bias, missing data, and ensuring equal follow-up and preventing detection bias (see Chapter 8 for details). Case-control studies (Chapter 9) are particularly susceptible to selection bias because of the (usual) built-in differential risk of inclusion based on disease status. Thus, minimising a differential response to study participation between cases and potential controls should be a major focus of study subject selection procedures. With regard to selection, the comparison group in case-control studies need not be similar to the case group in all respects except for the disease of interest, but rather just with respect to the factors related to the outcome that might lead to being included in the study. A key principle for control selection is that they should represent the proportion exposed, or the exposure time, in the non-diseased members of the source population. This is chiefly a problem in secondary-base studies and to circumvent it, we implement the guideline of selecting controls only from non-case diagnostic categories that are unlikely to be associated with the exposure. In addition, where possible, case-control studies should be based on only incident cases and the control subjects should come from the same source population as the cases (See Chapter 9 for details). Even with all these precautions, care must be taken in making broad inferences from a single case-control study using secondary databases.

12.4.1 Evaluating and correcting selection bias

For valid and effective control of selection bias, 1 of 2 conditions needs to be met: the factors associated with selection must be antecedents of both exposure and disease, or the distributions of exposure and disease must be known in the source population. Under the first condition, the

bias can be controlled in a manner similar to confounding; for example, if owner income might cause selection bias in a secondary-base case-control study, it can be measured and controlled in the analysis. Geneletti *et al* (2009), and Alonso *et al* (2006) describe methods to test for, and correct, selection bias in case-control studies, based on using data internal to the study, or in some instances data that are external to the study group. The variable(s) which is strongly related to selection, or study participation and produces the bias (called a **bias breaker** by Geneletti *et al* (2009)) needs to be identified so that unbiased estimates of its population distribution can be obtained (this is necessary so that these 'correct' estimates are not associated with 'selection'). We refer you to Geneletti *et al* (2009) for the actual calculations and the constraints needed to select valid adjustment factors.

However, as example, identifying and adjusting for the bias variable in case-control studies of the impact of smoking, the observed association can be adjusted for selection bias, if the prevalence of smoking in the source population (from which the cases were obtained) is available (*ie* it replaces the observed proportion of smokers in the study control subjects). In general, Berkson's fallacy can be prevented if estimates of the hospitalisation rates of the non-diseased subjects are available. Although this is quite difficult to implement, the potential impact of differential admission risks could be investigated in sensitivity analyses. Similarly adjustment for the potential effect of SES on participation can be made using information on the combined level of SES in the cases and controls (this approach uses data internal to the case-control study to 'adjust the biasing variable SES'). External data on the SES from a recent census in the source population could also be used. In the case of detection bias in the studies of estrogen use and uterine cancer, selection bias could be 'corrected' by using the prevalence of vaginal bleeding among women with uterine cancer in the source population (here again, this corrects for the distribution of the biasing variable—vaginal bleeding). Alonso *et al* (2006 and 2007) describe the use of inverse probability weighting to adjust for selection bias as a result of dropouts during a cohort study. Berger (2005) describes how to use reverse propensity scores to detect and 'correct' for selection bias.

In veterinary medicine, we rarely have solid estimates of the *sf*s or of the distribution of the bias variable. However, we can assess the potential bias from single estimates of sampling fractions or the bias from a distribution of sampling fractions using a stochastic approach. In Example 12.2, we use software developed by Orsini *et al* (2008), to demonstrate both deterministic and stochastic adjustment for potential selection bias based on estimating the sampling fractions in case-control studies. The examples given above of the bias-variables hopefully will help us to identify the key variable(s) that affect selection in our studies and assess their potential impact on the study results. Sensitivity analyses (using a range of parameter estimates) can be useful for this purpose (Sjolander *et al*, 2008).

12.5 INFORMATION BIAS

The previous discussion was concerned with whether the study subjects had the same exposure-disease association as that which existed in the source population, and we assumed that disease and exposure were correctly classified. We now move on to discuss the effects of incorrectly classifying, or measuring, the study subjects' exposure, extraneous factors and/or outcome status. When describing errors in classification of categorical variables, the resultant bias is referred to as **misclassification** bias; if the variables of interest are continuous, then we term the erroneous result as **measurement error** or bias. **Information bias** is a collective term for either

Example 12.2 The evaluation of potential selection bias based on estimates of sampling fractions

The table below displays the frequency of no-home-cooked food and home-cooked food that was fed to lactating dogs with and without atopic dermatitis in Sweden. Data from Nødtvedt *et al*, 2007.

	No home-cooked food	Home-cooked food
Cases	31	16
Controls	25	30

The authors noted an increased risk (*OR* 2.33; 95% CI 1.04-5.19) of atopic dermatitis in the high risk breeds of dogs that was associated with feeding 'commercial food' (*ie* no home-cooked food).

The remainder of this example is developed for pedagogic purposes only; we do not imply that the selection bias shown here actually exists. Suppose we know that selection bias is likely and we have a good idea of the relative selection probabilities. We will adjust the odds ratio, deterministically, using the following sampling fractions (*sf*) to assess the potential impact of the selection bias:

	Deterministic	Stochastic
sf exposed cases (E+D+)	0.5	triangular (0.4, 0.5, 0.6)
sf non-exposed cases (E-D+)	0.6	triangular (0.5, 0.6, 0.7)
sf exposed controls (E+D-)	0.05	triangular (0.01, 0.05, 0.1)
sf non-exposed controls (E-D-)	0.1	triangular (0.05, 0.1, 0.2)

The deterministic *sf*s were chosen to reflect our belief that those who prepare home-cooked food (*ie* the unexposed) would be more likely to participate in the study than those who do not. The cases have a higher participation level than the controls. While the observed *OR* was 2.33, the 'adjusted' *OR* (after accounting for the *sf*s) was 1.40 (95% CI of 1.04, 5.19); the strength of association is now considerably (67%) reduced.

To demonstrate stochastic sensitivity analysis, suppose we know the likely direction of selection bias but we don't have a precise idea of the actual *sf*s. We specified a triangular distribution for the *sf*s as shown above (*eg* the *sf* for E+ D+ was assumed to have a minimum value of 0.4, a maximum of 0.6 and a most likely value of 0.5). This keeps the same direction of bias as before but now we are uncertain about the actual sampling probabilities. The impact of considering the *sf*s was to reduce the *OR* to approximately 1.28 with 95% of the estimates falling between 0.43 and 3.13. Once again note the downward direction of the *OR* from what we would have expected to see in the absence of selection bias. Clearly, if selection bias was present, at about the same magnitude as we specified here, then the true association was considerably weaker than what was observed.

of these biases. Information bias can alter the magnitude and direction of estimates of association, in ways which might not be intuitive. Also, the errors in classification, or measurement, can affect different measures of association differently (*ie* risk ratio versus risk difference). Hence, for our purposes, we will focus primarily on the effects of information bias on relative measures of association (*RR*s and *OR*s). In the discussion that follows, the study subject could be an individual animal, or participant, or a group of individuals, such as a herd.

We will begin this topic with a review of the basics of misclassification—the most studied of information biases.

12.6 BIAS FROM MISCLASSIFICATION

Misclassification bias results from a rearrangement of the study individuals into incorrect categories because of errors in classifying exposure, outcome or both. Non-compliance with an assigned treatment in a clinical trial can also produce misclassification bias, because the subject was not actually receiving the treatment specified. With categorical measures of exposure, outcome, or other covariates, especially dichotomous measures (*ie* exposed or not, diseased or not), the errors of classification can be described in terms of sensitivity and specificity as shown in Chapter 5. Here, sensitivity (*Se*) for a given event (*eg* exposed) is the probability that an individual with the event will be classified as having the event. Specificity (*Sp*) is the probability that an individual without the event (*ie* not exposed) will be classified as being without the event.

12.6.1 Non-differential misclassification of exposure

The tabular data layout is the same as shown in Table 12.1. The true cell values for the study group are represented by a_1, b_1, a_0, and b_0, with m_1 diseased and m_0 non-diseased, n_1 exposed and and n_0 non-exposed subjects. The observed cell values will be denoted with the prime symbol) as a_1', b_1', a_0', and b_0'.

If misclassification of the exposure and outcome are independent (*ie* errors in classifying exposure are the same in diseased and non-diseased animals and vice-versa when classifying disease in exposed and non-exposed subjects) then the misclassification is called **non-differential**. With non-differential misclassification for exposure we have:

$$Se_{E|D+} = Se_{E|D-} = Se_{E} \qquad \text{and/or} \qquad Sp_{E|D+} = Sp_{E|D-} = Sp_{E}$$

where Se_E is the sensitivity of exposure classification and Sp_E is the specificity of exposure classification.

How do these errors relate to our observed data? We begin by assuming misclassification frequencies for exposure, denoted as Se_E and Sp_E, and assuming $Se_{D+}=Sp_{D-}=100\%$. The true cell frequencies are shown in the left column and the observed frequencies in the right column of Table 12.2. Clearly, the observed cell values are a mixture of correctly and incorrectly classified study subjects. Since we are only misclassifying exposure in this example, the number of diseased and non-diseased subjects represents the true number of subjects in each health category. With dichotomous exposures and outcomes, non-differential errors will bias the measures of association toward the null (given that the $Se_E+Sp_E>1$) (Jaffar *et al*, 2003,). Notwithstanding this, Jurek *et al* (2008) note that unless the classification errors are independent and exactly equal then bias away from the null can occur; thus the assumption that errors are approximately non-differential may not be predictive of bias toward the null. Thus assumptions about non-differential errors should be made only when it is logical that the conditions are met.

Table 12.2 Relationship between the number of correctly and incorrectly classified subjects by exposure status

True number	Incorrectly classified number
a_1	$a_1'=Se_E*a_1+(1-Sp_E)*a_0$
a_0	$a_0'=(1-Se_E)*a_1+Sp_E*a_0$
b_1	$b_1'=Se_E*b_1+(1-Sp_E)*b_0$
b_0	$b_0'=(1-Se_E)*b_1+Sp_E*b_0$

The impact of classification errors depend on their magnitude and the actual prevalence of the item (*ie* exposure or disease) being classified. Relatively small errors (10%–20%) can have sizable effects on relative risks. Nonetheless, Blair *et al* (2007) comment that some "exposure misclassification probably occurs in all studies". Thus, in judging the effects of misclassification the actual likelihood of that misclassification occurring and its magnitude should be considered. A numerical example of the impact of non-differential misclassification is shown in Example 12.3.

Whereas in cohort and cross-sectional studies, the assumption that any errors of exposure classification are non-differential may be logical and valid, in case-control studies, the assumption of non-differential errors is often open to question (see below).

12.6.2 Evaluating non-differential exposure misclassification

A few moments investigating small changes in the estimated sensitivity and specificity of exposure classification (based on Table 12.2) will convince you that they can produce large changes in the observed association. Indeed, the variability in the data arising from these small

Example 12.3 Impact of non-differential misclassification of exposure

In this (fictitious) example, we first assume that there is no misclassification, hence, the true study group structure in this example is:

	Exposed	Non-exposed	Total
Diseased	90	70	160
Non-diseased	210	630	840
Total	300	700	1000

The true *OR* is 3.86. If we now assumed an exposure sensitivity of 80% and an assumed exposure specificity of 90%, we would expect to have the following observed cell numbers (calculations shown):

	Exposed	Non-exposed	Total
Diseased	90*0.8+0.1*70=**79**	70*0.9+90*0.2=**81**	160
Non-diseased	210*0.8+630*0.1=**231**	630*0.9+210*0.2=**609**	840
Total	300* 0.8+700*0.1=**310**	700*0.9+300*0.2=**690**	1000

Note Exposure misclassification does not affect the disease status totals, only the exposure category totals. As predicted, with non-differential errors the odds ratio has been reduced from 3.86 to 2.57.

changes can be much more dramatic than changes that would be expected from sampling variation. Jurek *et al* (2006) stressed that quantitative methods are available to estimate the effect of, or correct for, these errors. Given that we often lack knowledge of the true Se_E and Sp_E values, we view this process more as evaluation than 'correction'. However, the quantification of potential effects provides valuable information that aids interpretation of study results.

As an introduction to this process (see Fox *et al* (2005)), if the most likely values of Se_E and Sp_E are known, we can correct the observed classifications for the errors. As we noted elsewhere, because we rarely know the true values of Se_E and Sp_E, we use this approach to evaluate the likely direction and magnitude of bias a range of reasonable estimates might produce, not necessarily to 'correct' for classification errors. Nonetheless, knowing the 'algebra' behind these methods should aid our understanding of the process. Assuming non-differential errors, we can use the following approach to reclassify the study group. Since $b_1'+b_0' = b_1+b_0 = m_0$, we can solve for the number of exposed controls b_1 as:

$$b_1 = \frac{b_1' - (1 - Sp_E) * m_0}{(Se_E + Sp_E - 1)} \qquad \textbf{Eq 12.3}$$

Similarly, we can solve for the number of exposed cases a_1 as:

$$a_1 = \frac{a_1' - (1 - Sp_E) * m_1}{(Se_E + Sp_E - 1)} \qquad \textbf{Eq12.4}$$

with b_0 and a_0 determined by $b_0=m_0-b_1$ and $a_0=m_1-a_1$. We now complete the 'adjusted' 2X2 table cell values and compute the estimate of the true *OR*. This process also can be used to assess the effect of differential errors in exposure status by repeating the process separately in each of the case and control groups using the appropriate estimates of Se_E and Sp_E .

Fox *et al* (2005) and Orsini *et al* (2008) have implemented this approach for evaluating and correcting misclassification errors in case-control studies with appropriate software code. Thus, we can 'plug-in' reasonable estimates of Se_E and Sp_E to ascertain the deterministic impact of classification errors. Example 12.4 shows an evaluation of the effect of misclassification of exposure using the data presented in Example 12.2.

In this process, if we obtain 'impossible' results; this means that the 'plug-in' values used are not consistent with the data, so the actual error risks must differ from the values being used for 'corrections'. In attempting to obtain better estimates of actual *Se* and *Sp* from our own validation, or external datasets, Lyles *et al* (2007) provide a test of 'transportability' which ascertains if the estimates of errors in different datasets are similar. They also provide a likelihood ratio test to ascertain if the errors should be considered to be differential.

In general, when exposure prevalence is low, lack of specificity produces more errors than lack of sensitivity. Walter (2007) notes that the attributable fraction is not biased if sensitivity is perfect; however, if perfect sensitivity is achieved at a cost of substantially reduced specificity then the precision of the attributable fraction estimate can be decreased. Frost and White (2005) describe methods for correcting errors in time-varying risk factors in longitudinal studies, and demonstrate that some frequently used methods do not work adequately in this context.

Example 12.4 Evaluating exposure misclassification

The original data (Nødtvedt *et al,* 2007) are shown in Example 12.2. As examples for evaluating potential misclassification bias, we chose the 3 different scenarios (ranges of errors and approaches to evaluation) shown in the table below; the impact of these errors on the observed *OR* is also included.

	Scenario 1 Deterministic with Non-differential errors	Scenario 2 Deterministic with Differential errors	Scenario 3 Stochastic with Differential errors
Se Cases	0.8	0.9	uniform(0.85-0.95)
Sp Cases	0.95	0.85	uniform(0.82-0.88)
Se Controls	0.8	0.8	uniform(0.7-0.9)
Sp Controls	0.95	0.95	uniform(0.92-0.98)
Observed *OR*	2.33	2.33	2.33
Adjusted *OR*	3.71	1.81	1.94 (median)

Scenario 1 The Se_E and Sp_E were assumed to be non-differential (*ie* equal in the case and control groups) and were assumed to be a single set of values. **Note** The adjusted (assumed closer to true) *OR* is larger than the observed *OR*. As expected, misclassification bias reduced the *OR*.

Scenario 2 The Se_E and Sp_E were assumed to be differential (Se_E higher in cases and Sp_E higher in controls) and were assumed to be a single set of values. **Note** Now the adjusted *OR* is closer to the null value than the observed value. Misclassification bias has resulted in a bias away from the null.

Scenario 3 The Se_E and Sp_E were assumed to be differential (Se_E higher in cases and Sp_E higher in controls) and but were now randomly selected from the uniform distributions shown. (A uniform distribution is one which any value within the range specified is equally likely.) The median value for the adjusted *OR* from 2,000 simulations was 1.94 (similar to the assumed to be a single set of values). **Note** The adjusted *OR* is now closer to the null value than the observed value. Misclassification bias has resulted in a bias away from the null, and 95% of the adjusted values fell in the range of 1.72 to 2.16. A distribution of estimated values is shown in Fig. 12.2

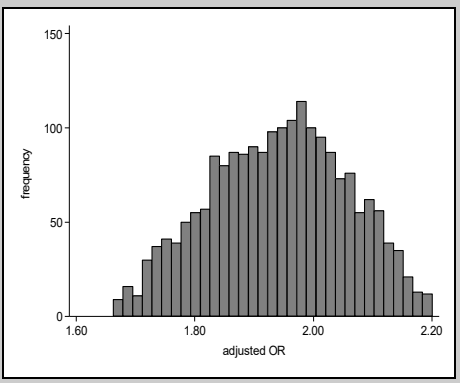

Fig. 12.2 Distribution of adjusted OR from stochastic analysis

As these scenarios demonstrate, misclassification can produce considerable bias. However, one needs to have reasonable estimates of the 'error rates' in order to assess the direction and extent of bias the errors produce.

12.6.3 Non-differential misclassification of disease-cohort studies

Here the same concepts of classification errors arise as with exposure misclassification except that we now focus on errors in classifying health status in cohort studies. With non-differential misclassification for disease we have

$$Se_{D|E+} = Se_{D|E-} = Se_D \qquad \text{and/or} \qquad Sp_{D|E+} = Sp_{D|E-} = Sp_D$$

where Se_D is the sensitivity of disease classification and Sp_D is the specificity of disease classification. There are 2 components to disease classification in cohort studies and they have different impacts on the association measure. First, we need to establish the health status of all study subjects at the initiation of follow-up in order to exclude prevalent cases, and second we need to identify the new cases of the outcome that develop during the study period.

With respect to establishing the initial health status, Pekkanen *et al* (2006) demonstrate that imperfect assessment of the disease status at the start of a cohort study can bias subsequent measures of association. Imperfect sensitivity fails to exclude subjects with the outcome at the study outset; imperfect specificity has less of an impact. The equations to estimate the impact of this bias are very complicated and have no simple arithmetic solution. However, Pekkanen *et al* (2006) showed that non-differential misclassification of disease at baseline can lead to over- or underestimation of the true incidence risk ratio, because the observed incidence risk ratio reflects both the association at baseline and at follow-up. This underscores the need to carefully exclude all prevalent diseased subjects from the study using a sensitive test for disease at the initiation of the study.

The impact of errors in the diagnosis of the outcome during follow-up is similar to the impact of exposure errors. For binary outcomes non-differential errors bias the association measure toward the null; the impact of differential errors in classifying the outcome are more difficult to predict. Adjusting for these errors is similar to the process discussed in Section 12.6.2 for exposure-related errors. Luan *et al* (2005) note that it is not always beneficial to adjust binary outcomes for misclassification because the increase in variance of the *OR* offsets the correction for bias.

12.6.4 Non-differential misclassification of disease case-control studies

Because of the often unknown *sf*s in case-control studies, the approach to correcting for diagnostic errors that are applicable in cohort studies do not apply to this case-control studies unless Sp_D=1.00. In that instance, imperfect disease sensitivity does not bias the *RR* or *IR*, and only biases the *OR* if disease frequency is common. The key here is that it pays to verify the diagnoses of the cases so that there are no false positive cases, as the association measures will not be biased even if the diagnostic Se_D is less than 100%.

When Sp_D <1, non-cases will be included in our case series. Hence, in a case-control study, if we take all the apparent cases for our study, we will be including $Se_D * M_1$ of the true cases and $(1-Sp_D) * M_0$ false positives as cases. Usually, we take only a fraction (*sf*) of the apparent non-cases as controls, hence ultimately, we will include a small number of false negative cases ($sf * (1-Sp_D) * M_1$) and a much larger number of true non-cases ($sf * Sp_D * M_0$). Thus, in the study group, the case-control sensitivity will be

$$Se_{cc} = Se_D / (Se_D + sf * (1 - Sp_D)) \qquad \textbf{Eq 12.5}$$

and the case-control specificity will be

$$Sp_{cc} = sf * Sp_D / ((1 - Sp_D) + sf) * Sp_D \qquad \textbf{Eq 12.6}$$

Both of these could be very far from the true population values of sensitivity and specificity.

Thus, external estimates of Se_D and Sp_D cannot be used to correct misclassification in case-control studies. Also, estimates of diagnostic Se_D and Sp_D obtained from case-control study subjects cannot be used to estimate the population Se_D and Sp_D values.

12.6.5 Misclassification of both exposure and disease

As noted earlier, if one works through many examples using realistic error rates, it becomes clear that misclassification bias can create much more uncertainty in our measures of association than sampling variation. Thus, we need to pay a great deal of attention to reducing these errors whenever possible. Although it is possible to conduct simultaneous adjustment for errors in exposure and outcome, in cohort or cross-sectional data, most researchers prefer to evaluate (what if?) for the more important errors or make the adjustments for one set of errors at a time.

12.6.6 Differential misclassification of exposure or outcome

If the errors in exposure classification are related to the status of the outcome under study, the errors are called **differential**. Here, the Se_E and Sp_E differ by disease status

$$Se_{E|D} \neq Se_{E|D-} \qquad \text{and/or} \qquad Sp_{E|D+} \neq Sp_{E|D-}$$

In a similar manner, for outcome classification, with differential errors, the Se_D and/or Sp_D of classifying disease status differs over exposure levels

$$Se_{D|E+} \neq Se_{D|E-} \qquad \text{and/or} \qquad Sp_{D|E+} \neq Sp_{D|E-}$$

The resulting bias in the measure of association might be in any direction (*eg* an association might either be exaggerated or underestimated). A few minutes with a spreadsheet playing 'what-if' will help convince you of this.

In case-control studies, **recall bias** is one illustration of (likely) differential errors in that 'affected' subjects (*ie* cases) might be expected to have an increased sensitivity, and perhaps a lower specificity than non-affected subjects in recalling previous exposures. We developed an example of this bias in Example 12.3. Chyou (2007) studied the impact of differential misclassification of exposure among cases and controls, and confirmed that differential errors make the direction of bias difficult to predict.

12.6.7 Reducing misclassification errors

The specific ways that can be used to reduce misclassification errors are highly context specific. Nonetheless, in general, the frequency of errors can be reduced by using clear and explicit guidelines, having well-trained consistent research personnel and 'double-checking' the exposure and disease status whenever possible. Seek confirmation of information whenever possible through laboratory results, or other confirmatory records of exposure or disease. Validating the test or survey instrument prior to its widespread use is certainly preferable (see Chapter 3 for some suggestions) to trying to correct for misclassification errors after the fact. As examples of approaches to minimise errors, it is important to collect specific rather than general exposure data as the latter often lead to attenuation of the true association between exposure and outcome (Friesen *et al*, 2007). When attempting to obtain specific exposure information (*eg* pesticide or antibiotic use) either ask detailed questions, or ask for bottle labels,

or have the participant identify the exposure item from a portfolio of pictures (Acquavella *et al*, 2006). Be aware that self-reported exposures may not correlate very well with objective measures of exposure (Radon *et al*, 2007), and don't make assumptions about exposures. For example, Jones *et al* (2006) found that household water supply was a poor indicator of drinking water source for subjects on private water systems.

In addition to reducing errors, because the results of non-differential misclassification generally are predictable, we often recommend 'blind' techniques for survey personnel to help ensure that the errors are equalised. This is a good general strategy, and can be applied to the perusal of case records, interview information *etc*.

12.6.8 Misclassification of extraneous variables

If a confounder is measured with error, it is impossible to fully control for its confounding effect. The bias can be large if the true effect of the exposure is weak and the confounder is strongly related to exposure and the outcome. In the face of misclassification of the confounder, it becomes difficult to know whether or not one should control for the confounder (see Chapter 13). A general recommendation is that the impact of controlling an extraneous variable should only be investigated when little misclassification of the confounder exists, or until after adjustments for the errors have been made. Berry *et al* (2005) demonstrate that using a badly misclassified confounder to control a bias can lead to incorrect conclusions. Similarly, Murad and Freedman (2007) used 'corrected' estimates of misclassified variables before examining for interaction. Clearly, one must focus on reducing measurement error in all variables, not just exposure and outcome, if valid analyses and inferences are to be made.

12.6.9 Misclassification of multinomial exposure or disease categories

With several levels of exposure, the effects of classification errors are less predictable than with dichotomous variables. Fosgate (2006) demonstrated that the likelihood ratio could be biased away from the null when categorising a continuous outcome into categories. A consistent finding was that non-differential error reduced the value (*eg* sensitivity or specificity) of the assessment tool at each level that was measured with error. Non-differential misclassification might bias measures of association in intermediate exposure levels away from the null, and might even reverse the direction of the *OR*s for these levels. This becomes an important issue when we use regression models because while these models allow for error in the measurement of the outcome they assume no error of measurement of the predictor variables. Non-differential underestimation of exposure at high levels might cause a threshold effect of exposure to appear as a dose-response relationship. Likewise, non-differential misclassification of both *E* and *D* status when the errors are **not independent** might lead to bias away from the null, particularly when the prevalence of both exposure and disease are low. Leeflang *et al* (2008) noted that data driven choices of cut-points often lead to overly optimistic assessments of error levels, but the bias tends to decrease with increasing sample size.

12.7 VALIDATION STUDIES TO CORRECT MISCLASSIFICATION

A thorough review of the use of validation studies to correct misclassification is given by Thurigen *et al* (2000) especially as they relate to case-control studies (see this paper for details).

The 4 main approaches reviewed are regression calibration, maximum likelihood, semi-parametric and Bayesian methods. One summary finding is that we need to be aware of the limitations in using 'simple' approaches to correct for misclassification, but unfortunately the more advanced methods are not user-friendly. Two-stage samples, mentioned in Chapter 11, are useful for validation purposes and this approach is also elaborated in Section 12.8. For validation, we select a subsample of study subjects and verify their exposure and/or disease status. Recall that, for direct estimates of sensitivity and specificity, we are determining the probability of the observed state (D'), given that we know the true state of the individual (D). That is:

$$p(D'=1|D=1)$$

whereas when correcting for misclassification, we are attempting to determine the probability of the true state, given knowledge of the observed state:

$$p(D=1|D'=1)$$

As noted previously, a major problem with post-hoc adjustments of misclassification is that they are very sensitive to changes in the estimates of the error rates used in the correction process. Thus, unless there is an extremely thorough validation procedure, the estimates of error might vary sufficiently such that very different 'corrected' results could arise from applying a range of apparently sensible choices of the correction factor. A few minutes trying 'what-if' adjustments should convince you of this. Lyles *et al* (2007) discuss correcting for misclassification using internal data (as in a 2-stage validation study) and also using data external to the study. The authors note that it is very important for the sensitivity and specificity of misclassification to be equivalent in the 2 datasets (transportable) before attempting to adjust for the errors. Validation to correct for measurement error is described in the next section.

12.8 MEASUREMENT ERROR

Errors in measuring quantitative factors can lead to biased measures of association and this fact seems to be ignored frequently (Jurek *et al*, 2006). The bias can arise either because the variable is not measured **accurately** (*ie* a systematic bias), or due to a lack of **precision** (see Section 5.2.2). In turn, lack of precision might arise from either variability in the test *per se*, or because the substance being measured varies within an individual (for physiological reasons) and consequently, repeated measures are needed to provide a valid overall indicator of the status of the individual (*eg* a mean of 2 or more samples).

Considerable work on the issue of measurement error and the general approach to correcting measurement bias has been published in recent years (Freedman *et al*, 2008, Guolo, 2007). To introduce the concepts of correcting these errors, let's suppose that we have 2 quantitative exposure factors and we wish to estimate their association with either a binary or continuous outcome. Allowing that the Y-variable could represent the logistic transform of a binary outcome, or a continuous outcome variable in a linear model, we could express the uncorrected 'naive' model as:

$$Y=\beta_{0u}+\beta_{1u}X_1'+\beta_{2u}X_2' \qquad \textit{Eq 12.7}$$

where the subscript 'u' indicates that the coefficients are biased because the predictor variables, here denoted as X', are measured with error. There is a variety of approaches to correcting for errors; one robust and relatively simple method is called the **regression calibration estimate**

(RCE). To obtain the RCE, we take a random subset of the study subjects and perform a validation study so that the true values for X_1 and X_2 are obtained. Now, assuming non-differential measurement errors, we regress each true X variable on the set of observed predictor variables. That is:

$$X_1 = \beta_0 + \lambda_{11} X_1' + \lambda_{12} X_2' \qquad \textit{Eq 12.8}$$

$$X_2 = \beta_0 + \lambda_{21} X_1' + \lambda_{22} X_2' \qquad \textit{Eq 12.9}$$

Then, we calculate the estimated (*ie* the predicted) X values for all the study subjects, denoted here as X_{1rc} and X_{2rc} using the coefficients from these equations. Then, we regress Y on these estimated values.

$$Y = \beta_{0rc} + \beta_{1rc} X_{1rc} + \beta_{2rc} X_{2rc} \qquad \textit{Eq 12.10}$$

The coefficients β_{1rc} should provide less biased estimates of the true X-Y association than the naive estimates. The standard errors need to be adjusted for the calibration process and are explained in Freedman *et al* (2008), and implemented in Hardin *et al* (2003). The above approach has a crucial assumption; namely non-differential measurement errors. If differential errors are suspected, the approach need to be modified (Freedman *et al*, 2008). The regression models chosen for the X variables depend on the assumed distribution of the X variables (*ie* continuous or binary), and the validity of the approach to correcting measurement errors, in part, depends on the fit of the above models. Murad and Freedman (2007) apply regression calibration to correcting measurement error before examining interactions in linear models. Wang *et al* (2008) describe methods to adjust for missing data, measurement error and misclassification in longitudinal studies.

12.9 ERRORS IN SURROGATE MEASURES OF EXPOSURE

Often, epidemiologists focus on the effects of a complex exposure factor. For example, in studies of the impact of air pollution from oil and gas-processing emissions on cattle or wildlife health, what is the appropriate measure of air pollution? In this, and other examples, the exposure might be a complex mixture of agents (or factors), doses and duration, and it will take considerable thought as to what components of exposure to measure and which to ignore (Waldner, 2008). For example, which of the hundreds of compounds in air pollution does one measure? The most abundant, the least expensive to monitor, or the most toxic? If a number of agents are measured, how will they be modelled? The answers to these questions (yes, there undoubtedly will be more than one correct answer) will largely involve knowing context-specific biological background information.

The decisions about surrogate measures must then be translated into what will be measured, and how the various axes of exposure will be analysed in order to achieve the study objectives. For example, will the exposure be measured and analysed on a continuous scale (the preferred option) or will it be categorised into a dichotomous or ordinal exposure variable? If levels of specific agents are highly correlated, which one should be analysed, or should a composite variable be created? Although categorising continuous data is not the preferred choice, it might reflect the reality of the exposure measurements better than the more refined measures. For example, if most levels of exposure are at or near the laboratory sensitivity of the test procedure, it might be best to dichotomise into non-exposed (for most of the data) and exposed for the limited number of measurements that are clearly above accepted levels of exposure. Of

course the measured factors, being surrogates, might still fail to reflect the actual exposure. Thus, even if the variables measured are, in fact, measured without error, we need to be aware that because the variables are surrogates, we could still be left with measurement error in respect of the true exposure.

One solution might be to change the questions asked. Instead of asking about the effects of 'air pollution', ask about the effects of the measurable component(s) (*eg* sulphur dioxide, then factors such as H_2S or particulates would be extraneous variables). More focused questions still require the measurement and control of other factors that might confound or interact with the exposure but the more focused answers might allow better progress toward solving the issue(s).

12.10 THE IMPACT OF INFORMATION BIAS ON SAMPLE SIZE

It is apparent that classification and measurement errors can have a serious impact on the measures of association. With non-differential misclassification of categorical variables, the measures are biased toward the null. And, under classical measurement error models, the same is true for continuous variables. This has led some to conclude that in planning a study, the projected loss of power due to these errors should be considered and the sample size increased accordingly (Devine, 2003). The formulae used in Chapter 2 for sample size estimates assumed that the p_1 and p_2 were true population levels. However, because the outcomes might be measured with an imperfect test, survey question, or diagnostic procedure the observed disease frequencies would be as follows:

$$p_1' = Se\, p_1 + (1 - Sp)(1 - p_1) \quad \text{and} \quad p_2' = Se\, p_2 + (1 - Sp)(1 - p_2)$$

The difference p_1' - p_2' is usually less than the difference p_1 - p_2 and it is the adjusted estimates (and their variances) that should be used to estimate sample size to account for the misclassification. Some care is needed, however, because, if we are using the observed outcome levels from previous studies where outcomes were measured with error, these would already represent p_1' and p_2' and need not be adjusted further. Obuchowski (2008) generalises sample size estimation to account for misclassification, response bias and other features of clinical trials with emphasis on evaluating screening programs. Huzurbazar *et al* (2004) used adjustment for the costs of misclassification errors in herds of cattle when screening for bovine virus diarrhea infection.

REFERENCES

Acquavella JF, Alexander BH, Mandel JS, Burns CJ, Gustin C. Exposure misclassification in studies of agricultural pesticides: Insights from biomonitoring. Epidemiology 2006;17:69-74.

Alonso A, de Irala J, Martinez-Gonzalez MA. Representativeness, losses to follow-up and validity in cohort studies. Eur J Epidemiol 2007;22:481-482.

Alonso A, Segui-Gomez M, de Irala J, Sanchez-Villegas A, Beunza JJ, Martinez-Gonzalez MA. Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. Eur J Epidemiol 2006;21:351-358.

Beck CA. Selection bias in observational studies: Out of control? Neurology 2009;72:108-109.

Berger VW. The reverse propensity score to detect selection bias and correct for baseline imbalances. Stat Med 2005;24:2777-2787.

Berry RJ, Kihlberg R, Devine O. Impact of misclassification of in vitro fertilisation in studies of folic acid and twinning: Modelling using population based Swedish vital records. Br Med J 2005;330:815.

Bjertnaes OA, Garratt A, Botten G. Nonresponse bias and cost-effectiveness in a Norwegian survey of family physicians. Eval Health Prof 2008;31:65-80.

Blair A, Stewart P, Lubin JH, Forastiere F. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med 2007;50:199-207.

Braback L, Hjern A, Rasmussen F. Trends in asthma, allergic rhinitis and eczema among Swedish conscripts from farming and non-farming environments. A nationwide study over three decades. Clin Exp Allergy 2004;34:38-43.

Braback L, Hjern A, Rasmussen F. Selective migration contributes to a healthy worker effect in the farming population. J Clin Epidemiol. 2006;59(1):102-3; author reply 3.

Chyou PH. Patterns of bias due to differential misclassification by case-control status in a case-control study. Eur J Epidemiol 2007;22:7-17.

Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. Int J Epidemiol 2006;35:1074-1081.

Devine O. The impact of ignoring measurement error when estimating sample size for epidemiologic studies. Eval Health Prof. 2003;26:315-339.

Ducrot C, Calavas D, Sabatier P, Faye B. Qualitative interaction between the observer and the observed in veterinary epidemiology Prev Vet Med 1998;34:107-113.

Ducrot C, Roy P, Morignat E, Baron T, Calavas D. How the surveillance system may bias the results of analytical epidemiological studies on BSE: Prevalence among dairy versus beef suckler cattle breeds in France. Vet Res 2003;34:185-192.

Elwood M. Forward projection—using critical appraisal in the design of studies. Int J Epidemiol 2002;31:1071-1073.

Fosgate GT. Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null. Emerg Themes Epidemiol 2006;3:7.

Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34:1370-1376.

Fraser G, Yan R. Guided multiple imputation of missing data: Using a subsample to strengthen the missing-at-random assumption. Epidemiology 2007;18:246-252.

Fraser GE, Yan R, Butler TL, Jaceldo-Siegl K, Beeson WL, Chan J. Missing data in a long food frequency questionnaire: Are imputed zeroes correct? Epidemiology 2009;20:289-294.

Freedman LS, Midthune D, Carroll RJ, Kipnis V. A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in

regression. Stat Med 2008;27:5195-5216.

Friesen MC, Davies HW, Teschke K, Ostry AS, Hertzman C, Demers PA. Impact of the specificity of the exposure metric on exposure-response relationships. Epidemiology 2007;18:88-94.

Frost C, White IR. The effect of measurement error in risk factors that change over time in cohort studies: Do simple methods overcorrect for 'regression dilution'? Int J Epidemiol 2005;34:1359-1368.

Geneletti S, Richardson S, Best N. Adjusting for selection bias in retrospective, case-control studies. Biostatistics 2008.

Geneletti S, Richardson S, Best N. Adjusting for selection bias in retrospective, case-control studies. Biostatistics. 2009;10(1):17-31.

Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. J Chronic Dis 1981;34:433-438.

Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet. 2002;359:248-252.

Guolo A. Robust techniques for measurement error correction: A review. Stat Methods Med Res 2008;17:555-580.

Hardin JW, Schmiediche H, Carroll RJ. The regression-calibration method for fitting generalized linear models with additive measurement error. Stata J 2003;4:361-372.

Harris AD, Carmeli Y, Samore MH, Kaye KS, Perencevich E. Impact of severity of illness bias and control group misclassification bias in case-control studies of antimicrobial-resistant organisms. Infect Control Hosp Epidemiol. 2005;26:342-345.

Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15:615-625.

Huzurbazar S, Van Campen H, McLean MB. Sample size calculations for Bayesian prediction of bovine viral-diarrhoea-virus infection in beef herds. Prev Vet Med 2004;62:217-232.

Jaffar S, Leach A, Smith PG, Cutts F, Greenwood B. Effects of misclassification of causes of death on the power of a trial to assess the efficacy of a pneumococcal conjugate vaccine in the Gambia. Int J Epidemiol 2003;32:430-436.

Jones AQ, Dewey CE, Dore K, Majowicz SE, McEwen SA, Waltner-Toews D. Exposure assessment in investigations of waterborne illness: A quantitative estimate of measurement error. Epidemiol Perspect Innov. 2006;3:6.

Jurek AM, Greenland S, Maldonado G. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? Int J Epidemiol 2008;37:382-385.

Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. Eur J Epidemiol 2006;21:871-876.

Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity

caused by data-driven selection of optimal cutoff values: Mechanisms, magnitude, and solutions. Clin Chem. 2008;54:729-737.

Luan X, Pan W, Gerberich SG, Carlin BP. Does it always help to adjust for misclassification of a binary outcome in logistic regression? Stat Med 2005;24:2221-2234.

Lyles RH, Zhang F, Drews-Botsch C. Combining internal and external validation data to correct for exposure misclassification: A case study. Epidemiology 2007;18:321-328.

Mezei G, Kheifets L. Selection bias and its implications for case-control studies: A case study of magnetic field exposure and childhood leukaemia. Int J Epidemiol 2006;35:397-406.

Miller LG, Tang AW. Treatment of uncomplicated urinary tract infections in an era of increasing antimicrobial resistance. Mayo Clin Proc. 2004;79:1048-53; quiz 1053-4.

Morabia A. Case-control studies in clinical research: Mechanism and prevention of selection bias. Prev Med. 1997;26:674-677.

Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: A survey of practice. Am J Epidemiol. 2006;163:197-203.

Murad H, Freedman LS. Estimating and testing interactions in linear regression models when explanatory variables are subject to classical measurement error. Stat Med 2007;26:4293-4310.

Nødtvedt A, Bergvall K, Sallander M, Egenvall A, Emanuelson U, Hedhammar A. A case-control study of risk factors for canine atopic dermatitis among boxer, bullterrier and West Highland white terrier dogs in Sweden. Vet Dermatol. 2007;18(5):309-15.

Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? Epidemiology 2006;17:413-418.

Obuchowski NA. Estimating sample size for a randomized clinical trial of lung cancer screening. Contemp Clin Trials. 2008;29:466-477.

Orsini N, Belleco R, Bottai M, Wolk A, Greenland S. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. Stata J 2008;8:29-48.

Pekkanen J, Sunyer J, Chinn S. Nondifferential disease misclassification may bias incidence risk ratios away from the null. J Clin Epidemiol. 2006;59:281-289.

Radon K, Schulze A, Ehrenstein V, van Strien RT, Praml G, Nowak D. Environmental exposure to confined animal feeding operations and respiratory health of neighboring residents. Epidemiology 2007;18:300-308.

Robinson KA, Dennison CR, Wayman DM, Pronovost PJ, Needham DM. Systematic review identifies number of strategies important for retaining study participants. J Clin Epidemiol. 2007;60:757-765.

Sadetzki S, Bensal D, Novikov I, Modan B. The limitations of using hospital controls in cancer etiology—one more example for Berkson's bias. Eur J Epidemiol 2003;18:1127-1131.

Sandler DP. On revealing what we'd rather hide: The problem of describing study participation. Epidemiology 2002;13:117.

Schwartzbaum J, Ahlbom A, Feychting M. Berkson's bias reviewed. Eur J Epidemiol 2003;18:1109-1112.

Shahar E. Causal diagrams for encoding and evaluation of information bias. J Eval Clin Pract 2009;15:436-440.

Singer RS, Atwill ER, Carpenter TE, Jeffrey JS, Johnson WO, Hirsh DC. Selection bias in epidemiological studies of infectious disease using *escherichia coli* and avian cellulitis as an example. Epidemiol Infect 2001;126:139-145.

Sjolander A, Humphreys K, Palmgren J. On informative detection bias in screening studies. Stat Med 2008;27:2635-2650.

Stang A. Nonresponse research—an underdeveloped field in Epidemiology Eur J Epidemiol 2003;18:929-931.

Tam CC, Rodrigues LC, O'Brien SJ. The study of infectious intestinal disease in England: What risk factors for presentation to general practice tell us about potential for selection bias in case-control studies of reported cases of diarrhoea. Int J Epidemiol 2003;32:99-105.

Thurigen D, Spiegelman D, Blettner M, Heuer C, Brenner H. Measurement error correction using validation data: A review of methods and their applicability in case-control studies. Stat Methods Med Res 2000;9:447-474.

Tongue SC, Pfeiffer DU, Warner R, Elliott H, Del Rio Vilas V. Estimation of the relative risk of developing clinical scrapie: The role of prion protein (PrP) genotype and selection bias. Vet Rec 2006;158:43-50.

Waldner CL. Western Canada study of animal health effects associated with exposure to emissions from oil and natural gas field facilities. study design and data collection III. methods of assessing animal exposure to contaminants from the oil and gas industry. Arch Environ Occup Health 2008;63:201-219.

Walter SD, Hsieh CC, Liu Q. Effect of exposure misclassification on the mean squared error of population attributable risk and prevented fraction estimates. Stat Med 2007;26:4833-4842.

Wang CY, Huang Y, Chao EC, Jeffcoat MK. Expected estimating equations for missing data, measurement error, and misclassification, with application to longitudinal nonignorable missing data. Biometrics 2008;64:85-95.