13

CONFOUNDING: DETECTION AND CONTROL

OBJECTIVES

After reading this chapter, you should be able to:

1. Apply a working set of criteria to identify potential confounders in an observational study.
2. Use restricted sampling to prevent confounding.
3. Determine appropriate variables for control of confounding using matching and implement the matching process in a cohort study.
4. Determine appropriate variables for control of confounding using matching and implement the matching process in a case-control study.
5. Use matching based on propensity scores in a cohort study.
6. Implement a valid plan for the control of confounding using analytic procedures.
7. Use a causal diagram to identify factors (confounders) needing control.
8. Apply a stratified analysis to a set of categorical variables to evaluate the presence of interaction and assess the extent of confounding, while estimating causal effects.
9. Apply a stratified analysis based on propensity scores.
10. Understand the link between inverse probability of treatment weighting (IPTW) and standardised risk ratios (SRRs) and the manner in which they estimate the causal effect.
11. Evaluate the potential of a non-measured confounder to bias the outcome measure using sensitivity analysis.
12. Interpret the likely effect of ‘controlling’ extraneous factors having specified their causal associations with the outcome and exposure.

13.1 INTRODUCTION

A central focus of epidemiologic research is to identify factors (*ie* causes) that contribute to the occurrence of disease, poor productivity or lower animal welfare status under 'real-world' conditions. In Chapter 1, we noted that it is generally agreed that a randomised controlled trial (RCT) is the best way to evaluate the effect of these factors (these 'factors' usually are referred to as **treatments** in the jargon of field experiments, whereas, in observational studies they are referred to as **exposures**). The use of randomisation in experiments provides a probabilistic basis for the balancing of factors, known and unknown, between the exposed and non-exposed groups. It prevents confounding and makes the groups 'exchangeable', in the sense that it does not matter which group gets assigned to receive the treatment. Thus, an ideal experiment would allow us to contrast the true frequency of outcome in the exposed (R_1) and non-exposed (R_0) subjects and closely approximate the true causal effect in an unbiased manner. Although it is not always feasible, ethical, or desirable, to randomly assign study subjects to receive or not receive an exposure, often observational data may be available to support a comparison of outcomes in exposed and non-exposed subjects. However, a difficulty in drawing causal inferences from these data is that exposed subjects are likely to differ from non-exposed subjects with respect to factors that can influence both whether or not the subject is exposed and the risk of the outcome. These factors bias (or confound) our observed measure of association. Put another way, the study groups being compared may differ in the frequency of the outcome for reasons other than the exposure of interest. Our challenge is to identify the factors that 'cause' this difference and prevent them from producing a biased result. This chapter is intended to help researchers using observational studies prevent confounding and obtain valid estimates of causal effects. As stressed in the earlier chapters on study design, it is also necessary to improve the description of our approaches to control confounding when reporting our findings so that others may learn from, and assess, our efforts (Groenwold *et al*, 2008; Klein-Geltink *et al*, 2007) .

Confounding can be described as the mixing together of the effects of 2 or more factors. Thus, when confounding is present we might think we are measuring the association between an exposure factor and an outcome, but the observed association measure also includes the effects of one or more extraneous factors. Hence, the measure of association is **biased**, or **confounded**. For our purposes of explaining confounding, we will assume that we have identified one factor as the main **exposure** of interest; this is our general preference in terms of research strategy and study design. One or more other factors that are of interest will be included in the study because they might help explain the frequency, or level, of the outcome—these will be denoted as **extraneous factors**. Some of the extraneous factors can have an association with both the exposure and outcome of interest, and failing to 'control' or 'adjust for' these relationships can produce a biased measure of association between the exposure factor of interest and the outcome. The extraneous factors that produce the bias are called **confounders** or **confounding factors**. Example 13.1 demonstrates confounding of an association.

13.1.1 Which extraneous factors are confounders?

Confounders might be defined based on their having distributional differences between study groups. This is a necessary but insufficient criterion of confounding. In addition, it is difficult to implement because we rarely know the true state, and the data from our study groups could

Example 13.1 A demonstration of confounding

We will begin by using a fictitious example with *Mannheimia haemolytica* (Mh) as the exposure of interest and bovine respiratory syncytial virus (BRSV) as the extraneous factor that we wish to control. The outcome is bovine respiratory disease (BRD), and the context is respiratory disease in cattle feedlots. We will assume BRSV, whose distribution we plan to 'control', is a confounder in the population. In our example, BRSV fulfills the criteria of being a confounder variable because it is associated with the exposure and the outcome, it is not intermediate between Mh and BRD on a causal pathway, and it is not an effect of BRD. Our summary of the fictitious population structure, ignoring BRSV status, is shown below:

	Mh+	Mh-	Totals	OR
BRD +	240	40	280	3.3
BRD -	6260	3460	9720	
Total	6500	3500	10000	
Risk (%)	3.7	1.1		

Based on observing the risk of BRD by Mh status and ignoring sampling variation, it appears that individuals with an active Mh infection (Mh+) have 3.3 times higher odds (think of this as 'risk') of contracting BRD than Mh- individuals (this assumes that 1.1% of the Mh+ individuals would have developed BRD in the absence of Mh—an assumed argument about exchangeability). But what about the effect(s) of BRSV? If BRSV is a confounder, then some of the crude association attributed to Mh might be due to BRSV.

Historically, one commonly used way to 'control' confounding is to stratify the data according to the levels of the confounding variable(s), or their combinations. Assuming that there are no other confounders, when the data are stratified on BRSV status, the 'true' association between Mh and BRD becomes apparent within strata. In this instance, it appears that Mh exposure doubles the risk of BRD in feedlot calves.

Population structure		Mh			Stratum-specific ORs	Crude OR
BRSV	**BRD**	**1**	**0**			
1	1	220	10	230	2	
1	0	5280	490	5770		
		5500	500	6000		
	Risks	0.04	0.02			3.3
0	1	20	30	50	2	
0	0	980	2970	3950		
		1000	3000	4000		
	Risks	0.02	0.01			

Note Ignoring issues of non-collapsibility of *OR*s (see Section 13.5.2), the crude *OR* differs from the stratum-specific *OR*s (by more than 30%), indicating that confounding is present so we need to use the stratum-specific *OR*s to estimate the causal association of Mh with BRD.

themselves be confounded. Nonetheless, based on a working set of criteria, we could conclude that a factor is a confounder if:

1. it is a cause of the disease, or a surrogate for a cause, and
2. it precedes and is associated with the exposure in the source population. In a cohort study, this means that the confounding factor must be associated with the exposure at the start of the study. In a case-control study, it means that the confounding factor must be associated with exposure in the population from whence the cases came (*ie* it must be associated with the exposure status in the control group), and
3. its distribution across exposure levels cannot be determined by the exposure (*ie* it is not an intervening factor) or by the disease (*ie* it is not a result of the disease). We stress that an **intervening factor** (also **intermediate factor**) should not be treated as a confounding factor, whether it is totally determined by the exposure or not, because this would modify (bias) the association between the exposure and the disease such that the true causal effect is not obtained. Similarly, if the disease produces an outcome such as another disease or change in production, that outcome should not be deemed to be a confounding factor.

It is useful to differentiate between a **population confounder** and a **sample** (*ie* study group) **confounder.** For example, if the factor is known (or regularly reported) to be a confounder in the target population, it should be treated as such in the sample (*ie* controlled) regardless of whether it appears to be a confounder in the sample or not. Conversely, if it is known not to be a population confounding factor, then it should not be controlled in the sample, even though it appears to be a confounder in the study subjects. Often, we do not know the true state of nature so we must use the data from the study group, or knowledge of the likely causal structure (Section 13.5.1) to make inferences about whether or not a factor is a confounder.

A statistical approach to defining confounding variables is based on the difference(s) in the distribution of the confounding factor(s) between the groups being studied. More precisely, if we have an exposure factor E, an outcome Y, and an extraneous factor Z (that is not an intervening variable or an effect of the outcome), factor Z is a confounder in a cohort study if:

- Z and E are associated unconditionally, and
- Z and Y are associated in exposure negative animals.

In a case-control study, factor Z is a confounder if:

- Z and E are associated in the controls (not just unconditionally), and
- Z and Y are associated in exposure negative animals.

Although these statistical criteria help us understand the necessary basis for confounding, these statistical criteria are insufficient to determine confounding without some additional assumptions about the lack of other confounders. Hence, we do not use statistical criteria to determine if a factor is a confounder or not. Rather, confounding is said to be present when our measure of association differs from the true value. Since the true value is usually unknown, the measure of association obtained after control of all identifiable potential confounders is deemed to be the best estimate of the **true causal association**. Usually we would say there is confounding when there is a noteworthy difference between the crude and adjusted (after control of the confounders) measures of association. If there is only a small difference, the crude measure will suffice. Because the identification and control of confounders in observational studies is rarely perfect, some confounding is invariably present, thus the important issue is how large the confounding effect is, not whether or not it is present. This becomes a matter of judgement (see Section 13.5.2) which we will elaborate subsequently.

13.2 CONTROL OF CONFOUNDING PRIOR TO DATA ANALYSIS

As noted here, and in the chapters on observational study design, we can prevent or control confounding from the extraneous factors that we can identify, or measure, by using one or more of 3 general procedures: **exclusion** (restricted sampling) (Section 13.2.1), **matching** (Sections 13.3 and 13.4), or **analytic control** (Section 13.6) (Mamdani *et al*, 2005; Normand *et al*, 2005). The use of these methods to control confounding can be traced directly, or indirectly, back to the idea of defining causal effects based on counterfactual outcomes. In order to obtain valid estimates of the causal effect, the groups being compared must be 'balanced' with respect to all factors that could bias the observed association between the exposure and outcome. Thus, exclusion and matching can be used to accomplish this prior to data analysis. The third approach includes a number of ways of statistically (*ie* **analytically**) balancing the groups in order to develop measures of association that are adjusted for any differences in the distribution of confounders. We would stress that all approaches rely on an implicit assumption of no residual confounding given the identified, or measured, confounders—an assumption that cannot be validated using the observed data but rather must rely on extant knowledge about the biology and context of the issue being studied.

13.2.1 Exclusion (restricted sampling)

Because confounding is the result of the differential distribution of an extraneous factor between the 2 (or more) groups being compared, a simple way to prevent confounding is by excluding subjects except those who possess only one defined level of the extraneous factor(s) for our studies. This is called **exclusion** or **restricted sampling**, and because every study subject has the same level of the potential confounder, no bias is present. Exclusion is a frequently used technique when selecting study subjects. Some restricted sampling is natural; for example, we would exclude males and only select females for a study of mastitis. In other instances, we might deliberately want to restrict our study population to a single breed of study subjects, or farms that use a specific production-recording scheme (Olde Riekerink *et al*, 2008). The former would prevent confounding by breed, whereas the latter could prevent confounding from differences in herd characteristics across recording schemes as well as help ensure that all required data would be easily available. Similarly, we could restrict our study population to those possessing a limited range of production, disease status, or being between specified ages *etc*. As examples of restricted sampling that produced several advantages for study quality by controlling confounding and in some cases enhancing data quality:

- Manske *et al* (2002) prior to a field trial of the effects of hoof-trimming on claw health in dairy cattle, restricted their study group of herds to selected herd sizes, breed compositions, and membership in an official milk-recording scheme.
- Cramer *et al* (2008) in a study of foot problems in dairy cows, restricted the study herds to those recruited by 5 professional hoof trimmers to minimise confounding by hoof trimmer.

When considering the use of restricted sampling based on dichotomous extraneous variables we would usually prefer to admit the low-risk group to the study. Even in the absence of confounding, admitting subjects only from the high-risk group could make data interpretation more difficult if interaction between the exposure and potential confounder were present. Thus, as an example, we normally would prefer to select cows without mastitis (and exclude those

with mastitis) in a study on the impact of foot lesions on milk production in case having both conditions produced biological synergism (and hence interaction (Section 13.6.2)).

13.3 MATCHING ON CONFOUNDERS

Matching is the process whereby we make the distribution of the extraneous factor(s) in the groups being compared the same. By making the distributions of these factors the same in both groups, we prevent confounding and in some instances increase the power of the study. We will discuss 2 broad approaches to matching, one focuses on matching for the variables of concern directly, both individually and collectively at the time of study subject selection. The second approach matches on a multivariable summary of the confounders, called a **propensity score,** prior to outcome data analysis. We will first focus on selecting and matching for confounders directly.

In randomised trials, matching on selected variables prior to randomisation of the treatment (also called **blocking**) is used to reduce the residual variance and thus give the study more power per study subject. Usually, matching is not used for prevention of bias, although in experiments with few subjects, it might help achieve this because randomisation is not likely to balance all the extraneous variables when the sample size is limited. As an example, in a field trial of hoof-trimming and claw health in dairy cows, Manske *et al* (2002) 'blocked' on breed, parity and stage of lactation before allocating, randomly, the treatment (hoof-trimming) to each cow.

In cohort and cross-sectional studies, matching on one or more confounding variables can prevent confounding bias and also result in increased power/precision of the study. Matching on host characteristics such as age, breed and sex is used frequently (because these variables often are strong confounders). As examples, Glickman *et al* (2009) matched on age when studying the progression of periodontal disease in dogs, and Bicalbo *et al* (2008) matched on parity, calving date and lactation status in a study of the impact of lameness on milk production. An example of the effects of matching in a cohort study is shown in Example 13.2.

Although some gains in power can result from matching, in observational studies, any gains in statistical efficiency come at a substantial cost. Most importantly,

- in case-control studies, it is not possible to estimate the effect of the matched factor(s) on the outcome because its distribution has been forced to be identical in the outcome groups. We can, however, investigate whether the matching factor acts as an effect modifier (*ie* if it produces interaction with the exposure of interest).
- matching by some global (*ie* very general) surrogate factors, such as farm, might 'match out' other potentially important exposures in hypothesis-generating studies.
- if matching is to be conducted on several factors, it can be quite difficult to find controls that have the same distribution of matching factors.

Matching is used frequently in case-control studies to increase the validity and efficiency of the study. As examples, in a study of avilamycin resistance in poultry bacteria, Chauvin *et al*, (2005), matched case and control broilers by the slaughterhouse, the time of sampling and the production type. McCarthy *et al*, (2004) used temporal matching in a study of equine grass sickness to prevent confounding by season, and Pinchbeck *et al*, (2004) matched on race type and jump number in a study to identify risk factors for falls in horses. However, matching in case-control studies has some potential disadvantages. For example, matching will actually

Example 13.2 Matching in a cohort study

In our 'pretend' cohort study, we will sample 500 exposed (Mh+) and 500 non-exposed (Mh-) individuals with frequency matching of the Mh- group for the distribution of the confounder (BRSV) in the exposed study group. Based on the population structure in Example 13.1, among the 500 Mh+ subjects, 85% (*ie* 5500/6500) of the Mh+ group will be BRSV+, and their risk of disease will be 4%. So, ignoring sampling variation, 17 of the 425 Mh+ and BRSV+ individuals in our study will develop BRD. Of the 75 Mh+ individuals without BRSV, 2% or 2 will develop BRD (expected numbers have been rounded to the nearest whole number).

Now, we need to select the Mh- subjects to match their distribution of BRSV to that in the Mh+ group. Normally, 14% (500/3500) of the 500 Mh- subjects would be BRSV+, but we need to have 85% (425) of them BRSV+. So after determining the BRSV status of the Mh- cattle, we select them to achieve this level of BRSV+ calves. Of the 425 BRSV+ Mh- calves, 2% develop BRD. Of the 75 Mh- calves who are BRSV-, 1% or 1 develops the disease. The numbers of matched Mh- subjects are italicised in the table below.

Note The observed stratum-specific odds ratios are equal to 2 (except for rounding errors), the same as in the source population (Example 13.1), as is the overall odds ratio. No control of the matched confounder is necessary in the analysis, and there is no bias present in the summary table. However, matched cohort data should be analysed using a stratified approach to ensure that the variance estimates of the adjusted odds ratio are correct.

Observed association between Mh and BRD in a cohort study following matching for BRSV

		Mh			Stratum-specific ORs	Crude OR
BRSV	**BRD**	**1**	**0**			
1	1	17	*9*	26	2	
1	0	408	*416*	824		
		425	*425*	850		2
0	1	2	*1*	3	2	
0	0	73	*74*	147		
		75	*75*	150		

In contrast to these results, in the next example, we pretend to conduct a case-control study using all 280 cases and 280 controls frequency matched by the confounder BRSV (see Example 13.3).

introduce a selection bias into the data. The stronger the exposure-confounder association in the source population, the greater the bias that is introduced. This bias is generally in the direction of the null effect, regardless of the direction of the exposure-confounder association, and must be controlled by carrying out an appropriate matched data analysis (see Section 13.3.4 for stratified matched analysis).

Why does matching have different effects in case-control studies than in cohort studies? In a cohort study, matching makes the exposure independent of the matched extraneous variable so there can be no confounding. The matched variable(s) can exert an effect on the outcome but it has the same effect in both exposure groups. Further, because the outcome (*eg* disease) has not

happened at the time of matching, the matching process is independent of the outcome. In contrast, in case-control studies, the disease has already occurred when the matching takes place. Hence, if the exposure is related to the matched variable (as it would be if the extraneous variable is a confounder), and if we make the distribution of the matched variable(s) the same in cases and controls, we will alter the distribution of exposure in the controls so that their exposure level is more like that in the cases and less like that in the source population. An example of this selection bias in a case-control study is presented in Example 13.3. This example also shows that we can prevent this type of selection bias, by stratifying on the matched variable(s).

13.3.1 General guidelines for matching

The following guidelines should be considered when contemplating the use of matching (Rothman *et al*, 2008). First, do not match unless you are certain that the variable is a confounder. This is particularly important in case-control studies—especially so if the extraneous variable and exposure are strongly associated. Matching in this situation leads to **overmatching**, because it gives the distribution of the exposure in the cases and controls greater similarity than the corresponding distributions in the source population. This occurs even if the extraneous variable is only related to the exposure and therefore not a confounder in

Example 13.3 Matching in a case-control study

In our case-control study, we will include all 280 cases from the source population in Example 13.1 as study subjects. This group will have the exposure and confounder distribution shown in Example 13.1. Now, after determining the BRSV status of the non-cases, we need to select the controls to match the distribution of BRSV in the cases. In this regard, we note that 82% (*ie* 230/280) of the cases will be BRSV+, so 230 of the controls will need to be BRSV+. Of these 230, 91.5% (5280/5770) will be Mh+ (n=210). Of the 50 BRSV- controls, 24.8% (980/3950) will be Mh+ (n=12). The numbers of matched controls are italicised in the table below.

Observed association between Mh and BRD in a case control study following matching for BRSV

Case-control structure		Mh			Stratum-specific ORs	Crude OR
BRSV	**BRD**	**1**	**0**			
1	1	220	10	230	2.1	
1	0	*210*	*20*	*230*		
						1.6
0	1	20	30	50	2.1	
0	0	*12*	*38*	*50*		

Note The stratum-specific *OR*s are equal to 2 (except for rounding error) but the crude *OR* is 1.6. This bias, induced by matching in a case-control study, is a form of selection bias. For example, in the population p(Mh+|BRD+)=86% (240/280) and p(Mh+|BRD-)=64% (6260/9720). In our study population, p(Mh+|BRD+)=86%, as it should, but p(Mh+|BRD-)=79%. The controls no longer represent the level of exposure in the source population. Clearly, analytical control (*eg* stratified analysis) of the matched confounder is necessary to prevent this selection bias in the overall measure of association.

the source population. In addition, with pair-matching (see Section 13.3.4), information will be lost because cases and controls with the same value of the exposure variable do not contribute useful data to the analysis, hence, effectively reducing the sample size and decreasing precision.

In some situations, however, matching will increase the efficiency of an analysis. For example:

- matching ensures that the dataset contains a control for every case when the matched factor is rare, or if it is a nominal variable with many categories (*eg* farm, age *etc*). Random sampling in this instance might lead to marginal zeros and the data from such tables are of no value in the analysis.
- matching might optimise the amount of information obtained per subject, if exposure information is expensive to obtain.
- matching might be the easiest way to identify controls in a case-control study using a secondary base (*eg* matching on admission time by selecting the next non-case admitted, or listed in the registry). This is one of the most common uses of matching and if matching is used only for this convenience, and the frequency of exposure is constant throughout the study period, the matching often is ignored and an unmatched analysis of the data performed.

If matching is not needed for one of these reasons, only consider matching in a case-control study if you anticipate a strong association in the source population between the outcome and the confounder and a relatively weak association between the exposure and the confounder. In case-control studies, any gains in efficiency from matching are likely to be modest at best.

13.3.2 Frequency and pair matching

In applying frequency-matching to categorical variables, the overall frequency of the potential confounder(s) is made the same in the 2 outcome (case and control) or exposure (cohort) groups. In pair- or individual-matching, one or more (*eg m*) control(s) is individually matched, with respect to the confounder, to each case. Relative to frequency-matching, pair-matching requires a more complex analysis, is generally less efficient (statistically), and makes it difficult to assess interaction between the exposure and confounder. However, pair-matching might be the only alternative when categories are very refined. For example if we want to match on age, gender and breed in a case-control study of equine lameness, we will have to identify an individual non-lame horse of the same age, gender and breed. Generally, we select between 1 to 4 controls matched to each case. There is minimal gain in efficiency if the control-to-case ratio exceeds 4:1. Although not necessary, it is simplest to use a fixed control-to-case ratio.

13.3.3 Caliper-matching

If the confounding variable to be matched on is continuous, we must specify how close, on the continuous scale, the subject must be in order to be considered matched and hence, this is called caliper-matching. Caliper-matching often produces a problem for analysis in that, if the individual-match must be within, say 2 years of age, then 2 case (exposed) subjects of the same age could be matched with controls (non-exposed) whose ages differ by almost 4 years. In this instance, we either have to live with the ‘wider’ match and chance residual confounding or decide to use strata in our analyses that are no wider than the ‘matching’ criteria even if that shifts the ‘matched’ subjects into different strata.

13.3.4 Analysing matched data

Frequency-matched data
In general, frequency-matched data should be analysed using a stratified method (as shown in Examples 13.1 to 13.3) to account for the matching. If pair-matching is used, with few categories of the confounder, and many pairs present within each category, the data could be analysed by creating a group identifier for the matched set of subjects and analysing the data as for a frequency-matched dataset using the group identifier to form the strata. The strata formed by the matching process must be preserved and an overall measure of association developed. Note that we cannot assess the main effect of the matched variable(s), but interaction between the confounder(s) and exposure should be evaluated in the usual manner. When additional confounders are present, a multivariable analysis using conditional logistic regression (see Section 16.15) can be used for the analysis.

Pair-matched data
If pair-matching is used, and there are many categories of the confounder and very few pairs within each category, the data must be analysed using a matched-pair analysis. For these analyses, we use the frequencies of matched sets in the 4 exposure and outcome patterns to estimate the odds ratio. In a case-control study, with 1 control matched to each case, there are 4 possible exposure patterns: both the case and its matched control were exposed; both non-exposed; case exposed and control non-exposed; case non-exposed and control exposed. The data layout is shown in Table 13.1.

Table 13.1 Data layout for matched-pair case-control analyses

		Control pair		Case totals
		Exposed	Non-exposed	
Case pair	Exposed	t	u	$t+u=a_1$
	Non-exposed	v	w	$v+w=a_0$
	Control totals	$t+v=b_1$	$u+w=b_0$	

The crude *OR* is:

$$OR_{\text{crude}} = \frac{a_1 b_0}{a_0 b_1} \qquad \textit{Eq 13.1}$$

The **Mantel-Haenszel matched *OR*** uses only the data in the discordant cells and is:

$$OR_{\text{match}} = \frac{u}{v} \qquad \textit{Eq 13.2}$$

The **Mantel-Haenszel χ^2 test** (which in the case of 1:1 matching equals **McNemar's test**), should be used for hypothesis testing with 1 df. The formula is:

$$\text{McNemar's}\,\chi^2 = \frac{(u-v)^2}{u+v} \qquad \textit{Eq 13.3}$$

Note Only the values in the discordant cells contribute to both the estimate of the *OR* and the McNemar's χ^2 test. Concordant pairs provide no useful information for the analysis. As with frequency matched data, conditional logistic regression (Section 16.15) can be used if multivariable modelling is required.

13.4 MATCHING USING PROPENSITY SCORES

Propensity scores have been used most commonly in cohort studies to evaluate the effects of treatments (or other exposures) when evidence from a randomised controlled trial is not available. In this situation, treated (in observational studies 'exposed') individuals may be substantially different from non-treated individuals and this difference needs to be accounted for. Propensity scores can be used in case-control studies with some limitations (see Månsson *et al*, (2007) for details).

Propensity scores were first proposed by Rosenbaum and Rubin, (1983). A **propensity score** (PS) is the conditional probability of being treated/exposed (*ie* the probability that an individual with certain characteristics will be treated/exposed) given the measured covariates. We denote this as p($E+|C$) (See Table 13.4). Once computed, propensity scores can be used for matching, as the basis for a stratified analysis, as a basis for doing a weighted analysis or they can be included in a modelling procedure as a covariate. One approach may prove to be more precise or less biased than another depending on the context (Austin, 2007; Austin, 2008b; Austin, 2009; Austin *et al*, 2007).

13.4.1 Computing propensity scores

With only 1 or 2 categorical confounders, propensity scores could be calculated manually, using the observed distribution of $E+$ within levels of the confounder C. With more confounders and/or continuous confounders, propensity scores are derived from a logit or probit model with treatment (observed exposure) allocation as the outcome (See Chapter 16). The question of what predictor variables to include in the model has been the subject of considerable recent research. In general, we accept that including potential confounders (*ie* variables, known or suspected to be related to both exposure and the outcome) and their interactions as necessary, is the most appropriate approach. The inclusion of a number of non-confounding extraneous variables can lead to problems of over fitting according to Senn *et al* (2007), but this problem was not noted by Austin *et al* (2007).

13.4.2 Balancing of exposure groups

Propensity scores can be used to help ensure equivalence of confounder distribution by 'balancing' the characteristics of the exposed and non-exposed individuals across all strata (also called 'levels' or 'blocks') of the PS (Austin, 2008c, 2008d). A study is considered balanced if 2 conditions are met. First, the average value of the PS is the same in exposed and non-exposed individuals within each stratum of PS. If this is not true, the stratum should be reconfigured until equality is achieved. Second, the mean value of all covariates making up the PS should be equal in the exposed and non-exposed groups within each stratum. As part of the balancing process, it is recommended to examine the distribution of each of the original confounders in the groups matched by PS score.

As noted, balancing does not require that individuals within a stratum be homogeneous. For example, mastitic cows treated with antibiotics may be a mixture of early lactation (treatment designed to minimise impact on milk production) and late lactation cows (treatment designed to eliminate the infection prior to the dry period if dry cow therapy is not practised). Consequently, cows with a high PS will be a mixture of early and late lactation cows. The study

would be considered balanced if within a stratum of PS (*eg* cows with high PS values), the treated and non-treated cows have the same average PS and the same mean days in milk (DIM).

Computation and evaluation of PSs is often limited to observations falling in the range of PSs that includes both exposed and non-exposed individuals (called the region of **common support**). Non-exposed individuals with PS values lower than the lowest value observed for a exposed individual are ignored as are exposed individuals with higher PSs than any non-exposed subject. Not accounting for these may seem wasteful or potentially biased; however, in the context of trying to assess the causal impact of an exposure, these individuals are so different from the others in the study group that regardless of their exposure and disease experience, it is virtually impossible to validly implicate exposure as a cause of the outcome. However, it is important to note the characteristics of these subjects because they may provide a clue about potential causes that can be addressed in future studies.

13.4.3 Matching on propensity scores

Usually, matching begins with obtaining the PS on the potential study subjects. Then, since exposed individuals usually are less frequent than non-exposed subjects, we begin by identifying an exposed individual and obtaining their PS. Then, one or more non-exposed individuals with a similar PS are selected from the available potential study group. Selection of matches is usually done with replacement (so a non-exposed individual can serve as a matched control more than once), but there are several ways to select the matched individuals. **Nearest-neighbour matching** selects the one or more individuals (*ie* 1:1 or 1:m matching) with PSs closest to the exposed subject. However, there is no guarantee that the matched individuals will have a PS that is very close to that of the exposed subject. **Radius matching** selects all non-exposed individuals with a PS within a certain distance of the value of the exposed individual (*eg* ± 0.05 PS units). In **kernel matching**, all non-exposed individuals serve as controls for each exposed individual, but they are weighted according to the closeness of their PS.

13.4.4 Analysis of propensity score matched data

Once the final groups are selected, it is generally recommended that the analysis of the matched data use procedures which take the matching into account although the need to do this has been questioned (Stuart, 2008). Software is available to implement the matching process, ensure the balancing of measured confounders within strata, and conduct the analysis on cohort or cross-sectional data (Becker & Ichino, 2002).

With cohort or cross-sectional-derived data, the most common measure of effect computed is the **average treatment effect in treated individuals** (*att*). It is the difference in the outcome measure between the treated (exposed) and non-treated (non-exposed) individuals. The outcome measure may be dichotomous (*eg att* is the difference in the proportion developing the outcome in the treated (exposed) compared with the non-treated (non-exposed) groups) or a quantity (*eg* average somatic cell count in treated minus the average SCC in non-treated). Analytic solutions exist for standard errors of the *att* if nearest neighbour or radius matching or stratification is employed. Bootstrap methods need to be used to estimate standard errors if kernel matching is used.

Example 13.4. shows the computation of PSs for an evaluation of the effect of Mh on the risk of

Example 13.4 Matching using propensity scores
data = feedlot

We wanted to evaluate the effect of seroconversion to *Mannheimia* cytotoxin (-mhcysc) on the risk of bovine respiratory disease (-brd-), but needed to control for the potential confounding effects of bovine respiratory syncytial virus infection (-brsvsc-), *Histophilus somni* infection (-hssc-), province of origin (-prov-) and weight at entry into the feedlot (-wt0-) using data collected in beef feedlots (Martin *et al*, 1999). Unconditionally, the risks of -brd- in -mhcysc- positive and negative cattle were 0.358 and 0.244, respectively, giving a *RR* of 1.47 and a *RD* of 0.115.

Propensity scores (PSs) were computed using a logistic regression of -mhcysc- on -brsvsc- -hssc-, -prov- and -wt0- with the analysis limited to the region of common support. Of the 587 individuals with completed data, 585 fell in the region of common support with 2 animals with very low PS being eliminated. Strata (blocks) were chosen to have a width of 0.1 and all PSs fell between 0.614 and 0.924. The balancing properties (see text) were satisfied with this stratification of PS.

Nearest-neighbour matching was used with many of the 109 of the 121 non-exposed animals being used multiple times as controls for the 466 exposed animals. The risks of -brd- in exposed and matched non-exposed cattle were 0.358 and 0.286, giving an *att* of 0.072 (SE=0.061, P=0.24). Nearest-neighbour matching has apparently increased the risk of -brd- in non-exposed animals and reduced the apparent effect of -mhcysc-.

Radius matching (with a radius of 0.01) results in 456 exposed animals being matched with 118 non-exposed. The risks of -brd- in exposed and matched non-exposed cattle were 0.364 and 0.221 giving an *att* of 0.143 (SE=0.052, P=0.006). This *att* represents an increase in the apparent effect of -mhcysc- compared with the crude *RD* (0.115). As we will see later, this probably represents a better evaluation of the effect of -mhcysc- than nearest-neighbour matching provided.

BRD. These PSs were then used to carry out both nearest-neighbour- and radius-matched analyses. Use of PSs on these data is a little unusual given that the exposure (Mh) is not a treatment or exposure that is 'assigned'. However, these data have been used for consistency with the other examples in this chapter.

We now move on to describe the main ways of detecting confounding. The first is through the use of causal diagrams, the second is a 'change-in-measure' approach, and the third (which we do not recommend) is to use statistical criterion.

13.5 DETECTION OF CONFOUNDING

Effective control of confounding during data analysis involves both identifying the potential confounders and then carrying out procedures to effectively control for confounding effects (although the 2 are closely linked). The identification of potential confounders usually commences with knowledge of the context and additional information about the biology of the outcome gained from a thorough review of the literature. Further consideration of whether or not a specific variable should be treated as a confounder, by exclusion, matching, or analysis, can be guided by the use of causal diagrams. Detection and control of confounders during analysis requires that data on the potential confounders have been collected during the early phases of the study (hence the value of a causal diagram, in addition to a thorough review of the literature). The next sections focus on identification of potential confounders and assessing whether or not they indeed are confounders.

13.5.1 Using causal diagrams to identify potential confounding variables

The identification of which potential confounders need to be controlled can be accomplished at the planning stage of the research project, or shortly after the final list of variables becomes available. In Chapter 1, we introduced the use of causal diagrams, and here we extend this approach as a method of determining whether or not a variable should be controlled. First, of course, we need to draw the **causal diagram** (also referred to as a **directed acyclic graph (DAG)** (see Greenland *et al* (1999) and Hernan *et al* (2002) for examples) using the principles explained in Chapter 1. VanderWeele *et al* (2008) describe the conditions necessary for causal modelling using DACs. Weinberg (2007) makes suggestions about how to incorporate interaction effects into a causal diagram.

Within the diagram, we identify the exposure factor and the outcome of interest, as specified in the major objective of the study. Any factor causally prior to the exposure factor that is on a pathway connecting the exposure and outcome is a likely candidate for control as a confounder. Factors that are causally (or temporally) after the exposure variable should not be controlled, nor should variables that are causally after the outcome. We formalise the process as follows:

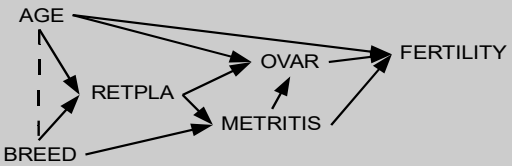
1. Draw the diagram using the guidelines outlined in Chapter 1 and as shown in Example 13.5.
2. Then eliminate all arrows emanating (*ie* leading away) from the exposure factor of interest on the graph.
3. If there are any paths that still connect the exposure and outcome, then the causally prior factors and other non-intervening variables in these paths should be controlled, otherwise these factors can bias the measure of association. In causal terminology, these factors produce spurious causal effects.
4. There is a final twist that is needed to complete this process. Suppose that there are 2 or more factors that 'cause' a third factor that is prior to the exposure factor and the initial assumption was that these 2 (or more) factors were unrelated, causally, to each other (*ie* these factors would be **marginally independent** statistically). AGE and BREED have this structure in Example 13.5 as they both cause RETPLA, but are independent of each other. However, when we control for a factor that they cause, this act makes these factors conditionally associated, and we will need to control for at least one of them to prevent bias. (This **conditional association** is shown as a dashed line on the diagram). To ascertain this, we need to connect all marginally independent factors with a 2-headed, or dashed, line. In tracing out pathways between the exposure and outcome we can go either way on this line. In order to 'close' this pathway, we will need to control for one (or more) of these factors in our modelling process. Thus, knowledge of the likely causal structure becomes very important in selecting factors for control, as control of one factor might necessitate control of others.

13.5.2 Change in measure of association as an indication of confounding

Once we have identified potential confounders, we can proceed to discover the magnitude (if any) of their effects. Since some confounding is almost always present in data from observational studies; the important question is when does it become sufficiently important to identify it as a problem? Suppose we begin our analysis of the study data with an unconditional (crude) association between our exposure and outcome variables and observe a **crude odds ratio**, OR_c. We then stratify the data based on a potential confounder, or a set of potential confounders. After having ensured that the stratum-specific odds ratios are deemed to be

Example 13.5 Identifying confounders using a causal diagram

We can use the causal diagram from Chapter 1 to demonstrate the application of the criteria for identifying confounders. Recall the example concerned studying the potential impact of selected diseases on infertility in dairy cows. We will add another variable to the diagram, BREED, and we assume that breed effects are transmitted through RETPLA and METRITIS. The causal diagram is:



where RETPLA is retained placenta, OVAR is cystic ovarian disease.

If we were interested in estimating the causal association between METRITIS and FERTILITY:

- omit the arrows leading forward from METRITIS to OVAR and FERTILITY.
- this leaves causal paths to FERTILITY from OVAR and AGE.
- the spurious causal path from METRITIS back to RETPLA through OVAR means that RETPLA needs to be controlled.
- once RETPLA is controlled we need to show that AGE and BREED have become conditionally associated (which we do by adding the dashed line). Although the original diagram shows them to be independent, controlling for RETPLA makes them conditionally associated.
- at this point, the only connection from METRITIS to FERTILITY is the pathway through BREED and AGE (because RETPLA has been controlled). This means that either BREED or AGE needs to be controlled to break that pathway. Controlling both would not be incorrect but is unnecessary.
- note that OVAR, being an intervening variable, is not controlled in the analysis.

Of course, there are more complex causal diagrams (see Hernan *et al* (2002)) but this example should convey the basics of their use.

approximately 'equal' to each other, we obtain the **adjusted odds ratio** OR_a. Almost always OR_a differs somewhat from OR_c, but if we deem the difference, relative to the unadjusted measure, to be 'large' (in some practical sense), we say that some or all of the factors we stratified on (or controlled) were confounders. Thus, when using the change in odds ratio between the crude (the baseline) and the adjusted values to determine if confounding is present, we need to specify a difference (*eg* >20-30% change in the odds ratio) that would be deemed important given the context of the study. (**Note** When computing the % change from the unadjusted to the adjusted 3 issues need to be considered. First, we would recommend always using the unadjusted values as the baseline. Second, for ratio measures (*eg OR*) the % change should probably be computed on the log scale (% change in $\ln OR$). This has the advantage that it works equally well for risk factors and protective factors ($OR>1$ and $OR<1$ respectively). However, the rule is commonly applied to *OR* directly and, for simplicity, we will do so throughout this chapter. Third, the % change criterion should only be applied to statistically significant variables. Non-significant variables for which $\ln OR \approx 0$ can have very large % changes with very small absolute changes. If this difference is exceeded, then we say confounding is present and the adjusted measure is preferred. Conversely, if there is virtually no

difference between the crude odds ratio and the adjusted odds ratio, we say that confounding was not present and the crude measure suffices.

In part, this inference and the **change-in-estimate** approach to identifying confounders are based on the fact that without confounding, if the stratum-specific measures are equal to *X*, then when the data are collapsed over that confounder, the crude measure will also be *X*. If the data meet this criterion, they are called **collapsible** (Kass & Greenland, 1991).

Non-collapsibility of odds ratios
The measure of association used can affect our interpretation of confounding. In particular, the odds ratio, which is our most frequently used measure, suffers from the problem that it is not always collapsible. If we are using risk difference or risk ratio measures of association, the crude measure will always be a weighted average of the stratum-specific measures; they are collapsible. And, as a result, in the absence of interaction, if no confounding is present, the data can be collapsed (*ie* summed over the levels of the confounder) and the stratum-specific risk ratios will be the same as the crude risk ratio. However, this is not true when the odds ratio is the measure of association. In this instance, even in the absence of confounding, the crude odds ratio can be closer to the null than the stratum-specific odds ratios; this is called **non-collapsibility**. This problem usually shows up when the outcome in one or more strata is very common as shown in Example 13.6. As a result, because the crude and adjusted measures differ, it might look as if confounding is present when it really isn't. Be aware of this situation. Notwithstanding the problem of non-collapsibility, the 20-30% change in odds ratio (or other measure of association) has become the standard method of identifying confounding.

13.5.3 Statistical identification of confounders

In this approach, a statistical algorithm is used to either select (*eg* through **forward selection** or **backward elimination**, with or without **stepwise** methods—see Section 15.8.2) variables from a regression model based on their statistical significance. This approach has become very convenient especially with the advancement of powerful statistical routines to select variables when building models, but it has rapidly lost favour in recent years for the control of confounding and the estimation of causal effects. An assumption underlying this method is that most confounders will be selected as 'statistically significant' by this process thereby preventing confounding. The major problem is that, in using this approach, we cannot (or do not) distinguish between intervening and other types of extraneous variable. Furthermore, the process flies in the face of statements that the extent of confounding bias is a matter of judgement, not a matter of statistical significance. Thus, we do not recommend using this approach for anything other than initial pilot studies of a particular problem, or preliminary analyses of a complex dataset.

To explain our reticence to rely on this method, we need to recognise that when we use a statistical algorithm to search for multiple risk factors simultaneously, we can break a number of 'rules' about what variables to control as confounders. With multiple factors under study, the causally prior factor that might need controlling to obtain valid effect estimates of one exposure factor could be an intervening variable for another exposure factor. Hence, the 'adjusted' measures of association we obtain from multivariable models are **direct effects** only, not **total causal effects**. The latter (the sum of the direct and all indirect causal pathways) is deemed to be the best estimate of the true causal effect.

Example 13.6 Non-collapsibility of odds ratios and disease frequency

An example of non-collapsibility of *OR*s between exposure (*E*) and disease (*D*) in the presence of a non-confounding extraneous variable (*Z*)[a]. Overall disease risk=0.55

	Z+		Z-		Totals	
	E+	E-	E+	E-	E+	E-
D+	870	690	430	200	1300	890
D-	130	310	570	800	700	1110
Totals	1000	1000	1000	1000	2000	2000
Risk	0.87	0.69	0.43	0.20	0.65	0.45
Risk ratio		1.26		2.15		1.44
Risk difference		0.18		0.23		0.20
Odds ratio		3		3		2.3

Note Variable *Z* is not a confounder because it is not associated with exposure (within both levels of *Z*, 50% of the individuals were exposed); it is however associated with the outcome *D*. Because the stratum-specific odds ratios are equal to each other, and hence to the OR_{MH}, but differ from the crude odds ratio, we might be tempted to conclude that confounding by *Z* is present. However, the difference in these odds ratios relates to the use of 'odds' as a measure of outcome frequency; there really is no confounding present in this example.

Non-collapsibility is a greater problem for interpretation when the outcome frequency is high (55% in this example). In the table below, the average risk is much lower at 8.3%, the data are 'virtually' collapsible.

An example of near-collapsibility of odds ratios between exposure (*E*) and disease (*D*) in the presence of a non-confounding extraneous variable (*Z*)[a]. Overall disease risk=0.083

	Z+		Z-		Totals	
	E+	E-	E+	E-	E+	E-
D+	211	82	29	10	240	92
D-	789	918	971	990	1760	1908
Totals	1000	1000	1000	1000	2000	2000
Risk	0.21	0.08	0.03	0.01	0.12	0.05
Odds ratio		3		3		2.8

[a] Example based on Greenland and Morgenstern (2001).

As this example indicates, in practical terms, confusing confounding and non-collapsibility is only a problem when the outcome frequency is high.

For example, in Fig. 13.1 (on page following), X_1 has a direct causal effect on *Y* and an indirect causal effect via X_2. X_2 has a direct effect on *Y* and a spurious effect via the X_1-*Y* pathway.

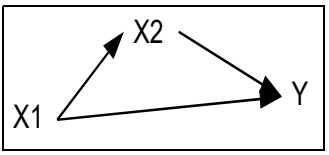


Fig. 13.1 Causal effects of 2 predictor variables on outcome Y

Assuming a reasonable sample size, a statistical algorithm is likely to identify the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

From this model, we could estimate the direct effect of X_1 using the β_1 coefficient. With respect to X_2, its direct (and total) causal effect on Y also could be estimated using the β_2 coefficient from this regression model. However, in order to correctly estimate the total effect (direct + indirect) of X_1 we should use the β_1 coefficient from the following regression (not including X_2.).

$$Y = \beta_0 + \beta_1 X_1$$

Thus, in using the first model, for estimating the causal association between X_1 and Y, (which would likely result from using statistical criteria to control confounding), we will have 'over-controlled' for intervening variables (and perhaps effects of other factors).

We might also note that the β_1 coefficient from the following regression is biased for the causal effect of X_2 since the coefficient contains some of the confounding effects of X_1.

$$Y = \beta_0 + \beta_1 X_2$$

When using statistical criteria to select variables, the lack of clarity about what the coefficients actually estimate increases with the complexity of the model. One conservative approach to managing more than one exposure variable in a dataset is to take the set of 'significant' variables and then conduct a separate analysis (as shown above) for each factor as the exposure of interest and use this measure of association as the best estimate of the causal association.

Now that we have the tools to identify factors needing control, we will move on to describe processes for implementing analytic control. In this respect, the details about these multivariable models are described in Chapters 14-23. The relationships between assumed causal structures and observed risks of disease will be elaborated in subsequent sections (beginning in Section 13.11).

13.6 ANALYTIC CONTROL OF CONFOUNDING

A variety of analytic procedures can be used to control for confounding. In Sections 13.6.1 through 13.6.3, we will describe methods that can be applied to stratified data (stratified by levels of confounder variable(s). These include the Mantel-Haenszel procedure (the most commonly used method for stratified data; Sections 13.6.1 and 13.6.2) and stratification by PS (Section 13.6.3). These approaches assume homogeneity of the association measure (*eg* odds ratio) across the strata in order to validly summarise the exposure-outcome data. In Section 13.7, we introduce standardisation (Section 13.7.1), and marginal structural models (Section

13.7.2). These approaches produce summary measures of association for specified populations regardless of the presence or absence of interaction. Subsequently, we will describe analytic control of confounding in multivariable models (Section 13.8). Instrumental variables are discussed in Section 13.9. External adjustment and sensitivity analysis for unmeasured confounders is covered in Section 13.10.

There are a number of approaches to the analysis of stratified data. In all of these approaches, we physically stratify the data by creating a 2X2 table for each level of the confounder (or combination of the confounders). The notation for each stratum is shown in Table 13.2.

13.6.1 Stratified analysis to control confounding: The Mantel-Haenszel Estimator

This procedure was first described by Mantel and Haenszel in 1959 and it revolutionised the work of epidemiologists. It became the most widely used stratified analytic approach for categorical data with dichotomous exposure and is known as the **Mantel-Haenszel (MH) procedure** (or estimator). The procedure is straightforward, easy to use, and its use can help inform the researcher of details of the data that otherwise might be missed. Indeed, we advise researchers to use this approach in initial analyses, whenever possible, even when planning to use more complex analyses such as logistic regression.

This method relies on physically stratifying the data according to the combination of levels of the confounding variables (see Table 13.2), examining the stratum-specific measures of association (odds ratios for now) and, if these are deemed to be equal (apart from sampling variation), creating a pooled 'weighted' or 'adjusted' estimate of the association. The equality of the stratum-specific measures can be evaluated visually, or statistically using a test for homogeneity (described below). Demonstrating this equality is a prerequisite to calculating a valid overall measure of association.

In order to describe the Mantel-Haenszel procedure, we will assume that we have dichotomous exposure and outcome variables and a one or more confounders. If a single dichotomous confounder is present, we will have 2 tables (*ie* strata), one for those with the confounder and one for those without the confounder: here we assume there are '*J*' strata. Recall that the n_j or m_j in Table 13.2 might not have a population interpretation depending on the study design (*eg* n_j is not an estimate of a population denominator in case-control studies). Nonetheless, the values in the cells are used for purposes of calculating the measure of association and its variance.

Table 13.2 Data layout for stratified analyses

	Exposed	Non-exposed	Total
Cases	a_{1j}	a_{0j}	m_{1j}
Non-cases	b_{1j}	b_{0j}	m_{0j}
Total	n_{1j}	n_{0j}	n_j

Note *j* is the stratum designator.

Eqs 13.4 to 13.9 show the necessary formulae for analysing binary data (*ie* risk, not rate, data) based on the *OR* as a measure of association. **Note** The MH procedure can also be used based on *RR*, *RD* and *IR* as measures of association.

We begin by stratifying the data as shown in Table 13.2 and calculating the stratum-specific *OR*s. The *OR* for the j^{th} stratum is:

$$OR_j = a_{1j} * b_{0j} / a_{0j} * b_{1j} \qquad \textit{Eq 13.4}$$

We also need the expected values and the variance of the exposed-diseased cell value. Under the null hypothesis (as opposed to the counterfactual basis used in Section 13.7.1), the expected number of exposed cases in the j^{th} stratum is:

$$E_j = m_{1j} * n_{1j} / n_j \qquad \textit{Eq 13.5}$$

and the variance of E_j is:

$$\text{var}(E_j) = V_j = m_{1j} * m_{0j} * n_{1j} * n_{0j} / n_j^2 * (n_j - 1) \qquad \textit{Eq 13.6}$$

The 'adjusted' or Mantel-Haenszel odds ratio is a weighted average across the strata:

$$OR_{\text{MH}} = \frac{\sum (a_{1j} * b_{0j} / n_j)}{\sum (a_{0j} * b_{1j} / n_j)} \qquad \textit{Eq 13.7}$$

from which we can obtain $\ln OR_{\text{MH}}$ for use in testing homogeneity (Eq 13.6).

As stated earlier, from a practical point-of-view, if the adjusted (pooled) measure differs from the crude measure of association (by an amount deemed to be important), then confounding is said to be present. If confounding is deemed to be present, the adjusted measure of association is always preferred to the crude measure.

Before interpreting the adjusted odds ratio as a valid summary measure of association, we need to examine the stratum-specific odds ratios and see if they are 'approximately' equal. Otherwise, the adjusted odds ratio oversimplifies the association. Inequality of stratum-specific odds ratios is an indicator of the possible presence of interaction—we say possible presence because confounding by an unknown factor can produce effects that resemble interaction. There is a Wald-type χ^2 test for interaction; but in general, it has low power, so we might benefit from relaxing the P-value for significance to the 10-15% level. The **Wald χ^2 test** for homogeneity with (j-1) df is:

$$\chi^2_{\text{homo}} = \sum \left(\frac{[\ln OR_j - \ln OR_{\text{MH}}]^2}{\text{var}[\ln OR_j]} \right) \qquad \textit{Eq 13.8}$$

where $\text{var}[\ln OR_j] = \frac{1}{a_{1j}} + \frac{1}{b_{1j}} + \frac{1}{a_{0j}} + \frac{1}{b_{0j}}$.

Whether or not interaction is deemed to be present depends in part on the scale of measurement of association. Here we present only odds ratios but we could use risk difference, relative risk, or rate ratio as measures. The finding of interaction in one scale does not necessarily translate into the presence of interaction in another (see Section 13.6.2).

An overall test statistic, with 1 df, for the significance of the summary odds ratio is:

$$\chi^2_{\text{MH}} = \frac{\left(\sum a_{1j} - \sum E_j\right)^2}{\sum V_j} \qquad \textit{Eq 13.9}$$

An example of the use of this approach is given in Examples 13.7 and 13.8. Formulae for stratified analyses of risk and rate data from cohort studies are available elsewhere (Rothman *et al*, 2008).

13.6.2 The Mantel-Haenszel estimator when interaction is present

In Chapter 1, we demonstrated how 2 or more factors that were members of the same sufficient cause exhibited biological synergism which, in turn, could lead to differences in risk depending on the presence or absence of other component causes. In the section just completed, we indicated that the exposure of interest had to have the same association across all levels of the confounder (or PS) in order to support the use of a single summary measure of association. A test of the equality of the stratum-specific measures of association served to assess this feature. If the stratum-specific measures were declared different, this was an indication that interaction was present and that the stratum-specific measures should not be averaged into a single overall measure such as the OR_{MH}.

Interaction is a somewhat confusing term. Its presence could provide clues about biological mechanisms or pathways of action, but whether it is deemed to be present or not depends on the statistical model and the scale of measurement. However, regardless of the scale or measure of association, **interaction** is said to occur when the combined effect of 2 variables differs from the sum of the individual effects in that scale. For current purposes, there are 3 types of joint effect that 2 (or more) exposure factors can produce: additive, synergistic (if the combined effect is greater than the sum of the individual effects) and antagonistic (if the combined effect is less than the sum of the individual effects). VanderWeele and Robins, (2007) have described the identification of synergism in the context of the sufficient-component-cause framework. In order to explain interaction, it will be helpful to return to some basic measures of single and joint-exposure risks. For this discussion, we will assume that we use the risk of disease as the outcome. Risk will be denoted as:

- R_{11} when the study subjects have both exposure factors 1 and 2; as
- R_{10} when the study subjects have only exposure 1; as
- R_{01} when the study subjects have only exposure 2; and, as
- R_{00} when the study subjects have neither exposure factor.

Now, the effect of each variable can be measured by either a difference measure such as the risk difference (*ie* $RD_{10}=R_{10}-R_{00}$) or a relative measure such as the risk ratio (*ie* $RR_{10}=R_{10}/R_{00}$). With these as the basis, we can examine the joint effects of 2 variables. Example 13.9 indicates some possible joint-exposure results when stratification is used to control confounding in the presence of interaction.

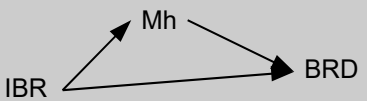
Additive scale of association
Using risk difference as the measure of association, additive interaction would be present if

$$RD_{10} + RD_{01} \neq RD_{11} \qquad \textit{Eq 13.10}$$

Generally, if the effects are measured as *RD*, and the effects are additive (scenario b in Example 13.9), this might be taken to indicate that the 2 factors operate through different biological pathways or mechanisms (*ie* they are not members of the same sufficient causes). The risk difference describes the excess number of cases that an exposure might cause.

Example 13.7 Stratified analysis of respiratory agents and bovine respiratory disease: no confounding
data = feedlot

In this dataset, there are data on the titres to a variety of putative respiratory pathogens in feedlot calves and on the occurrence of bovine respiratory disease (BRD). Experimentally, an interaction has been demonstrated between infectious bovine rhinotracheitis (IBR) virus and *Mannheimia haemolytica* (Mh), and as we have data on these, we can summarise the relationship of each of these agents, alone and together, on the occurrence of BRD. The exposure of interest is Mh, and our proposed causal model is:



We include a direct causal arrow from IBR to BRD because of our belief that IBR could enhance the respiratory pathogenicity of other unmeasured agents, besides Mh, and hence cause BRD. Thus, to ascertain the causal association of Mh with BRD, we need to control for IBR. The unconditional relationship of Mh with BRD has an *OR* of 1.69 and the χ^2 test is 5.19 with a P-value of 0.023. Hence, when we ignore the effects of IBR, seroconversion to Mh is associated with an increased risk of BRD of about 1.7 times.

In order to obtain the adjusted *OR*, we use the joint distribution of -Mh- and -IBR- to create the strata shown below:

Stratification of BRD by Mh and IBR, prior to Mantel-Haenszel analysis

IBR	BRD	Mh+	Mh-	Total
1	1	83	18	101
1	0	85	48	133
Total		168	66	
0	1	84	12	96
0	0	215	43	258
Total		299	55	

The layout of the essential calculations for the Mantel-Haenszel procedure is:

Stratum	OR	lnOR	var(lnOR)	a_j	E_j	var(E_j)	$a_{1j}*b_{0j}/n_j$	$a_{0j}*b_{1j}/n_j$
1	2.6	0.96	0.10	83	72.51	11.67	17.03	6.54
2	1.4	0.34	0.12	84	81.08	9.21	10.20	7.29
Totals				167	153.60	20.88	27.23	13.83

The 'adjusted' or Mantel-Haenszel odds ratio is:

$$OR_{\text{MH}} = \frac{27.23}{13.83} = 1.97$$

Based on these calculations, it appears that the strength of the association is slightly increased in the presence of IBR virus but perhaps not to the extent of being declared different from the effect when IBR virus is absent. However, we will perform a formal test of equality (or homogeneity) of the stratum-specific *OR*s.

(continued on next page)

Example 13.7 (*continued*)

The Wald test for homogeneity is:

$$\chi^2_{\text{homo}} = \frac{(0.96-0.678)^2}{0.100} + \frac{(0.34-0.678)^2}{0.120} = 0.795 + 0.952 = 1.747$$

where 0.678 is the ln(1.97). Although the stratum specific *OR* appears different (2.6 vs 1.4), this test result is non-significant (P=0.189); thus, we conclude that they do not differ statistically. An overall test statistic of the null hypothesis that OR_{MH}=1 is:

$$\chi^2_{\text{MH}} = \frac{(167-153.6)^2}{20.88} = 8.6$$

with 1df, P=0.003 so we can accept that OR_{MH} >1.

Based on this test, because P=0.003, we can reject the null hypothesis and conclude that there is good evidence that seroconversion to Mh increases the risk of BRD, after controlling the effects of IBR.

Compared with the crude *OR* of 1.69, the increase in size of OR_{MH} is only about 17%, so with our guideline of a change greater than 30%, we might say that serious confounding was not present and we might choose to use the crude *OR* to describe the causal association.

Multiplicative scale of association

Using a ratio measure of association, multiplicative interaction would be said to be present if:

$$RR_{10} * RR_{01} \neq RR_{11} \qquad \textit{Eq 13.11}$$

As this involves multiplying the relative measures, it is known as the **multiplicative model** or scale. Returning to our earlier definition of interaction, if we take logarithms of Eq 13.9, we have $\ln RR_{10} + \ln RR_{01} = \ln RR_{11}$ showing that additive effects on the logarithmic scale are equivalent to multiplicative effects (*ie* interaction) on the original scale. As we pointed out in Example 13.9, the risks of disease in jointly exposed individuals that are consistent with an additive arithmetic-scale model (scenario b) differ greatly from those that are consistent with an additive multiplicative-scale model (scenario c).

Returning to the stratified approach to data analysis, when the multiplicative-scale model holds, it can be shown that the *RR* for the primary exposure of interest will be the same in all strata of the extraneous variable(s). Thus, the equality of stratum-specific *RR*s, provides a convenient test for interaction in the **multiplicative scale**. This is also the basis of the test of homogeneity of *OR*s in the Mantel-Haenszel procedure (Eq 13.6)—we used *RRs* in Example 13.9 instead of *ORs* to keep the arithmetic simple. A significant test result indicates that the stratum-specific ratios are not equal, or equivalently, that the joint effect of the 2 factors is not what would be predicted based on the singular effects of the 2 variables (*ie* the effect of one exposure factor **depends on** the level of the other exposure). This phenomenon is referred to as interaction or **effect modification** (Susser, 1973) in the multiplicative scale.

The multiplicative model is widely used for assessing associations between dichotomous outcomes and exposures. It is applicable in a variety of contexts and study designs and appears to 'fit' observed data well. As Example 13.9 demonstrates, when the stratum-specific *OR*s are equal, the *RR* and *RD* measures will not be, and conversely if the *RD* measures were equal, then *RR* and *OR* would not be. Thus, in large sample-size studies, if the data are consistent with the additive model in one scale, they will be consistent with interaction in another scale.

Example 13.8 Stratified analysis when confounding is present
data = feedlot

Here we use the same dataset but control for Province (this is a surrogate for location of feedlot, partly for source of calves, and weight of calves on arrival). Our causal diagram is:

The data summary is:

Stratification of BRD by Mh and Province prior to Mantel-Haenszel analysis

Province	BRD	Mh+	Mh-	OR
1	1	84	21	2.75
1	0	80	55	
2	1	83	9	1.51
2	0	220	36	

The test of homogeneity of the stratum *OR*s had a χ^2 (1 df)=1.47 (P=0.23), so it is probably legitimate to calculate and interpret a weighted average *OR* as a summary measure. The crude *OR* is 1.69, and the OR_{MH} is 2.19. This is a 30% change in the coefficient and certainly suggestive of moderate confounding by Province being present. The test that the OR_{MH}=1 had a χ^2 (1 df)=11.20 with a P-value of <0.001 so we conclude that Mh and BRD are associated (or that OR_{MH} >1) after controlling for Province.

Thus, based on the crude *OR*, we might suggest that seroconversion to Mh was associated with an increased risk of BRD. After controlling for province where the feedlot was located, the relationship gets considerably stronger; thus, we would say that confounding by Province was present and the larger OR_{MH} (2.2) is the better indicator of the true causal association.

In Chapter 1, we demonstrated clear evidence of interaction arising from the sufficient cause model. Indeed, the sufficient cause model implies synergism which can show up statistically as interaction. However, we also demonstrated that, with the presence of unknown or unmeasured extraneous variables, interaction is not always detectable (even though the occurrence of synergism is the basis for the causal model). We also know that confounding can produce data that looks as if interaction is present, or conversely hide it. Thus, it is important to control confounding from other factors while trying to identify if interaction is present between 2 factors of interest. Example 13.10 demonstrates the detection of interaction when attempting to adjust for the effects of a confounder.

A biological example of known synergism is the combined effect of viral exposure of the respiratory tract of calves 4-6 days prior to exposure with Mh. Experimentally, this was a useful 'model' for reproducing the disease using aerosol challenges. Notwithstanding this, when the disease is observed in feedlots, even when a large number of organisms are measured and included in the model, it has not been possible to detect interaction (Examples 13.7 and 13.8).

Example 13.9 An example of identifying interaction between exposure factors for BRD using different scales of measurement

	Mh	BRSV	BRD (cases/1000)	Risk	RD	Additive scale interaction	RR	Multiplicative scale interaction
Effects of individual factors by scale of measurement								
	+	-	10	0.01	0.009		10	
	-	+	20	0.02	0.019		20	
	-	-	1	0.001				
Four possible scenarios (*ie* levels of combined risk) for joint effects								
a	+	+	100	0.100	0.099	synergism	100	antagonism
b	+	+	29	0.029	0.028	none	29	antagonism
c	+	+	200	0.200	0.199	synergism	200	none
d	+	+	300	0.300	0.299	synergism	300	synergism

Note Any joint effect above 29/1000 would be considered as synergism on the additive scale (a and c); whereas a joint risk of 200/1000 indicates no interaction on the multiplicative (*ie* log) scale (c).

13.6.3 Stratification using propensity scores to obtain average treatment (exposure) effects

This stratification involves dividing the observed data into strata (blocks) that were used to evaluate the 'balancing properties' of the PS procedure. A stratified analysis is then carried out and a summary estimate of *att* (see Section 13.4.4) obtained. An example of stratified analysis based on PS is shown in Example 13.11.

13.7 Other approaches to control confounding and estimate causal effects

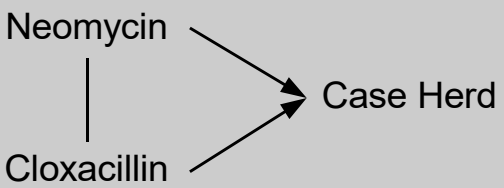
The next 2 methods are related although one uses standardisation to estimate the expected number of cases (or risk) and the other uses 'weights' to produce an unconfounded pseudo-population from which we can estimate the causal effect of interest using a crude (*ie* marginal) measure of association such as a risk ratio (we could also use risk difference or odds ratio as our effect measures). These approaches provide a valid summary of the effect of exposure in a specified population whether or not interaction is present; the stratum specific measures of association do not need to be homogeneous. These 2 features—that the population of interest is specified and that the summary measure is valid in the presence of interaction—are key elements for choosing this approach to estimating causal effects, although this putative benefit has been debated (Shah *et al*, 2005b).

13.7.1 Using standardised risks/rates to estimate causal coefficients

The use of direct and indirect standardisation was introduced in Chapter 6 as a descriptive

Example 13.10 Detection of interaction when controlling for a confounder
data = Nocardia

The data for this example are from a case-control study of dairy farms with and without *Nocardia* mastitis (these data are used extensively in Chapter 16). The exposure of interest was neomycin-containing dry-cow treatments. However, it was believed important to examine other dry-cow treatments also, both as possible risk factors and as potential confounders. Our causal model is:



We use a non-headed line between the 2 types of dry-cow treatment to indicate a non-causal correlation, likely because of a third common-cause factor such as management style. Even though the association is unlikely causal, using the rules of causal diagrams set out in Section 13.5.3, we need to control for cloxacillin to determine the causal effect of neomycin-containing treatments.

Stratification of case/control herds by neomycin and cloxacillin

Cloxacillin	Nocardia mastitis	Neomycin+	Neomycin-	Stratum-specific ORs
1	1	5	3	1.5
1	0	10	9	
0	1	44	2	29.3
0	0	15	20	

In the herds not using cloxacillin, the *OR* between neomycin use and case status was 29.3, whereas, in those herds using cloxacillin, the *OR* was 1.5. The test of homogeneity had a χ^2 of 6.44 (1 df) with a P-value of 0.011. This is considerable evidence of a difference in *OR* and is consistent with the presence of interaction. Hence, controlling for confounding is moot; we should not compute an adjusted *OR* because the association between neomycin use and case-control status (*Nocardia* mastitis) **depends on** the presence or absence of cloxacillin use on the farm. Thus, when interaction is present we should not interpret the summary measure because it varies with the level of other extraneous variables.

Example 13.11 Stratification using propensity scores
data = feedlot

The propensity scores computed in Example 13.4 were used as a basis for a stratified analysis of the effects of -mhcysc- on -brd-. Four blocks of PS contained data (0.6-0.7, ..., 0.9-1.0) and all 585 observations that fell in the region of common support were used in the analysis. The summary estimate of *att* was 0.144 (SE=0.049, P=0.003). This result was similar to the one obtained using radius matching (Example 13.4).

means to summarise data and adjust for confounders. However, this approach also can be used to estimate causal effect coefficients (Sato & Matsuyama, 2003). Essentially, whenever we have a risk factor (*eg* age) that has a different risk of the outcome at some or all of its levels, the overall number of cases for the group (or population) is the stratum specific risk multiplied by the number of subjects in that stratum as shown below:

$$\text{obs num cases} = \sum n_j * r_j$$

For purposes of demonstrating this method, we will use the data from Table 1.3. in which we discussed counterfactual approaches to the investigation of causal effects. The summarised data are shown in Table 13.3.

Table 13.3 Exposure, disease (cases) and confounder distribution in study-subject population

Stratum 1; confounder=0 (subjects have a low risk of infection; p(D+)=2/8=0.25)

	Exposed (vaccinated)	Non-exposed (non-vaccinated)	Total	Risk Ratio
Cases	1	1	2	
Non-cases	3	3	6	1
Total	4	4	8	

Stratum 2; confounder=1 (subjects have a high risk of infection; p(D+)=8/12=0.67)

	Exposed (vaccinated)	Non-exposed (non-vaccinated)	Total	Risk Ratio
Cases	6	2	8	
Non-cases	3	1	4	1
Total	9	3	12	

Using the levels of the confounder to form strata and, applying the previous formula, we would have (8)*0.25+(12)*0.67=10 observed cases overall.

Now, suppose we have 2 exposure groups (say the vaccinated and the non-vaccinated in Table 13.3). Their overall risks could differ even if their strata specific risks are equal. One way to obtain a fair comparison of the risks (or number of cases) by group is to indirectly standardise the groups such that they use a common set of risks (or rates) and apply these to the number of subjects in each strata as shown below

$$\text{exp num cases} = \sum n_j * std\ r_j$$

This allows us to contrast the observed and expected number of cases and then estimate a **standardised risk ratio** (*SRR*).

$$SRR = (\text{obs num cases}) / (\text{exp num cases})$$

We can use this approach to standardise a stratified set of counterfactual risks also; all we need to decide is what population we will choose to be the 'standard'. This will allow us to calculate the *SRR* which is a non-parametric way of estimating causal parameters. Because it relies on

physical stratification based on the combination of levels of the measured confounders, it can suffer from sparse data problems which can lead to unstable estimates. Little and Rubin (2000) published a review on the use of potential outcomes for causal modelling, and the standardisation process is described by Hernan & Robins (2006b); Newman (2006); Sato & Matsuyama (2003).

Using the data from Table 13.3, let's contrast the observed number of cases (n=7) in the vaccinated group (n=13) with the expected number of cases had exposure (*ie* vaccination in our example) not occurred. The counterfactual number of cases, if everyone in the vaccinated group was non-vaccinated, is found by using the risks a_{0j}/n_{0j} derived from the non-vaccinated group. (**Note** This is not the expected number of cases if vaccination and disease were independent of each other which could be estimated using $m_{1j}/n_{j\cdot}$.) To begin the standardisation, we note that the risk of disease in the non-vaccinated subjects with a low risk of infection was 0.25 and in non-vaccinated in the high risk of infection group it was 0.67 and these become our standard set of risks (see data=ind_vacc_summ). Thus, in terms of number of cases, we would have expected to see (4*0.25)=1 case in the 4 vaccinated members of the group with a low risk of infection. We would have expected to see (9*0.67)=6 cases in the 9 vaccinated members of the group with a high risk of infection. The standardised risk ratio in the vaccinated is SRR_{E+} is (Σ obs # cases)/(Σ exp # cases)=7/7=1.0. In general, the SRR_{E+} is the proportionate increase in risk in the exposed (here the vaccinated) due to being exposed—none in our example. This is the most common population to standardise for when making causal inferences and it allows us to estimate the average causal effect.

Using a similar approach, we could standardise the observed number of cases in the unexposed (non-vaccinated) group for the expected number of cases had vaccination occurred using the same approach as above. This would indicate the proportionate change in risk that would have occurred in the non-vaccinated group if they had been vaccinated. We also can standardise the observed number of cases in the total group for the expected number of cases under complete vaccination and complete non-vaccination by combining the above findings. The SRR_{tot} describes the proportionate increase in risk in the total population due to exposure (if everyone was exposed) compared with the risk if no one was exposed (see Sato & Matsuyama (2003) for worked examples).

13.7.2 Marginal structural models

In the marginal structural model (Robins *et al*, 2000; Suarez *et al*, 2008) the marginal distribution of the counterfactual risks is modelled as:

$$\log \mathrm{p}(D+) = \alpha_0 + \alpha_1 X$$

where X is the dichotomous counterfactual exposure variable and $\exp(\alpha_1)$ is the causal risk ratio.

The corresponding model for the observed data would be:

$$\log \mathrm{p}(D+) = \beta_0 + \beta_1 X$$

where X is the dichotomous exposure factor and $\exp(\beta_1)$ is the crude risk ratio. However, $\alpha_1 \neq \beta_1$ unless exposure is unconfounded. Robins *et al* (2000) have proposed a weighted analysis that gives unbiased estimates of the causal parameter α_1. Note that this is a 'marginal' model in that we do not need to condition on potential confounders; their effect has been removed through the construction of the **pseudo-population**. Thus, we can pursue the analysis in a 2X2 table format.

We will first develop the weights then explain how constructing the pseudo-population works to prevent confounding.

The idea behind this approach is to describe and account for the distribution of 'exposure' (vaccination in our example). For example, the data in Table 13.3 likely would not have arisen from a totally random distribution of vaccination, and we would have drawn the wrong conclusion about the effect of vaccination if the data were analysed with this randomisation process in mind. However, if it was recognised that the researchers had stratified their study group into those with high and those with low risk of infection and had randomly assigned 75% of the high-risk group and 50% of the low-risk group to receive vaccination then an appropriate analysis would have reached the correct conclusion about the effects of vaccination. Now, if this was an observational study, and we were sufficiently clever to identify the confounder (risk of infection from column 1 of Table 13.4), we could use that fact to 'explain' the distribution of vaccination in the 2 groups. Once recognised, we could 'control' the confounder (thereby explaining the exposure) and obtain the correct causal effect of vaccination. One method of achieving the 'control' is to create a pseudo-population by 'weighting' the groups.

The first component of the weight is the probability of receiving the exposure (*ie* $E+$ or $E-$) each subject actually received, conditional on the confounder information which is $p_E=p(E=e|C_j)$ with 'e' taking the values 1 or 0 depending on whether the subject was exposed (1) or not (0), and 'j' representing the strata formed by different levels of the confounder (or combinations of the confounders) (see Table 13.4). The weight W_i we assign to each subject is equal to the inverse of this probability which is $1/p_E$. The resultant estimator is called the **inverse probability of treatment weighted (IPTW) estimator** (Cole & Hernan, 2008; Hogan & Lancaster, 2004). Recall that, in Chapter 2, we introduced sampling weights and stated that they described the number of subjects each study subject represented. Well, the same approach applies here, and we use the weights to construct the pseudo-population. Not surprisingly, the total pseudo-population is twice the size of the observed population because it contains information on the counterfactual outcome (1 per subject) in addition to the observed outcome for each subject. This IPTW measure contrasts the outcome frequency if everyone in the study group was exposed versus the outcome if no one was exposed, and is equivalent to the SRR_{tot} (See Hernan & Robins (2006b) for a worked example). We can also obtain an estimate of the SRR_{E+} (our more usual population of interest) using weights (W_E) of $W_{E+}=1$ if the subject is exposed and W_{E-} = the odds of exposure in each level of the confounder if the subject is unexposed, as shown below:

$$W_{E-}=\frac{n(E=1|C_j)}{n(E=0|C_j)}=\frac{b_{1j}}{b_{0j}}$$

In using these weights to create the pseudo-populations we are assuming no confounding within the levels of the measured confounders which produces exchangeability and allows us to estimate the causal effects. However, we need to remind ourselves that this assumption is not verifiable from the available data and is an assumption that must be defended on other substantive grounds by the researcher. Recall the discussion in Chapter 7, about deciding on the exact composition of the groups that we wish to compare before seeing the outcome data (as suggested by Rubin, (2007)). Given that we cannot verify the exchangeability, it is vitally important that we at least have a consensus about what constitutes 'comparable groups' before potentially being biased by seeing the outcome data.

Table 13.4 Conditional probability of exposure, p(E=e|C), inverse probability of total exposure weights (W_T) and pseudo-population (pop_T), exposed group weights (W_E) and pseudo-population (pop_E) compositions and propensity scores for data in Table 13.3
data = vacc_factual

C	E	D	Obs no. n_j	p_E= p(E=e|C)	W_T =1/p_E	Pseudo pop_T no. =W_T * nj	W_E	Pseudo pop_E no. =W_E * n_j	Propensity Score
1	1	1	6	0.75	1.33	7.98	1	6	0.75
1	1	0	3	0.75	1.33	3.99	1	3	0.75
1	0	1	2	0.25	4	8	3	6	0.75
1	0	0	1	0.25	4	4	3	3	0.75
0	1	1	1	0.5	2	2	1	1	0.5
0	1	0	3	0.5	2	6	1	3	0.5
0	0	1	1	0.5	2	2	1	1	0.5
0	0	0	3	0.5	2	6	1	3	0.5

If we collapse the total pseudo-population over the confounder we obtain the total 'crude' pseudo-population data (Table 13.5). The marginal (or crude) risk ratio estimates the causal risk ratio which as we have seen earlier is 1 (apart from rounding error). Here, the IPTW estimate is the same as the SRR_{tot} estimate.

Table 13.5 The crude total pseudo-population composition and risk ratio

	Exposed	Non-exposed	Risk Ratio
Cases	9.98	10	1
Non-cases	9.99	10	

If we wanted to use the exposed (*ie* vaccinated) population as our standard (which is the 'usual standard'), we use the weights W_E and this leads to the exposed pseudo-population shown below in Table 13.6. Again, the marginal standardised risk ratio estimates the causal risk ratio in the exposed population which as we have seen earlier is 1.

Table 13.6 The crude exposed pseudo-population composition and risk ratio

	Exposed	Non-exposed	Risk Ratio
Cases	7	7	1
Non-cases	6	6	

As with other examples of applying this method, if we use the marginal structural model approach with the exposed as the population of interest on the data in Example 13.8, the exposed have W_{E+}=1 and the unexposed W_{E-}=0.76 (stratum 1) and W_{E-}=0.36 (stratum 2). Using these weights to create a pseudo-population (multiply the weights by the observed number of subjects), we can create a marginal table. The odds ratio in this table is 2.19, the same as was obtained by the Mantel-Haenszel approach, which implies that the risk of BRD was increased 2.2 times in the exposed subjects relative to what it would have been had they remained unexposed.

Similarly, if we use the marginal structural model with the exposed as the population of interest, on the data in Example 13.10, the exposed have W_{E+}=1 and the unexposed W_{E-}=0.42 (stratum 1) and W_{E-}=1.31 (stratum 2). Using these weights to create a pseudo-population (multiply the weights by the observed numbers), we create a marginal table. The odds ratio in this table is 12.3. We would interpret this to mean that the population of farms who used neomycin had an average of 12.3 times increase in their risk of *Nocardia* mastitis relative to what their risk would have been had they not used neomycin. (**Note** This is an estimate of the effect of neomycin for this set of farms. It does not provide any insight into the interaction between neomycin and cloxacillin.)

Newman (2006) shows how to extend the marginal-structural approach to the analysis of case-control studies and demonstrates the relationship between the standardised odds ratio (where the strata specific odds ratio is weighted by b_{0j}) and the Mantel-Haenszel odds ratio (where the strata specific odds ratio is weighted by n_j). Kurth *et al* (2006) compared the results of standardising, using IPWT, and propensity scores in the analysis of a large dataset; some of the measures differed greatly and reasons for these discrepancies were investigated and recommendations about the choice of analysis given.

13.8 MULTIVARIABLE MODELLING TO CONTROL CONFOUNDING

The most commonly used analytical method for controlling confounding is to include confounders in a multivariable model such as a linear regression model (Chapter 14) or other type of multivariable model (Chapters 16-24). In all these models, the effect of an exposure of interest is estimated given that other factors are held constant (or controlled). For example, in a logistic model in which -mh- was examined as a risk factor for BRD, along with IBR as a predictor, the effect of -mh- would be an estimate of its effect when comparing animals of comparable IBR status (effectively controlling confounding from IBR). This approach to controlling confounding is discussed in much more detail in Chapter 15. However, we will discuss the use of PSs and instrumental variables in multivariable models in the next 2 sections.

13.8.1 Multivariable modelling using propensity scores

As noted in Section 13.4, propensity scores can be used as an alternative to including individual covariates to control confounding in a multivariable model. At this point the question arises, "does using a PS in a multivariable model do a better job than controlling for confounding by including all of the potential confounders directly in the model?" (Austin, 2008a; Seeger *et al*, 2007). Martens *et al* (2006) summarised the findings of 2 recent major reviews of manuscripts which analysed data using both approaches and reported finding little evidence of a difference (Shah *et al*, 2005a; Stürmer *et al*, 2006). Closer analysis indicated that in general, studies using PS to control for confounding produced estimates closer to the null and this was especially true when the odds ratio was >2 or <0.5. The difference was exacerbated as the incidence proportion of the outcome increased and as the number of prognostic factors increased (Austin, 2007). The non-collapsibility of the odds ratio (Section 13.6.2) seemed to be the reason for most of these differences. Further, a simulation study of logistic regression models suggested that, if there are fewer than 7 outcome events per confounder, controlling confounding using a PS (included in the model as a categorical variable based on quintiles of the PS) was preferred. If there were 8 or more outcome events per confounder, a logistic model with the original confounders was the

technique of choice (Cepeda *et al*, 2003). In any event, the more factors that we try to match on, the greater the value of using the PS approach. If interaction is present (*ie* the effect of treatment varies with level of PS), then the way in which the PS is used may have a big impact on the overall effect estimate (Kurth *et al*, 2006). However, if interaction is present, the value of a summary measure of effect is questionable in the first place.

Thus, there is merit in considering the use of PSs instead of the traditional multivariable regression approaches, at least in the selected instances mentioned above. As Månsson *et al* (2007) have stated "if it is sufficient to adjust for individual covariates, then it is sufficient to adjust for the propensity scores". Perhaps the biggest benefit of this approach is that it changes the strategy of analysis. In traditional approaches, we focus on relationships between predictors and the outcome from the early stages of investigation and expend much energy on 'getting the association correct' (*ie* linearity *etc*). However, with propensity scores we place our emphasis on getting the groups 'comparable' so that our subsequent comparison of the outcome frequency in each group is valid. The focus on comparability is not biased (or should not be) by knowledge of predictor-outcome associations. However, as Stuart (2008) notes "Applied researchers wish to know 'best practices' for the use of propensity score methods in practice, but unfortunately clear advice does not yet exist". So we advise the reader to 'stay tuned' to future publications on the topic. Example 13.12 shows the use of PSs in multivariable models.

13.9 INSTRUMENTAL VARIABLES TO CONTROL CONFOUNDING

We begin this discussion by assuming that we wish to estimate the true causal effect of an exposure (or treatment) in a randomised controlled trial. In a perfect experiment (random selection of study subjects, randomisation of treatment (*ie* randomised exposure; *Z*), complete

Example 13.12 Use of propensity scores in a multivariable model
data = feedlot

The PSs from Example 13.4 were used in logistic models evaluating the effect of -mhcysc- on -brd-. Four models were fit:

- an unconditional model,
- a model in which the PS was included as a continuous variable,
- a model in which the PS was included as a categorical variable (based on the blocks generated when evaluating the balancing properties of the PS, and
- a model in which the covariates that were used to compute the PSs were included directly.

The resulting *OR* and CI were:

Model	OR	95% CI	
Unconditional	1.733	1.095	2.744
PS - continuous	2.236	1.380	3.622
PS - categorical	2.389	1.464	3.898
Original covariates	2.250	1.387	3.652

All models suggest that -mhcysc- increased the risk of -brd-. In all cases, controlling for the covariates increased the strength of this association, with very little difference between the model with PS as a continuous variable and the model with the original covariates.

compliance and follow-up of study subjects and the lack of measurement error in outcome status), the causal effect of a realised exposure (*E*) can be estimated as the difference (or ratio) in the mean value of the outcome (*D*) in the assigned exposed (treated) and non-exposed (non-treated or placebo) groups. However, one of the reasons that a field experiment (randomised controlled trial) can be 'imperfect' includes the lack of compliance—that is, not all subjects randomised to treatment (*Z*+) complete the treatment and some of the subjects randomised to the placebo (*Z*-) group may actually undergo the treatment (see Chapter 11 for a discussion of compliance). Hence, the difference (or ratio) in outcome between the assigned treated and placebo groups does not estimate the true causal effect of the exposure, but rather the likely effect of the exposure (the intention to treat analysis) if it were to be introduced to that population. If the data, on compliance, were available, we might wish to use this to estimate the true causal effect of treatment among subjects who actually complied; however, we might be concerned about the effect of confounding variables (*C*; measured and unmeasured) which might account for the failure to comply with the assigned treatment and also impact on the outcome risk and hence, bias the measure of association. A causal diagram of this scenario is shown in Fig. 13.2.

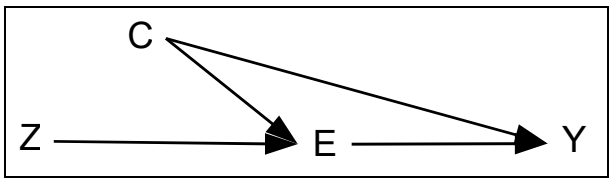


Fig. 13.2 Causal diagram showing instrumental variable Z

It turns out that we can estimate the true causal effect by using variable *Z* (the assigned 'or intent to treat' group) as an **instrumental variable** (IV). A valid IV (*Z*) must meet 3 requirements: it has a direct causal effect on exposure (or actual treatment; *E*); is unrelated to the outcome (*D*) except through its association with the exposure, and shares no common causes with the outcome. Here the IV is the randomised or intended exposure (Z); clearly this is related to the observed exposure and is unrelated to the outcome except through the observed exposure. And, *Z* shares no common causes with *Y* so it is unrelated to confounder(s) (*C*) be they measured or unmeasured. In the analysis of randomised controlled trials with incomplete compliance, the randomised treatment assignment serves as the IV for the actual exposure which is based on whether or not the subject complied with the randomisation process. The approach bypasses the need to adjust for confounders by estimating the **true causal effect** (*TCE*) (shown here on the difference scale) of the exposure based on the effects of the IV as shown below:

$$TCE = \frac{\mathrm{p}(D+|Z=1) - \mathrm{p}(D+|Z=0)}{\mathrm{p}(E+|Z=1) - \mathrm{p}(E+|Z=0)}$$

Note We use *D* here as a dichotomous realisation *Y*. The numerator estimates the effect of exposure as randomised (*ie Z*; this would also be the causal effect of the exposure with perfect compliance). The denominator reflects the association between the randomised exposure (*Z*) and the actual exposure (*E*). With perfect compliance the denominator becomes 1 and the *TCE* of *E* on *D* becomes the same as the effect of *Z* on *D* (*ie* the ratio estimates the causal effect of the exposure among those that actually were exposed in comparison to those who were not exposed). As the non-compliance increases, the denominator becomes smaller and inflates the

quotient so that the ratio consistently estimates the *TCE*. Most importantly, we do not have to correct for any potential confounders such as variable *C*.

Given the concern over unmeasured confounders in observational studies, finding an IV would clearly be an advantage. However, finding an IV that meets the criteria for a valid IV is rather difficult. Researchers have used context specific knowledge to try and identify suitable surrogate IVs, however, verifying the assumptions for a valid IV remains a challenge. Furthermore, specific IV methods are needed if the exposure is time-varying and if there is interaction between *E* and *D* (Bond *et al*, 2007). The situation is even more complex in that the direction of bias from the use of imperfect IVs is not intuitive (Bang & Davis, 2007; Hernan & Robins, 2006a; Johnston K *et al*, 2008; Martens *et al*, 2006; Rassen *et al*, 2009a; Rassen *et al*, 2009b). Terza *et al* (2008) caution researchers about using an IV that is adequate in a linear model, but then applying it in a non-linear model such as logistic regression. To date, we have not seen an application of IV methods in veterinary epidemiology.

13.10 EXTERNAL ADJUSTMENT AND SENSITIVITY ANALYSIS FOR UNMEASURED CONFOUNDERS

Sometimes we might have conducted a study without measuring or otherwise controlling the effects of one or more potentially important extraneous variables. We might have calculated a crude odds ratio between our exposure (*E*) and disease (*D*), but wonder what value it would have had if we had measured and controlled a particular confounder (*C*). Can we gain some insight into how much bias this unmeasured confounder might produce. The short answer is yes, but we would need to know 3 things, only one of which can be gleaned from the available data. They include the:

1. prevalence of the exposure variable, *E* (we can get an estimate of this from the control group in a case-control study)
2. strength of association between the confounding variable (*C*) and disease having adjusted for the exposure ($OR_{CD|E}$; sometimes we can obtain this value from other studies) and,
3. prevalence of the confounding variable among the exposed (P_{C1}) and non-exposed (P_{C0}) groups. We know these have to differ from each other, or else the factor would not be a confounder. We might obtain these estimates from other studies, or be able to make educated guesses about their values.

The adjustment procedure is as follows: first, we will assume the confounding variable is dichotomous, and thus, if we stratify on it, there will be 2 tables. These tables have the usual risk-based 2X2 structure, the first representing the data when the confounder is absent, and the second the data when the confounder is present. Now if the prevalence of the confounder is P_{C1} among the exposed and P_{C0} among the non-exposed, then within the exposed group, our predicted number of non-cases with the confounder ($C+$) will be $b_{11}'=P_{C1}b_1$. Within the non-exposed the predicted number of non-case subjects with $C+$ is $b_{01}'=P_{C0}b_0$ (see Example 13.13).

If it is reasonable to assume a common disease-confounding variable odds ratio (OR_{DC}), we can use these estimates of the number of non-cases to solve for a_{11} and a_{01} (*ie* the number of exposed and non-exposed cases with the confounder). The formulae (Rothman K *et al*, 2008) are:

$$a_{11}=\frac{OR_{DC}\,a_1 b_{11}'}{(OR_{DC}\,b_{11}'+b_1-b_{11}')} \quad \text{and} \quad a_{01}=\frac{OR_{DC}\,a_0 b_{01}'}{(OR_{DC}\,b_{01}'+b_0-b_{01}')} \qquad \textit{Eq 13.12}$$

Example 13.13 Effects of unmeasured confounders

Suppose we had observed the following hypothetical data on bovine respiratory disease (BRD) and *Mannheimia haemolytica* (Mh) in calves. Our interest was to ascertain if calves with Mh were at increased risk of BRD; however, we had not controlled for an important confounder such as BRSV. Our summary 2X2 table data would be:

	Mh+	Mh-	Totals
BRD+	78 (a_1)	11 (a_0)	89
BRD-	86 (b_1)	74 (b_0)	160
	164	85	249

The odds ratio would be 6.11 with a χ^2 statistic of 29.2 (P<0.001); it appears that Mh+ calves were at increased risk of BRD. But, perhaps this relationship was largely explicable by BRSV infection. What effect might this have on our observed association if we had measured it? Suppose there is evidence that BRSV (*Z*+) doubles (*ie* OR_{EZ} =2) the risk of BRD. We will also suppose that 60% of Mh+ calves and 40% of Mh- calves were infected with BRSV.

Based on this, the predicted number of non-case Mh+ calves that are also infected with BRSV is b_{11}=0.6*86=51.6 and the predicted number without Mh but with BRSV is b_{10}=0.4*74=29.6. Hence, solving for the expected number of Mh+ calves with BRD and BRSV we have:

$$a_{11}'=\frac{2*78*51.6}{(2*51.6+86-51.6)}=58.5$$

and for the Mh- cases with BRSV we have:

$$a_{10}'=\frac{2*11*29.6}{(2*29.6+74-29.6)}=6.3$$

We can now complete the first table for the BRSV-infected subjects (*ie* the *C*+ group).

BRSV+	Mh+	Mh-	Totals
BRD+	58.5	6.3	64.8
BRD-	51.6	29.6	81.2

The *OR* between Mh and pneumonia here is 5.3. Now, data for the second table for those without the confounder BRSV (*ie* the *C*- group) is obtained by subtraction from the original observed cell values (*eg* a_{10}= a_1-a_{11}).

BRSV-	Mh+	Mh-	Totals
BRD+	19.5	4.7	24.1
BRD-	34.4	44.4	78.8

The *OR* between Mh and pneumonia here is 5.4. The summary *OR* would be close to 5.3. Thus, at least with this set of estimates, the presence of BRSV infection in these calves would not explain very much of the observed crude association between Mh and BRD (*ie* the adjusted *OR* is only slightly smaller than the crude *OR*).

With these 2 cell numbers, we have complete information for the 2X2 table of subjects with the confounder. The table values for the subjects without the confounder can be obtained by subtracting the values for the subjects with the confounder from the original observed cell values. Given that we rarely know the true values of the parameters, the process should be viewed more as a 'what if' investigation than a true 'correction' of association measures. However, by substituting a reasonable range of prevalences and confounding-disease odds ratios, we can investigate the likely impact of this unmeasured confounding variable on the exposure-disease association. One 'what-if' example is shown in Example 13.14.

Similar approaches have been incorporated into software packages (Orsini N *et al*, 2008) and these allow a sensitivity analysis of confounding effects (see Example 13.14). Chiba *et al* (2007) and MacLehose *et al* (2005) have developed 'bounds' for confounding effects. McCandless *et al* (2008) demonstrate Bayesian sensitivity analysis for effects of unmeasured confounders. Yin *et al* (2006) discuss the use of information from secondary samples to control confounding.

13.11 UNDERSTANDING CAUSAL RELATIONSHIPS

In this section, we are interested in the effect of an extraneous variable given that we know the underlying causal structure. Hopefully, this will be of use for purposes of understanding the relationship between causal structures and the data we obtain in our studies. We do need to be careful however if, based on our analyses, we try to predict the causal structure. Although a number of researchers have tried to develop a general process for doing this successfully, regrettably, except in limited situations, our ability to infer causal structures from observed data is very limited, largely because we might be missing one or more important extraneous factors in our model (Thompson, 1991).

Example 13.14 Sensitivity analysis of unmeasured confounder effects

In Chapter 12, we used data from Nødtvedt *et al*, (2007), who reported that dogs born to bitches fed a commercial ration had a 2.3 times higher risk of atopic dermatitis than dogs from bitches fed home-cooked rations. For purposes of this example (and we do not claim this to be true), we might assume that the results are biased by the unmeasured confounder 'socio-economic class (SEC)' of owner. Specifically, we might posit that dogs from a high SEC (HiSEC) family are more likely (RR=2) to be taken to a veterinarian and diagnosed with atopic dermatitis than dogs from a lower SEC family. We could also posit that 50% of the dogs fed home-cooked meals were from the HiSEC and only 20% of those fed commercial rations were from the HiSEC. What impact might this have had on the findings if we had measured it and controlled for it? We used the Stata program -episens- (Orsini *et al*, 2008) to investigate this question.

Bearing in mind the observed OR and 95% confidence interval were 2.33 and [1.04, 5.19], the externally 'adjusted' OR was 1.86 giving a bias of 25%.

Our results suggest that if our assumptions about the strength of confounding were valid, the data would still have suggested an increased risk of atopic dermatitis in dogs fed commercial rations. However, if the detection bias by HiSEC dogs being taken to veterinarians was stronger (RR=5), then the authors most likely would have concluded there was no association of lactation diet and risk of atopic dermatitis because the adjusted OR would have been only 1.4.

While we rarely know the values of these 'what-if factors' we can at least posit a reasonable range of values and examine the likely affect on our results, and then interpret our results accordingly.

In the discussion that follows, we focus on causal structures and their impact on the disease frequencies that we observe. In reality, there are a number of ways in which factors can combine to produce disease and it is rare that we identify all of the component factors of particular sufficient causes. Thus, if we measure 2 potentially causal exposures, they might be members of the same or different sufficient causes or they might turn out not to be causes at all. Sometimes, because of the arrangement of some of the underlying causes, we might find **spurious relationships** (*ie* statistical associations when no causal relationship exists). Here we show some of the ways of detecting and understanding these relationships. Not all the relationships we demonstrate relate to confounding factors; however, they are intended to demonstrate the impact that different types of extraneous factors can have on the association between the exposure and outcome of interest. Because of their central value prior to and during analysis, we continue the discussion on causal diagrams that we began in Chapter 1 and elaborated on in Section 13.5.1.

13.11.1 Graphical aids to understanding multivariable systems

As a simple biological example we will continue to focus on identifying factors that might be of causal importance for bovine respiratory disease (BRD). We will suppose that our principal objective is to investigate the association between infection with the bacterium *Mannheimia haemolytica* (based on seroconversion) and the occurrence of BRD. Suppose the additional factor we measure is infection with bovine respiratory syncytial virus (BRSV; based on seroconversion). BRSV is only one extraneous factor but we can think of situations where there are numerous factors each with an underlying relationship with the exposure and/or the outcome. In the more general setting we are modelling relationships between an outcome (D) and an exposure of interest (E) in the presence of an extraneous variable (Z) which may or may not be a confounder or effect modifier).

The presumed causal relationship between pairs of variables will be shown using a **causal diagram**. In this instance, our predictor (or exposure) variables are BRSV and Mh. There are a number of possible causal models involving just 2 predictors that we will outline subsequently. When describing the causal (structural) relationships between variables using line diagrams, an arrow (directed edge) implies a cause-and-effect relationship, a double-headed arrow indicates unresolved causal correlation, a non-headed arrow (*ie* line) non-causal correlation (likely because of another unmeasured factor), and no arrow implies no causal relationship. In general, we would expect all relationships except the latter to result in significant statistical associations (exceptions will be noted subsequently).

We will describe the statistical results we expect, based on the causal structure in the line diagram, both visually using **Venn diagrams** and descriptively in the text. In the Venn diagrams, each circle represents a factor, or outcome, and the amount of overlap in the circles the extent (strength) of their association whether measured on a difference or a relative scale. If the circles do not overlap, this indicates that the factors are not associated statistically; it does not mean that they are mutually exclusive (*ie* do not occur together). The position (left to right) of each circle represents (where possible) the relative temporal (and potentially causal) positioning of the variables.

In describing these models we will assume all variables are dichotomous, similar to the factors used in Chapter 1, Example 1.1 where we use a relative measure of association (the risk ratio).

We continue to use that approach here except that we will use the *OR* as our measure of association (see Chapter 6). In the multivariable setting, when examining the Mh–BRD association, any factor that is not the exposure of primary interest is an **extraneous variable**. Susser (1973) named each type of extraneous variable based on their causal relationships with the exposure and outcome; we continue that practice with some revisions from his nomenclature. As noted previously, we can accomplish control of the extraneous variable(s) using matching, stratification or a multivariable regression approach—the latter are the subjects of detailed discussion later in this text (Chapters 14-23).

Hence,

1. *OR* is the unconditional (crude) *OR* between Mh and BRD. This is the measure we would obtain from a 2X2 table (or by analogy from a simple logistic regression model) when we ignore all other factors. When we 'adjust' or 'control' for other factors, the crude measure of association might change and it is referred to as a conditional, or adjusted, measure of association. Hence,
2. *OR*|BRSV is the conditional, or adjusted, *OR* (*eg* OR_{MH}) between Mh and BRD after controlling for the relationships with the extraneous variable BRSV. We could estimate this in a multivariable regression model by including BRSV in the model.

In each of the following sections we will:

- describe the causal relationships among the exposure, extraneous variable(s) and the outcome of interest,
- draw the causal relationships between the 2 predictor variables and the outcome to display the underlying causal structure,
- note the crude statistical association between Mh and BRD that we expect to observe given the causal model, and
- examine the association (in the absence of any sampling error) between the exposure and outcome after the extraneous variable is 'controlled' (*ie* through a stratified analysis of by addition of the extraneous variable to the regression model).

Mehio-Sibai *et al* (2005) provide a simple method for determining the direction of confounding, which builds on that of Susser. VanderWeele and Robins (2007) have described how to incorporate sufficient causes into a causal diagram and how this can help in deciding if a factor should be controlled. One of the constraints in using causal diagrams is how to incorporate interaction effects and we will describe some recent suggestions (Weinberg, 2007) in Section 13.11.9. VanderWeele *et al* (2008) have elaborated on the use of causal diagrams and the conditions necessary to infer the direction of bias from an unmeasured confounder. Streiner (2005) extends causal diagrams into a more formal path analysis with the important caveat that path analysis cannot prove causality. Bearing in mind the limitations of inferring causal structures from observed risks, we will now present a series of assumed causal structures and the most likely resultant statistical associations that researchers would observe.

13.11.2 Exposure-independent variable(s)

See the causal model in Example 13.15. The underlying causal structure is that both Mh and BRSV cause BRD but they are unrelated causally to each other; hence, BRSV is called an **exposure-independent variable**. Because of their lack of causal association with the exposure, unless they are correlated because of the effect of other factors, exposure-independent variables

are expected to be uncorrelated with the exposure. In observational studies, exposure-independent variables might arise naturally. In other situations the extraneous variables are causes of the outcome but also are related to the exposure of interest, and might be treated as a confounding variable. However, when matching is used to control these extraneous variables in cohort studies, the matched variables are converted into exposure-independent variables. Thus, they do not bias the measure of association and need not be 'controlled' analytically. In controlled trials (Chapter 11), we rely on randomisation to convert a number of causal extraneous cofactors into treatment-independent variables so they will not bias the measure of effect.

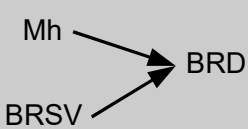
Exposure-independent variables do not distort the crude measure of association. This is displayed in Example 13.15 by noting that the portion of the outcome explained by BRSV does not overlap with the proportion explained by Mh. Thus, whether BRSV is included in the model or not makes no difference to the *OR*. However, exposure-independent variables account for some of the unexplained variation in BRD, often referred to as the residual variation. Thus, accounting for them in the analysis improves the precision of the estimate of association by reducing the unexplained variability in the outcome. In this context, the exposure independent variable may be manipulated to prevent future disease and may prove to be as, or more, important in this regard than the exposure of interest.

13.11.3 Simple antecedent variable

See Example 13.16. The underlying causal structure is that BRSV (the **simple antecedent**) increases susceptibility to -mh- which directly causes BRD. A simple antecedent is a variable that occurs temporally before the exposure variable, and is causally related to the outcome only through the exposure variable of interest. In our example, if BRSV is the simple antecedent, adding this variable to our model merely traces the sequence of causation backward in time.

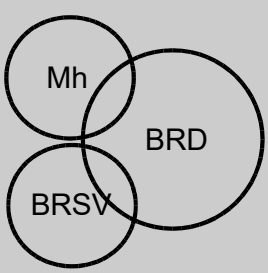
Example 13.15 An exposure-independent variable

Causal model



Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Statistical model



Comment
The two predictor variable circles do not overlap indicating their independence. Both exposure circles overlap with the outcome circle indicating their significant statistical association with BRD.

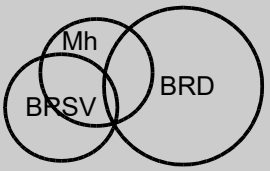
Example 13.16 A simple antecedent variable

Causal model

BRSV → Mh → BRD

Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Statistical model



Comment
Often there is a weak overlap between variables such as BRSV and the outcome, but statistical associations favour direct causes over indirect causes so the strength and significance level of the BRSV association might be low. The Mh–BRD association would not change when BRSV is controlled.

(This can be of importance in our understanding of the causal web, and in our attempts to control disease, so simple antecedents should not be dismissed as 'unimportant'.)

Assuming no sampling error, when BRSV is added to the model (*ie* its effects are controlled) it does not change the Mh–BRD association. By itself, BRSV might or might not be statistically associated with BRD; this depends on how much of Mh susceptibility is caused by BRSV and how much of BRD is attributable to Mh. However, when added to the model containing Mh, BRSV will not be statistically significant; any association it has with the outcome is already contained within the association explained by the exposure factor. Hence, in a forward model-building approach when -mh- is in the model, BRSV would not be added and the likely inference might be that it is causally unimportant. Technically however, it just means it has **no direct effect** on the outcome. The sample statistics are:

- Crude: *OR*(Mh) significant
- Crude: *OR*(BRSV) might or might not be significant—but *OR*(Mh) >*OR*(BRSV)
- Conditional: *OR*(Mh|BRSV)=*OR*(Mh)

Note When describing relative relationships with '>', we assume that the associations are positive, that is, producing odds ratios greater than 1. To include the possibility of both associations being negative, the > symbol might be read 'farther from 1' rather than just 'greater than 1'.

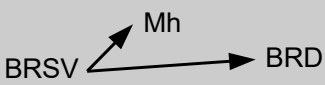
The *OR*(BRSV|Mh) is not a valid indicator of the causal association of BRSV with BRD; this *OR* reflects only the direct effect (which in this instance is 0). The crude *OR*(BRSV) is the correct estimate of the **total causal effect** of BRSV on BRD in this example.

13.11.4 Explanatory antecedent variable—complete confounding

See Example 13.17. The underlying causal structure is that BRSV precedes and causes (or predicts) both -mh- and BRD, but -mh- is not a cause of BRD. Statistically, we expect to observe a significant crude relationship between -mh- and BRD because of the common cause BRSV. This association is causally spurious. When BRSV is added to the model, the

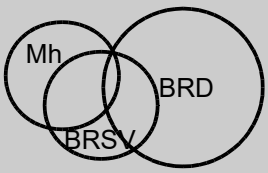
Example 13.17 An explanatory antecedent variable with complete confounding

Causal model



Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Statistical model



Comment
The Mh circle overlaps with the outcome, as they are statistically related until BRSV is added to the model (*ie* controlled). Then, the association becomes non-significant as all of the previous crude association between Mh and BRD is covered by the BRSV-BRD association.

association between -mh- and BRD becomes non-significant, because BRSV now 'explains' the original association. Thus, we would infer (correctly) that -mh- was not a cause of BRD. Adding BRSV to the model usually reduces the residual variance also. Many extraneous factors function as explanatory antecedents in this manner. The sample statistics are:

- Crude: *OR*(Mh) and *OR*(BRSV) are significant, usually with *OR*(BRSV) >*OR*(Mh)
- Conditional: *OR*(Mh|BRSV)=1, (BRSV biases the *OR* for -mh- if it is ignored), *OR*(BRSV|Mh) >1

Note The results of the model with both BRSV and -mh- included as predictors is not optimal for estimating the BRSV total causal effect. Once we remove all arrows emanating from BRSV, (item 2 in Section 13.6.3) there is no pathway from BRSV through -mh- to BRD, hence the model with BRSV only is preferred for estimating this causal effect. Controlling -mh- might not change the BRSV coefficient greatly, but it is better NOT to control unnecessary variables as controlling them can necessitate having to control even more variables.

13.11.5 Explanatory antecedent variable—incomplete confounding

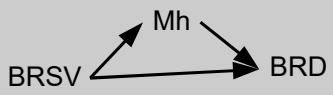
Example 13.18 shows a very common causal structure. The underlying causal structure is that BRSV causes (or predicts) both -mh- and BRD, but -mh- is also a cause of BRD. The sample statistics are:

- Crude: *OR*(Mh) and *OR*(BRSV) are significant
- Conditional: *OR*(Mh|BRSV) <*OR*(Mh) but *OR*(Mh|BRSV) ≠1

The model with both predictors included is appropriate for estimating the total causal effect of Mh. Statistically, as Mh still has an association with BRD after control of the confounder BRSV, this is the best estimate of its causal association with BRD. Thus, we would infer that Mh was a cause of BRD, and that the reduced 'strength' was the best estimate of magnitude of causal effect because the spurious causal component (from BRSV) was removed. Again, adding BRSV to the model usually decreases the residual variance of the model.

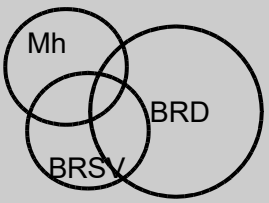
Example 13.18 An explanatory antecedent variable with partial confounding

Causal model



Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Statistical model



Comment
The Mh circle overlaps with the outcome. The association remains statistically significant when BRSV is added to the model (*ie* controlled), but some of the previous association is now attributed to BRSV. Thus, the Mh–BRD association is not as strong when BRSV is controlled as when it was not included. Adding BRSV to the model explains more of the variation in BRD than just knowing Mh status.

Note The results of the model with both BRSV and -mh- included as predictors are inappropriate to estimate the total causal effect of BRSV as only the direct effect would be reflected in the *OR* or regression coefficient. Mh would function as a partial intervening variable and should not be controlled when estimating the BRSV causal association with BRD. Again, the model with only BRSV is preferred for this purpose.

13.11.6 Intervening variable

See Example 13.19. An **intervening variable** is one that, in causal or temporal terms, intervenes in the causal or temporal pathway between exposure and disease. Now, although unlikely from a biological point of view (humour us on this), the underlying causal structure is that Mh causes (or predicts) BRSV and BRSV causes BRD. The sample statistics are:

- Crude: Likely both *OR*(Mh) and *OR*(BRSV) significant
- Conditional: *OR*(Mh|BRSV)=1

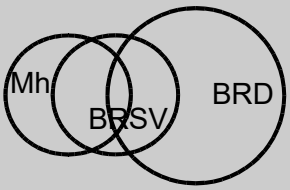
Although this conditional model is improper in the context of ascertaining the causal association of Mh on BRD, the model with both Mh and BRSV would provide a reasonable estimate of the causal association of BRSV with BRD. Nonetheless, the model with only BRSV included would be preferable for estimating the BRSV causal effect.

As noted, we recognise that this is, biologically, a silly example because we have no evidence that Mh would cause increased susceptibility to BRSV in the context of feedlot respiratory disease. However, often it is not so obvious. Thus, it is very important to identify intervening variables and not 'control' them (*ie* do not put them in the model). Intervening variables might be totally or only partly caused by the exposure but should not be 'controlled'. They are not confounders but they cause similar changes in the measure of association to explanatory variables; thus, we must know the likely causal structure and time sequence between variables to differentiate explanatory from intervening variables. They cannot be differentiated

Example 13.19 An intervening variable

Causal model

Statistical model



Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Comment
The Mh circle might or might not overlap with the outcome. However, any association of Mh with BRD disappears when BRSV is added to the model (*ie* controlled). It might well be that all of the effect of Mh on BRD is mediated through BRSV (and in that sense Mh is still a cause of BRD), but adding BRSV to the model would lead us to conclude that Mh was not associated with BRD and therefore we might infer that Mh was not a cause of BRD. Intervening variables should be identified and should not be controlled when estimating the causal effect of an exposure.

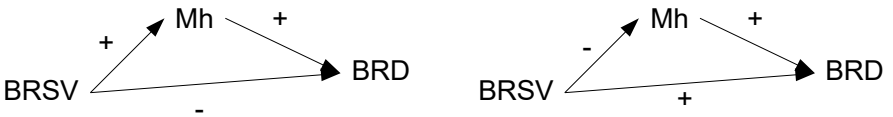
analytically. This is a major reason for our stressing the development and use of explicit causal diagrams before initiating analyses.

Before leaving this example, we wish to make note of the discussion of direct effects by Petersen *et al* (2006). Suppose our causal model had a direct arrow going from Mh to BRD, as well as the path through BRSV. If we did wish to estimate the direct effect of Mh we could achieve that by controlling for BRSV, provided Mh and BRSV did not interact in their effects on BRD. If they did interact, we would need to create an Mh*BRSV term to ascertain the direct effect when BRSV was absent and the direct effect when BRSV was present. Note that in either instance (interaction present or not) controlling for BRSV also blocks effects of other variables whose effect might be mediated through BRSV. This was termed the '**controlled direct effect**' Petersen describes an approach to estimate the direct effect of an exposure (*eg* Mh) when the effect of the exposure on the intermediate is blocked, but the effects of the intermediate and variables that cause it are not 'controlled'. This was termed the **natural direct effect**. In order to obtain this effect, a second regression of the intermediate (BRSV) on the exposure and confounders is necessary to obtain the likely level of the intermediate at the reference level of exposure. Then the effects obtained form the controlled direct estimates are weighted to obtain the natural direct effect. The necessary assumptions about confounding for this approach to be valid are explained (chiefly no unmeasured confounding of the exposure- intermediate association (*ie* Mh-BRSV) and no confounding of the intermediate-outcome (*ie* BRSV-BRD) association).

13.11.7 Distorter variable

Causally this is the same model setup as for explanatory variables except that at least one of the causal effects is of a different sign than the other 2 (*ie* one of the causal arrows reflects prevention not causation). In our example, there are 2 possible underlying causal structures, assuming Mh is a cause of BRD. In the model on the left, Mh is a cause of BRD and BRSV

prevents BRD but is causally and statistically positively correlated with Mh. In the model on the right, Mh is a cause of BRD and BRSV is also a cause of BRD, but BRSV is causally and statistically negatively correlated with Mh. Thus, the causal structures could be either:



The sample statistics for both models are:

- Crude: *OR*(Mh) and *OR*(BRSV) might be <1, =1 or >1
- Conditional: *OR*(Mh|BRSV) >1
  *OR(*BRSV|Mh)<1 (left side model)
  *OR(*BRSV|Mh)>1 (right side model)

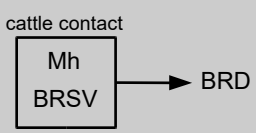
In either model, to estimate the causal association of Mh with BRD, we need to control for BRSV. Controlling BRSV will increase the strength of association between Mh and BRD (*eg* a non-significant *OR*(Mh) might become significant when BRSV is controlled). This potential for increasing the *OR* is of significance in model-building. When it occurs it signals an underlying relationship similar to that described here. It is also possible that a significant positive association can become a significant negative association, and only **distorters** can cause this reversal in the direction of association. The preferred model to estimate the total causal association of BRSV with BRD is the model with only BRSV included. When Mh is included, only the direct effects of BRSV are obtained.

13.11.8 Suppressor variables and refinement of exposure and outcome variables

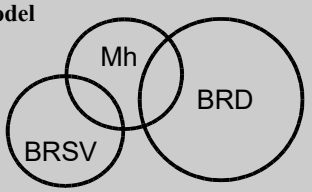
See Example 13.20. Here the underlying causal structure is that Mh is a cause of BRD and BRSV is not. What distinguishes this from the other examples of relationships with extraneous

Example 13.20 A suppressor variable

Causal model



Statistical model



Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Comment

Before control of BRSV the variable 'cattle contact' is not, or only weakly, associated with BRD. Once BRSV is controlled by refinement (usually) or analysis, the Mh circle overlaps with the outcome indicating an association of 'cattle contact' with BRD. By controlling the non-causal component of our global variable, we increase the strength of the remaining factors' association with the outcome.

variables is that both Mh and the **suppressor** BRSV are members of the same global variable as defined by the researcher. For example, we might have measured 'cattle contact' as a surrogate for exposure to infectious agents. However, because we are assuming that BRSV is not a cause of BRD (in this example), when BRSV is controlled, it will reveal or strengthen the suppressed association between Mh and BRD. BRSV is the (or one of the) irrelevant components of the global variable 'cattle contact'. The refined variable, without BRSV included, would have a stronger association with BRD. Control in situations such as this is usually by **refinement** of the predictor variable(s), but can be accomplished using analytical methods also.

Suppression often occurs with portemanteau-type (global) predictor variables (these are crudely defined or complex variables that contain a number of components). By refinement (stripping away the useless parts), the components of the original variable that are important can be identified. For example, 'ration' might need to be refined to locate which components (if any) of ration (type of roughage, length of roughage, amount of roughage *etc*) are related to abomasal displacement in dairy cows. We had suppression in mind when discussing combining length of exposure with dose of exposure to make a composite variable (in cohort and case-control studies; Chapters 8 and 9). Hence, we stated that it is best to examine the relationship of the components separately before assessing the composite variable for this reason.

Suppression of the dependent variable can also occur. As an example, perhaps only fibrinous pneumonia, not other types of respiratory disease, is related to Mh. Thus, if crude morbidity is the outcome variable, the association between Mh and BRD will be weak. If cause-specific BRD is used as the outcome, the stronger association between Mh and fibrinous pneumonia can be uncovered. Thus, whenever possible, refine the exposure factors and outcome variables to the point that suppression is unlikely. The extent of refinement used will, however, depend on the objectives of the study as well as practical constraints.

13.11.9 Moderator variable

See Example 13.21. **Moderator variables** produce statistical interaction. The underlying causal structure is that Mh causes BRD, but only when BRSV is present. Hence, the statistical strength of its association with BRD depends on the presence or absence of BRSV. In the first model, we show this with arrows of different density; in the second we use the approach of Weinberg (2007) to show that BRSV affects the strength of the Mh-BRD association. Recall from Chapter 1, that interaction is the statistical result of the joint causal effect of 2 or more factors on an outcome parameter. Interaction can, but doesn't necessarily, reflect a biological property of the joint effect of variables (*ie* either synergism or antagonism). Moderator variables might or might not be confounders, but since the summary measure of association is misleading we do not summarise over the strata. Assuming no residual or unmeasured confounding within strata, confounding is no longer of concern. The sample statistics are

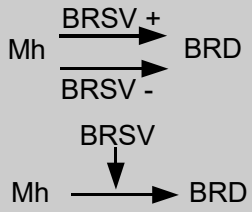
- Crude: *OR*(Mh) and *OR*(BRSV) usually≠1, but might=1
- Conditional: *OR*(Mh|BRSV) might not be meaningful because *OR*(Mh|BRSV+) ≠ *OR*(Mh|BRSV-), and χ^2_{homo} is significant (Eq 13.8)

13.12 SUMMARY OF EFFECTS OF EXTRANEOUS VARIABLES

We summarise the previous discussion in Table 13.7. We indicate the likely impact of adding each type of extraneous variable (*ie* BRSV) to an analysis of the Mh–BRD association on the

Example 13.21 A moderator variable

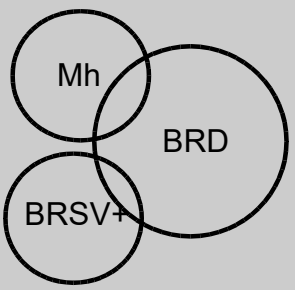
Causal model (Model 2 drawing is based on Weinberg (2007)



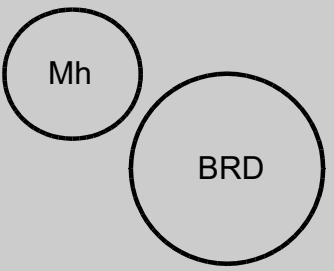
Mh = Mannheimia haemolytica
BRSV = Bovine respiratory syncytial virus
BRD = Bovine respiratory disease

Statistical model

When BRSV is present, the effect of Mh is present:



When BRSV is absent, the effect of Mh is absent:



Comment
The Mh circle overlaps with the outcome only when BRSV is present. This is the exact basis of the causal models shown in Examples 1.1 and 1.2. No disease occurs unless the two factors are present. Interaction is extremely important to identify as it has large implications for disease prevention.

magnitude (or direction) of the association of Mh with BRD. The association is measured as a regression coefficient (β_{Mh}) denoting the magnitude and direction of association in simple linear (Chapter 14), logistic (Chapter 16), ordinal (Chapter 17) and Poisson (Chapter 18) regression models, and in survival models (Chapter 19).

Table 13.7 Effect of controlling BRSV on Mh–BRD association as measured in a simple regression-type model

BRSV is a(n) ... variable	Effect on β_{Mh}	Comments (including impact on regression models)
Exposure independent	no change	BRSV explains some of BRD incidence, so the residual σ^2 is smaller and the significance of β_{Mh} increases
Simple antecedent	no change	No effect on the analysis by BRSV might be important to know about, from a preventive perspective, if it is easier to modify than Mh
Explanatory antecedent (complete confounding)	 becomes 0	Control of BRSV will remove any Mh association with BRD. The R^2 of the model should increase as the residual variance decreases
Explanatory antecedent (incomplete confounding)		Controlling BRSV will impact on the significance of β_{Mh} depending on the strength of the BRSV effect on Mh and on BRD. The R^2 of the model should increase
Intervening		Because BRSV is more closely related to BRD, it probably has a stronger association and explains more variability. The β_{Mh} is reduced in size and significance. If all of the effect passes through the intervener, it will remove all of the Mh effect on BRD
Distorter		Essentially the same impact as an explanatory-antecedent variable except the Mh effect is increased, or in the opposite direction, to the crude association
Suppressor		As the global variable containing Mh is refined, it will now have a stronger relationship with BRD, it will probably explain more of the variation in the outcome
Moderator	not applicable	In the presence of interaction, the effect of one variable depends on the level of the other variable, hence separate estimates of effect are required

REFERENCES

Austin PC. Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulations. Biom J. 2009; 51: 171-184.

Austin PC. Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score. Pharmacoepidemiol Drug Saf. 2008D; 17: 1202-1217.

Austin PC. Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. Pharmacoepidemiol Drug Saf. 2008C; 17: 1218-1225.

Austin PC. The performance of different propensity-score methods for estimating relative risks. J Clin Epidemiol. 2008B; 61: 537-545.

Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Stat Med. 2008A; 27: 2037-2049.

Austin PC. The performance of different propensity score methods for estimating marginal odds ratios. Stat Med. 2007; 26: 3078-3094.

Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. Stat Med. 2007; 26: 734-753.

Bang H, Davis CE. On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? Stat Med. 2007; 26: 954-64.

Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores Stata J 2002; 2: 358-77.

Bicalho RC, Warnick LD, Guard CL. Strategies to analyze milk losses caused by diseases with potential incidence throughout the lactation: a lameness example J Dairy Sci 2008; 91: 2653-61.

Bond SJ, White IR, Sarah Walker A. Instrumental variables and interactions in the causal analysis of a complex clinical trial Stat Med. 2007; 26: 1473-96.

Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders Am J Epidemiol. 2003; 158: 280-7.

Chauvin C, Gicquel-Bruneau M, Perrin-Guyomard A, Humbert F, Salvat G, Guillemot D, Sanders P. Use of avilamycin for growth promotion and avilamycin-resistance among Enterococcus faecium from broilers in a matched case-control study in France Prev Vet Med 2005; 70: 155-63.

Chiba Y, Sato T, Greenland S. Bounds on potential risks and causal risk differences under assumptions about confounding parameters Stat Med. 2007; 26: 5125-35.

Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models

Am J Epidemiol. 2008; 168: 656-64.

Cramer G, Lissemore KD, Guard CL, Leslie KE, Kelton DF. Herd- and cow-level prevalence of foot lesions in Ontario dairy cattle J Dairy Sci 2008; 91: 3888-95.

Glickman LT, Glickman NW, Moore GE, Goldstein GS, Lewis HB. Evaluation of the risk of endocarditis and other cardiovascular events on the basis of the severity of periodontal disease in dogs J Am Vet Med Assoc 2009; 234: 486-94.

Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research Epidemiology. 1999; 10: 37-48.

Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: a systematic review Ann Epidemiol. 2008; 18: 746-51.

Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology Am J Epidemiol. 2002; 155: 176-84.

Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006a; 17: 360-72.

Hernan MA, Robins JM. Estimating causal effects from epidemiological data J Epidemiol Community Health. 2006b; 60: 578-86.

Hogan JW, Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies Stat Methods Med Res. 2004; 13: 17-48.

Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research Stat Med. 2008; 27: 1539-56.

Kass PH, Greenland S. Conflicting definitions of confounding and their ramifications for veterinary epidemiologic research: collapsibility vs comparability J Am Vet Med Assoc. 1991; 199: 1569-73.

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

Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect Am J Epidemiol. 2006; 163: 262-70.

Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches Annu Rev Public Health. 2000; 21: 121-45.

MacLehose RF, Kaufman S, Kaufman JS, Poole C. Bounding causal effects under uncontrolled confounding using counterfactuals. Epidemiology. 2005; 16: 548-55.

Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, Rochon PA, Anderson

GM. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding BMJ. 2005; 330: 960-2.

Manske T, Hultgren J, Bergsten C. The effect of claw trimming on the hoof health of Swedish dairy cattle Prev Vet Med. 2002; 54: 113-29.

Månsson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case-control and case-cohort studies Am J Epidemiol. 2007; 166: 332-9.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease J Natl Cancer Inst. 1959; 22: 719-48.

Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations Epidemiology. 2006; 17: 260-7.

Martin SW, Nagy E, Armstrong D, Rosendal S. The associations of viral and mycoplasmal antibody titers with respiratory disease and weight gain in feedlot calves Can Vet J. 1999; 40: 560-7, 570.

McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding J Clin Epidemiol. 2008; 61: 247-55.

McCarthy HE, French NP, Edwards GB, Miller K, Proudman CJ. Why are certain premises at increased risk of equine grass sickness? A matched case-control study Equine Vet J. 2004; 36: 130-4.

Mehio-Sibai A, Feinleib M, Sibai TA, Armenian HK. A positive or a negative confounding variable? A simple teaching aid for clinicians and students Ann Epidemiol. 2005; 15: 421-3.

Newman SC. Causal analysis of case-control data Epidemiol Perspect Innov. 2006; 3: 2.

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, bull terrier and West Highland white terrier dogs in Sweden Vet Dermatol. 2007; 18: 309-15.

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

Olde Riekerink RG, Barkema HW, Kelton DF, Scholl DT. Incidence rate of clinical mastitis on Canadian dairy farms J Dairy Sci. 2008; 91: 1366-77.

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

Petersen ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects Epidemiology. 2006; 17: 276-84.

Pinchbeck GL, Clegg PD, Proudman CJ, Morgan KL, French NR. Whip use and race progress are associated with horse falls in hurdle and steeplechase racing in the UK Equine Vet J. 2004; 36: 384-9.

Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I:

instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships J Clin Epidemiol. 2009a; in press.

Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes Am J Epidemiol. 2009b; 169: 273-84.

Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology Epidemiology. 2000; 11: 550-60.

Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects Biometrika. 1983; 70: 41-55.

Rothman K, Greenland S, Lash T. Modern Epidemiology, 3rd Ed. Philadelphia: Lippincott Williams & Wilkins; 2008.

Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials Stat Med. 2007; 26: 20-36.

Sato T, Matsuyama Y. Marginal structural models as a tool for standardization Epidemiology. 2003; 14: 680-6.

Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. Med Care. 2007; 45: S143-8.

Senn S, Graf E, Caputo A. Stratification for the propensity score compared with linear regression techniques to assess the effect of treatment or exposure Stat Med. 2007; 26: 5529-44.

Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review J Clin Epidemiol. 2005; 58: 550-9.

Streiner DL. Finding our way: an introduction to path analysis Can J Psychiatry. 2005; 50: 115-22.

Stuart EA. Developing practical recommendations for the use of propensity scores: Discussion of 'A critical appraisal of propensity score matching in the medical literature between 1996 and 2003' by Peter Austin, Statistics in Medicine. Stat Med. 2008; 27:2062-65 .

Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods J Clin Epidemiol. 2006; 59: 437-47.

Suarez D, Haro JM, Novick D, Ochoa S. Marginal structural models might overcome confounding when analyzing multiple treatment effects in observational studies J Clin Epidemiol. 2008; 61: 525-30.

Susser M. Causal thinking in the health sciences: concepts and strategies of epidemiology. : Oxford University Press, Toronto (Out of print); 1973.

Terza JV, Bradford WD, Dismuke CE. The use of linear instrumental variables methods in

health services research and health economics: a cautionary note Health Serv Res. 2008; 43: 1102-1-20.

Thompson WD. Effect modification and the limits of biological inference from epidemiologic data J Clin Epidemiol. 1991; 44: 221-32.

VanderWeele TJ, Hernan MA, Robins JM. Causal directed acyclic graphs and the direction of unmeasured confounding bias Epidemiology. 2008; 19: 720-8.

VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework Epidemiology. 2007; 18: 329-39.

Weinberg CR. Can DAGs clarify effect modification? Epidemiology. 2007; 18: 569-72.

Yin L, Sundberg R, Wang X, Rubin DB. Control of confounding through secondary samples Stat Med. 2006; 25: 3814-25.