Statistics Session 8

CONFOUNDING AND STRATIFICATION

Many of the statistical methods and techniques that will be presented in this course are concerned with dealing with confounding. It is essential, however, that before these statistical methods are presented the basic concept of confounding is understood.

We will go on to demonstrate how to use the Mantel-Haenszel technique to obtain an odds ratio adjusted for a confounding factor. We will also perform a Mantel-Haenszel X^2 test to assess whether the adjusted odds ratio is significantly different to 1.

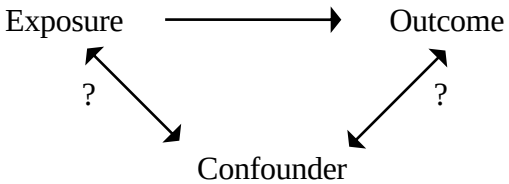
The final part of this session introduces the McNemar's test to compare two proportions when the observations in one sample are paired with the observations in the other.

References:
Hennekens and Buring (p287-293) & Chapter 12.
Rothman (2nd Edition) pages 59-62 for more details on confounding.
Breslow and Day Volume I (Chapter 3) for a deeper understanding.
Kirkwood, pp 96-102, p 180 for the Mantel-Haenszel techniques.

8.1 Reminder of confounding?

CONFOUNDING is about *alternative explanations*. Much of epidemiology is concerned with establishing associations between exposures and the risk of disease. For example in a previous session, the study found that the presence of storks was associated with the presence of young babies. We do not really believe that storks bring babies so we looked for an alternative explanation. Increased heating in the homes provides such an explanation. Houses with babies will tend to have more heating. Storks are likely to be attracted to warmer roofs. Heating is said to *confound* the relationship between the presence of storks and the presence of babies.

A confounder must be *associated with* the exposure and independently with the outcome. Neither of these relationships need be causal or statistically significant and they can go in either direction. What we are interested in is assessing whether an association we have found between an exposure and an outcome is likely to be causal or whether there are alternative explanations for it.
In general terms we are considering the relationship:



Example 1: A report was published that made the novel claim that coffee consumption is associated with the risk of cancer of the pancreas. Here, exposure is

coffee consumption, and outcome (disease) is cancer of the pancreas. The importance of this association was disputed because it was pointed out that coffee consumption was correlated with cigarette consumption (confounder), and cigarette smoking is believed to be a risk factor for cancer of the pancreas. Thus, cigarette consumption confounded the association between coffee and cancer of the pancreas.

It is because cigarette smoking has a real effect on cancer of the pancreas that it is credible as an alternative explanation for the reported association between coffee and cancer of the pancreas. There are many other things that are probably associated with coffee drinking. For any of them to provide a credible alternative explanation for the association between coffee and cancer of the pancreas, they would also have to be associated with the risk of cancer of the pancreas *independently of their association with coffee drinking*.

A confounder must be associated with both the exposure and the outcome. If cigarette smoking were only associated with coffee drinking but not pancreatic cancer, or only with cancer but not with coffee drinking then it would not act as a confounder in this relationship.

8.2 Dealing with confounding

In designing a study to determine whether there is a direct association between a particular exposure and a disease, one should anticipate the potential alternative explanations that might be offered for such an association. In other words one has to identify *potential confounders* at the outset. This means any factor which is thought likely to be associated with the disease and exposure(s) under study.. This is likely to include factors which are known to be causally linked to the disease (e.g. smoking and cancer of the pancreas), and factors that are good proxy measures of more direct causes (e.g. social class). Unless potential confounders are identified at the stage of study design, it will not usually be possible to reject alternative explanations for any association you may find.

Extending Example 1, confounding may be shown in numerical terms as follows:

Example 2: Suppose that the association between coffee consumption and cancer of the pancreas was detected in a case-control study, where the basic data was as follows:

	Coffee	No coffee	
Cases	450	300	
Controls	200	250	Estimated odds ratio = 1.9

8.3 Stratification

However, following the publication of this result, a reanalysis of the data was undertaken in which the study subjects were analysed according to smoking habit (the data were *stratified* by smoking habit). The new data were as follows:

	Non smokers		Smokers	
	Coffee	No coffee	Coffee	No coffee
Cases	50	100	400	200
Controls	100	200	100	50
Estimated odds ratios	=1.0		=1.0	

The reanalysis confirmed the suggestion that smoking confounded the association between coffee consumption and cancer of the pancreas. There is no effect of coffee on cancer of the pancreas seen for smokers or non smokers.

In any particular set of data, not all factors that are believed to have an effect on disease risk actually lead to confounding of the association between the exposure and the outcome under study. To be a confounder, a risk factor must also be correlated with the exposure to the factor under study.

In Example 2, it is clear that among the controls, smoking is correlated with coffee drinking:

	Coffee	No coffee
Smokers	100	50
Non smokers	100	200

Whereas only 1/5 of those who do not drink coffee are smokers, among the coffee drinkers 1/2 are smokers.

It should be noted that the association between smoking and coffee has been examined among the controls rather than the cases, or both taken together. This is, because controls represent the population from which the cases were drawn. In a cohort study the association would be looked at by constructing a similar table, replacing the number of controls with person-years at risk or alternatively numbers of persons at the start of the follow-up if a risk analysis was being undertaken.

Example 3: In a study of the effect of vaccination against pertussis, household contacts of cases were followed up. Among those aged under 6 the following results were obtained:

	Vaccine	No vaccine
Ill	63	125
Not ill	1108	699
	1171	824

Crude relative risk of disease in vaccinated compared to non vaccinated = 0.35 (95% CI 0.27-0.48). *Stratifying (or splitting)* the data by age gave the following results:

Age group	Vaccinated ill/total	Not vaccinated ill/total	RR
0	4/27	7/28	0.59
1	10/106	29/130	0.42
2-3	21/326	61/339	0.36
4-5	28/712	28/327	0.46

The crude estimate can be seen to be misleading - it is lower than that in any of the strata. A better estimate of the effect is obtained by taking a weighted summary estimate of the results from the different strata. This gives a relative risk of 0.41 (95% 0.31-0.56).

In this example there was confounding by age. *Adjusting for* age gave us a different estimate than that we obtained when considering only the totals (the crude estimate). Vaccination still has an effect after adjusting for age, but the magnitude of the effect is reduced. Age only partly explains the association seen.

So far we have considered confounding as fully or partially explaining an association that has been found: the relative risks after adjusting for the confounder were closer to one. Confounding can also act in the other direction. No association may be seen in the crude analysis, but after adjusting for confounding an association becomes apparent. This is sometimes called "negative confounding".

Example 4: The following data come from a study in Kenya. The association between marriage and schooling level was examined, using those with less than primary schooling as the baseline group:

	Women			Men		
Schooling level	% married	OR	OR adjusted for age	% married	OR	OR adjusted for age
< primary	75.0	1		49.5	1	1
Primary	73.1	0.91	0.76	54.7	1.2	0.59
Secondary or higher	70.9	0.81	0.28	61.9	1.7	0.31

For women there was little association between schooling level and marriage in the crude analysis, but a strong association became apparent after adjusting for age. For men the direction of the association was actually reversed by adjusting for age.

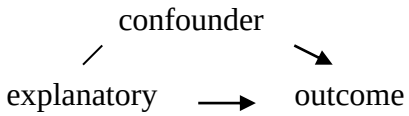
In general, confounding is present if the estimate obtained when you adjust for the potential risk factor is different from the estimate you get without adjustment. Of course estimates are likely to change slightly, and one problem is to know *how much difference matters*. This is a matter of common sense and experience. It is important to realise that there is no statistical test for confounding.

STEPS FOR DEALING WITH A POSSIBLE CONFOUNDER

1. Calculated crude X^2 and OR.

2. List possible confounders - socio-economic status, age, sex, season, village etc

3. Determine whether they are possible confounders: Are they
   - associated with explanatory variable
   - associated with outcome
   - not on causal path



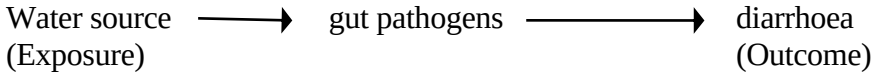
4. Do analysis stratified by possible confounder

5. Calculate pooled X^2 and OR (using Mantel Haenszel) i.e. look at association adjusted for confounder

6. If crude OR and pooled OR are different conclude that variable was a confounder.

8.4 Causal pathways

Example 5: Imagine a study to look at the association between source of drinking water and risk of diarrhoea. A risk factor for diarrhoea would be the presence or absence of gut pathogens in the water that was being drunk. However, it is clear that the presence of gut pathogens in the water does not provide an *alternative* explanation for any association between water source and diarrhoea. Instead, it may be regarded as an *elaboration* of the reason why water source may be associated with risk of diarrhoea. This example is one in which a factor (pathogens in the water) is on the *causal pathway* between the factor being studied (water source) and the disease outcome (diarrhoea).



Factors that are on a casual pathway between an exposure and a disease should not be regarded as confounding the association between the exposure and the disease, even if it is believed that they are directly related to disease risk. To control a factor that is on the casual pathway leads to an *underestimation* of the strength of the effect of the factor on the disease.

9.5 Interaction

Example 6: The following table shows data from the British doctors study on ischaemic heart disease mortality in non-smokers and current cigarette smokers by age.

Age	Annual death rate per 100,000 men (rate ratio compared to non-smokers)			
	Non-smokers	Current smokers (cigarettes per day)		
		1-14	15-24	25 +
< 45	7 (1)	46 (6.6)	61 (8.7)	104 (14.9)
45-54	118 (1)	220 (1.9)	368 (3.1)	393 (3.3)
55-64	531 (1)	742 (1.4)	819 (1.5)	1025 (1.9)

The rate ratios for mortality from ischaemic heart disease for the different levels of cigarette smoking compared to non-smokers are seen to be different in the different age groups.

In Example 3 we obtained a weighted summary estimate of the effect overall. This was a sensible thing to do because, although there were some differences between the strata they were not very large, and there was no evidence of a consistent trend. In Example 5 the differences between the strata are large and there seems to be a consistent trend with age: the adverse effect of smoking on ischaemic heart disease is less strong with increasing age. Where the effect of the exposure on the outcome is different for different levels of the confounder there is said to be *interaction* or *effect modification.* In these situations it is inappropriate to calculate an overall summary measure because important information would be lost. Statistical tests exist to assess whether the interaction is more than you would expect by chance. It is also helpful to think about whether the interaction seen makes biological sense. (Note that, in this example, although the rate ratio is much higher in the young, the rate difference is higher in the elderly. In terms of absolute death rates or numbers of deaths smoking could be said to be more important in the elderly.)

8.7 Residual confounding

Example 7. A case-control study compared *Helicobacter pylori* serology in patients with coronary heart disease and healthy controls. The authors suspected that chronic infection with *H pylori* might be a risk factor for heart disease. Since both heart disease and *H pylori* infection are associated with social class, social class was treated as a confounder. However, it is very difficult to measure social class accurately so several measures were used. The following results were obtained:

Effect of *H pylori*, adjusted for	Odds ratio (95% CI)
Unadjusted	2.28 (1.25-4.15)
Age	2.26 (1.15-4.44)
" + current social class	2.15 (1.07-4.29)
" + father's occupation	2.08 (1.03-4.20)
" + housing density, hot water in home	1.90 (0.91-3.97)

As each additional factor is adjusted for the effect diminishes a bit further. If they had only adjusted for current social class they might have stated that "there was an association that was independent of social class", but we can see from the further results that this would have been misleading. We cannot assume that even the bottom line where they have adjusted for everything that they have measured is the correct answer: perhaps if they had had more measures available, the association would have disappeared completely. This problem of incomplete adjustment for a confounder is common, especially when the confounder is difficult to measure, or when broad categories are used (eg broad age groups). There is said to be *residual confounding.*

8.8 Taking account of confounding

As we have said, confounding must be considered at the design stage of a study in order that appropriate information is collected. The effect of confounding can be minimised at the design stage and/or at the analysis stage.

Techniques used at the design stage are:

Restriction: Restricting a study to a subgroup will avoid confounding on that particular factor. For example, a study in women will avoid confounding by sex. However, within some subgroups, e.g age groups, there may still be some confounding, and restriction will reduce the numbers available for study and reduce generalisability.

Randomization: In a trial, properly conducted randomisation should avoid confounding, both by known and unknown confounders. However this is only possible in experimental studies. When numbers are small, chance variations between the groups may still give rise to confounding.

Matching: Subjects are selected in a way that ensures that potential confounders are equally distributed between the study groups. For example, in a study of the effect of dampness in bedrooms and acute respiratory infection, social class might be a confounder, so for each case a control could be chosen from the same social class. Matching tends to be expensive and time consuming and it is not possible to study the effect of the variable matched on. Adjusting for factors not matched on requires special techniques. On the other hand, it is very useful for adjusting for factors which are otherwise difficult to measure, such as neighbourhood. It is usually used only for factors which are likely to be very strong confounders (e.g. age and sex).

Techniques used in the analysis stage are:

Stratification: We have already seen examples of how stratification works. It is very useful when there are few confounders, but when there are many, the number of strata gets very large.

Standardization: This technique is used mostly to adjust for age and sex.

Logistic regression: This method allows adjustment for many confounders at once. It was used in Example 8, and will be outlined in a later session.

8.9 Further Analysis Of 2x2 Tables

We can do more that construct two-way tables to examine the joint distributions of two categorical variables. We can perform a X^2 test to assess the evidence for an association between those variables. We also saw that obtaining a weighted average of the odds ratios gives a better estimation of the adjusted association, but we need to formalize how to do this.

Another (hypothetical) case-control study of the effect of coffee consumption and cancer of the pancreas yielded the following results.

		case	control
coffee	yes	450	440
	no	300	410
	Total	750	850

The crude OR for the association between coffee and cancer of the pancreas is ad/bc = 450x410/440x300 = 1.40. This suggests an risk of cancer of the pancreas associated with coffee consumption. There are several possible explanations for these findings:

i) Chance: that the observed association between coffee consumption and cancer of the pancreas arose by chance. A chi-square test on this table gives $X^2 = (|ad-bc| - N/2)^2 \times N / efgh = 10.62$, p=0.001 so chance is an unlikely explanation for these findings.

ii) Bias: that the observed OR of 1.40 is biased and does not represent the true OR.

iii) Confounding: that the OR of 1.40 was real but was due to the effect of another variable. For example, it may be that those who consume coffee were more likely to smoke, and smoking (rather than coffee consumption) put them at an increased risk of cancer of the pancreas. This is called confounding and in this session, we will use the Mantel-Haenszel method to adjust for confounding.

iv) Causation: that coffee consumption increases the risk of cancer of the pancreas, and the estimated OR of 1.40 is accurate.

Remember:
A confounding factor is one that is related to both the response and explanatory variables. A further condition for a confounder is that it does not lie on the causal pathway between the explanatory and response variables. Common examples of confounding variables are age, sex and season. Ignoring confounding, when assessing the association between a response variable and an explanatory variable, can lead to spurious results.

In this example, it is possible that the association between coffee consumption and cancer of the pancreas arose because of the confounding effect of smoking. It is possible that the effect of coffee consumption on the risk of cancer of the pancreas varies between smokers and non-smokers. The results shown so far have been for all subjects (smokers and non-smokers) combined. In order to assess whether smoking is a confounder, we want to look at the

association between coffee consumption and cancer of the pancreas separately for smokers and non-smokers:

		Smokers		**Non-smokers**	
		case	control	case	control
coffee	yes	400	340	50	100
	no	200	190	100	220
	Total	600	530	150	320

We notice that among the non-smokers, the prevalence of coffee consumption is very similar in cases (50/150=33%) and controls (100/320 = 31%), and that among the smokers, the prevalence of coffee consumption is similar in cases (400/600=67%) and controls (340/530=64%). However, when the smokers and non-smokers are combined, there is a larger difference in the prevalence of coffee consumption in cases (450/750=60%) and controls (440/850=52%). The difference in these results arose because of two factors:

a) Smoking is associated with being a case (80% of cases and 62% of controls are smokers).

b) Smoking is associated with coffee consumption (among the controls, 31% of non-smokers and 64% of smokers consume coffee).

Since smoking is associated both with the exposure (coffee consumption) and with the outcome (cancer of the pancreas), smoking is a confounding factor. This is an example where we see the misleading effects of combining dissimilar data. The effect of combining smokers and non-smokers suggests an association far stronger than really exists (confounding). In other situations, combining in the presence of a confounder may mask a difference that really exists (negative confounding), or even show a difference opposite to the one that exists.

8.10. The Mantel-Haenszel odds ratio

What follows will refer to odds ratios rather than relative risks, although the analogous formulae for relative risks are given in the Appendix. Since smoking is a confounder, we should not combine the smokers and non-smokers for the analysis. The OR of 1.40 obtained from the combined table is not appropriate. It would be appropriate to calculate separate ORs of coffee consumption for smokers and non-smokers. Among the smokers, OR = 400x190/340x200 = 1.12and among the non-smokers, OR = 50x220/100x100 = 1.10. However, we may not really be interested in the results in these two groups, but would rather have a single overall result that takes account of the confounding factor. We would obtain a better estimate of the overall OR by taking an average of the ORs for the two groups (1.12 and 1.10). The Mantel-Haenszel odds ratio, denoted OR_{MH}, gives a weighted average of the ORs in the different strata, where the ORs from larger strata are given more weight. In this example, the two stratum-specific ORs are equal so the OR_{MH} will be equal to the stratum-specific ORs.

To calculate the OR_{MH}, first construct the 2x2 tables of exposure by disease for the separate strata of the confounder:

		STRATUM 1			STRATUM 2		
		disease			disease		
		Y	N		Y	N	
exposed	Y	a_1	b_1	$\|e_1$	a_2	b_2	$\| e_2$
	N	c_1	d_1	$\|f_1$	c_2	d_2	$\| f_2$
		---	---	---	---	---	---
		g_1	h_1	$\| n_1$	g_2	h_2	$\| n_2$

Then OR_{MH} is obtained as follows:

$$OR_{MH} = \frac{a_1d_1/n_1 + a_2d_2/n_2}{b_1c_1/n_1 + b_2c_2/n_2}$$

In our example:

$$OR_{MH} = \frac{(400x190)/1130 + (50x220)/470}{(340x200)/1130 + (100x100)/470} = 1.11$$

So, the odds ratio for coffee consumption adjusted for smoking is 1.11. Compare this to the odds ratio from the combined table (usually called the crude OR) of 1.40. Adjusting for smoking gives an OR much closer to 1, compared to the crude OR, which means that the risk associated with coffee consumption is not as strong as we initially thought (OR=1.11 versus OR=1.40).

The Mantel-Haenszel method of controlling for confounding is similar to the method of standardization used to control for age in the Session "Controlling For Age". In this session, we have obtained an "adjusted odds ratio" rather than a "standardized rate". The Mantel-Haenszel method can be used to obtain an adjusted relative risk and an adjusted rate ratio (denoted RR_{MH}) by calculating a weighted average of the RRs in the different strata. The formula for calculating a Mantel-Haenszel relative risk is given in the Appendix. These methods are referred to as stratified analyses because we look at an exposure by a response for the different strata (levels) of a confounder.

8.11 The Mantel-Haenszel chi-square test

The Mantel-Haenszel chi-square test is used to determine whether OR_{MH} (or RR_{MH}) is significantly different to 1. The test is rather like the usual chi-square test. The null hypothesis is that there is no association between the exposure and the disease within any of the individual strata. In order to perform this test, we must first obtain the following from each table:

i) The observed value of a.

ii) The expected value of a, assuming the null hypothesis of no association: $E_a = eg/n$.

iii) The variance of a, assuming the null hypothesis of no association:

$V_a = efgh/\{n^2(n-1)\}$.

We then sum each of these quantities over all the tables. In our example, we obtain:

	a	E_a=eg/n	V_a=efgh/{n^2(n-1)}
Smokers	400	392.9	63.7
Non-smokers	50	47.9	22.2
Total	450	440.8	85.9

If the null hypothesis is true, then we would expect the difference between our observed and expected values to be small. We test whether the differences obtained are greater than would be expected by chance, this time using the difference between Σa and ΣE_a. We calculate:

$$X^2 = (| \sum a - \sum E_a | - 0.5)^2 / \sum V_a$$

and obtain a P-value by referring our result to the X^2 distribution with 1 degree of freedom.

In our example, $X^2 = (|450 - 440.8| - 0.5)^2 / 85.9 = 0.88$ on 1 df. This gives P>0.30 from which we conclude that after adjusting for smoking, there is no evidence of any association between coffee consumption and cancer of the pancreas. So, $OR_{MH} = 1.11$ is not significantly different to 1.

8.12 Comparison of proportions from paired data

In this and other sessions, we have demonstrated various statistical tests for the analysis of categorical data. These methods are only appropriate for unpaired data, that is, when the observations in one group are not paired or related in any way to the observations in another group. For example, the cancer of the pancreas cases were not related or matched in any way to the controls. Examples of paired data are a matched case-control study, "before and after" measurements, and comparisons between two observers. Paired data is analysed using techniques which take the pairing of the data into account.

The table on the bottom of the page shows the results of a matched case-control study of the effect of exogenous oestrogens on the risk of endometrial cancer (taken from Statistical methods in cancer research, volume 1, on pp 162-168, N Breslow and N Day). The 63 cases of endometrial cancer occurring in a retirement community in Los Angeles were each matched to a control of the same age (within one year) and marital status, who had entered the community at the same time and had not had a hysterectomy prior to the time when the case was diagnosed,

and who were therefore still at risk of the disease. The table directly below shows the exposure status (exposed is "ever having taken any oestrogen") of the 63 case-control pairs:

		control		
		exposed	not exposed	Total
case	exposed	27	29	56
	not exposed	3	4	7
	Total	30	33	63

Since we are interested in comparing the proportion exposed to exogenous oestrogens among the cases and controls, we might have summarised the exposure status of the 126 subjects as follows:

	cases (n=63)	**controls** (n=63)	**Total**
exposed	56 (89%)	30 (48%)	86
not exposed	7 (11%)	33 (52%)	40
Total	63	63	126

We can see that 89% of the cases, but only 48% of the controls, were exposed to exogenous oestrogens. However, it would be incorrect to calculate the ordinary odds ratio and perform an ordinary X^2 test on the above table, since the table represents 63 pairs (63 cases each matched to a control), and the ordinary odds ratio and X^2 test take no account of the paired nature of the data. (Had the cases and controls been unmatched, then the above table and ordinary odds ratio and X^2 test would be appropriate).

Instead we rearrange the data as follows, showing how many of the case-control pairs were concordant (both case and control exposed, or both case and control not exposed) and how many were discordant (case exposed and control not exposed, or case not exposed and control exposed):

		control		
		exposed	not exposed	Total
case	exposed	27	29	56
	not exposed	3	4	7
	Total	30	33	63

We find that 27 case-control pairs were both exposed and 4 case-control pairs were both not exposed. These 31 pairs give us no information as to whether the exposure is associated with being a case. Of the remaining pairs, 29 had an exposed case and non-exposed control, and 3 had an exposed control and non-exposed case. If there was no association between the exposure

and being a case, then we would expect similar numbers in these two groups. We use McNemar's X^2 test to assess whether the observed difference between these two numbers is greater than we would expect by chance.

If we call the number of (case exposed, control not exposed) pairs r and the number of (case not exposed, control exposed) pairs s, then McNemar's X^2 is calculated as:

$$X^2 = (|r - s| - 1)^2/(r + s)$$

and is distributed as X^2 on 1 degree of freedom.

In the example, r = 29 and s = 3 so that:

$$X^2 = (|29-3| - 1)^2/(29+3) = 25^2/32 = 19.53 \quad (1 \text{ df}).$$

Thus, from the X^2 tables we obtain $P<0.001$ and we can conclude that there is strong evidence that being exposed to exogenous oestrogens is associated with an increased risk of endometrial cancer.

A matched odds ratio that measures the effect of exposure in a matched case-control study is calculated as r/s. In this example, the matched odds ratio = 29/3 = 9.67, which suggests almost a tenfold increased risk.

APPENDIX

The Mantel-Haenszel method can be used to obtain an adjusted relative risk and an adjusted rate ratio. To calculate the Mantel-Haenszel relative risk (RR_{MH}), first construct the 2x2 tables of exposure by disease for the separate strata of the confounder. Then:

$$RR_{MH} = \frac{a_1(c_1+d_1)/n_1 + a_2(c_2+d_2)/n_2}{c_1(a_1+b_1)/n_1 + c_2(a_2+b_2)/n_2}$$

STATA can be used to obtain a Mantel-Haenzsel odds ratio (see Practical).

Notes on the use of the Mantel-Haenszel test

Like the X^2 test that we have already met in Session 8, this is an approximate test, and should not be used if the sample size is very small. However, the procedure for assessing the adequacy of the sample does not depend on any individual table, but on the tables taken together. Details are given in Kirkwood (pp 96-102).