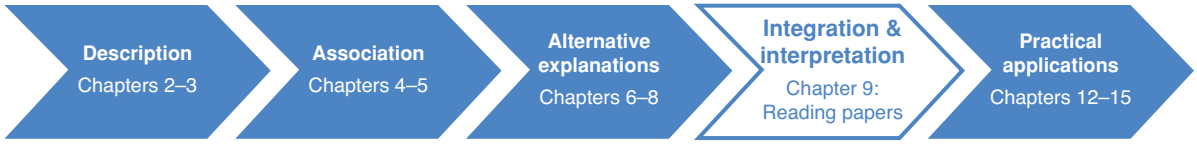
9

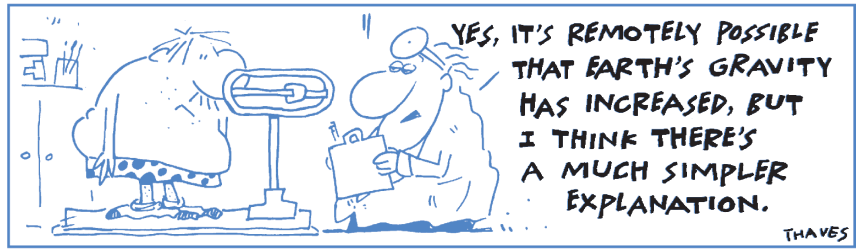
Reading between the lines: reading and writing epidemiological papers



In Chapters 4, 6, 7 and 8 we looked at the different epidemiological study designs and examined the various misfortunes that can befall them. Good studies are difficult to design and implement, and interpretation of their results and conclusions is not always as straightforward as we might hope. How, then, can we make the best use of this information? The central question we have to answer when we read a study report is '*Are the results of the study valid?*' If the authors report an association between exposure and outcome, is it real? If they find nothing,

do we accept this? Or could there be an alternative explanation for the results, namely chance, bias and/or confounding?

Frank and Ernest



Much of the following discussion will pick up and integrate the core epidemiological issues covered in the previous chapters. We will concentrate mainly on analytic studies looking for associations between 'cause' and 'effect', the study designs that you met in Chapter 4, but the same general principles apply equally to descriptive epidemiology. To extract the maximum information from a paper we need a systematic approach to identifying its strengths and weaknesses. Some quite detailed sets of guidelines for 'critical appraisal' of the health literature exist already and we do not intend to add to this list (although we do offer a flowchart for more general guidance). Instead we will focus on the essence of the challenge: what are the practical effects of the ways in which subjects were selected and information collected, and the likely influence of confounding and chance on the results we see? While the elements of the general strategy we propose are universal, the approach can (and should) be tailored to suit your own personal style. In practice you will almost certainly have to read individual papers and reports and, if you are involved in research, you may write some of your own. Both activities demand a very practical approach and this is what we will focus on here. We will emphasise the perspective of the reader, but the writer should be thinking about exactly the same things, since good writing demands that the readers' needs and perspectives are kept firmly in mind.

The research question and study design

When reading a paper, the first step is to identify the *research question* that the authors set out to answer and then the strategy they used to attempt to answer that question. Was the *study design* appropriate to answer the question posed? This involves consideration of what the *ideal* type of study would be and also what would be *practical* in that particular situation.

As you have seen, the *ideal* study to answer a question of cause and effect would often be some sort of randomised trial, but in many situations this will be impossible for numerous ethical and/or practical reasons. Next best would

generally be a cohort study in which exposure is measured prior to the development of disease, but again the resources, time and money required to conduct a large enough study often make it unfeasible. So, from a practical viewpoint the key question should be 'Was the research design the best that could have been done in the circumstances to answer that particular question?' If it was not the best, can it still provide useful information? Are there existing studies addressing the same issue that were of better design against which the findings of the current study can be compared?

Many studies are conducted not because they will provide the strongest possible evidence for a causal association between exposure and outcome, but because they can answer a range of other more indirect questions of interest. For example, the results from the ecological study of *Helicobacter pylori* infection and stomach cancer rates in China shown in Figure 3.7 cannot directly answer the question 'Does *H. pylori* infection cause stomach cancer?' We can, however, answer the question 'Are stomach cancer rates higher in areas where *H. pylori* infection is more common?' Trials, and to a lesser extent cohort studies and case–control studies, can address issues of causality more directly, whereas other types of study provide more circumstantial evidence, but if the results are valid each can increase our understanding of the relation between an exposure and outcome. As an example, ecological and migrant studies conducted across countries with widely differing levels of solar ultraviolet (UV) radiation have consistently revealed an association between sun exposure in childhood and melanoma rates. In contrast, case–control studies, which have generally been conducted within a single country or region with a narrow range of UV exposures, have not given consistent results (Whiteman *et al.*, 2001). In this particular situation ecological studies with their wide variety of exposure levels provide a valuable additional perspective to the case–control studies.

So how do we decide whether the results of a study are valid? We have to consider the three main alternative explanations that we discussed in the preceding chapters: bias (both the *selection of participants* for the study and the *information* that was *measured* or collected from or about them), confounding and chance.

The study sample: selection bias

Who was included in the study, how were they selected and are there possible sources of selection bias? Specific questions to ask when reading a paper include those below.

- Is the comparison group appropriate?
  In a case–control study are the controls really representative of the population from which the cases arose? In a cohort study where the comparison cohort

was recruited separately from the exposed cohort, are the two groups really comparable?

- What proportion of eligible participants actually took part in the study and, if appropriate, what proportion was lost to follow-up?
  Participation or follow-up rates less than 90% (some would say 80%) may be cause for some concern. If the rates are lower than this, could participation (or loss to follow-up) be related to either the exposure or the outcome of interest? That is, could those who refused to take part (or who were lost to follow-up) have differed in some way from those who did take part? If so, might this have led to an overestimation or underestimation of the level of exposure and/or outcome? Most importantly, could this have differed between study groups?
- Finally, what is the likely effect of any selection bias on the results of the study?
  Ideally the authors of the paper will have considered all of these issues in their discussion, but if they have not then it is up to the reader to decide whether bias might be present and, if so, what effect it may have had on the results. In practice there will almost certainly be some potential for selection bias. Participation rates are never 100% and in many developed countries it is becoming increasingly hard to persuade people to take part in research, especially when they see no benefit to themselves. This is a major issue in case–control studies when the motivation for a 'case' to take part may be much greater than that of an unaffected 'control'. Also, people are becoming increasingly mobile, so follow-up in a cohort study that runs for more than a few years is never likely to be 100%.

If we were to reject all studies with less than 100% participation or follow-up rates, we would be left with nothing to review. In practice, participation or follow-up rates greater than 80% or 90% are generally considered to be good, but rates lower than this do not necessarily invalidate the findings (see Example 2 below). The challenge for both investigator and reader is to think practically and to decide whether any potential biases related to selection might have compromised the study results (the **internal validity**) and, if so, how and to what degree the results might be biased. It is often impossible to quantify this, but **sensitivity analyses** making various assumptions about the size and direction of possible bias can be informative (see Chapter 7).

Example 1: case–control studies of blood transfusion and Creutzfeldt–Jakob disease

In five case–control studies of Creutzfeldt–Jakob disease (CJD) the controls were more likely to report having had a blood transfusion than cases (Riggs *et al.*, 2001). Does this tell us that blood transfusions might protect against CJD (a finding contrary to the causal hypothesis)? If we consider the control groups, we find that in three of the five studies they were selected from among hospitalised

patients and in another study more than 12,000 telephone calls were made in order to recruit just 784 controls.

The use of hospital controls and the very low participation rate among controls should ring alarm bells. Why?

People who are in hospital are more likely to have had a blood transfusion than those who are not; and in addition, given the publicity surrounding 'mad cow disease', people who have had a blood transfusion may well be more likely to agree to take part in a study of CJD. Indeed, in these four studies approximately 20% of controls reported having had a blood transfusion – an improbably high proportion, probably due at least in part to these selection pressures. So what can we conclude about the association between transfusion and CJD from these studies? Not much. The high transfusion rate in controls almost certainly overestimates the base rate in the population from which the cases came. We have no idea whether the true background rate is similar to that in cases (i.e. there is no association) or lower than in cases (i.e. there is a positive association). Our next example shows how external information can help resolve such dilemmas.

Example 2: a case–control study of oesophageal cancer and smoking in Australia

In an Australian case–control study of oesophageal cancer, the authors considered the relation with smoking. In this study approximately 70% of eligible cases but only 49% of the controls who were contacted agreed to participate – this is a fairly typical response rate in many countries these days, but is far from ideal. The authors found that current smoking rates were higher among cases with oesophageal adenocarcinoma than controls (OR compared to never smokers = 2.7; 95% CI 1.9–3.9), but could this be due to selection bias?

In general, smokers are less likely to agree to take part in a study than non-smokers. What effect might this have had on the odds ratio?

If smokers were less likely to take part the prevalence of smoking in the control group would be *lower* than that in the general population. This would exaggerate the difference between cases and controls and so increase the odds ratio, making it look as if smoking is associated with oesophageal adenocarcinoma when in reality it might not be. To address this issue the authors used data from a National Health Survey conducted at about the same time to estimate the likely prevalence of smoking in the controls who did *not* agree to take part in the study. If they assumed that the whole control population had a smoking rate equal to that seen in the national survey, they found that the odds ratio for the association between smoking and oesophageal adenocarcinoma was slightly weaker but still significantly greater than 1.0 (imputed OR = 2.4; 95% CI 1.7–3.4). This suggested that even though only about half of the controls invited to take part

in the study actually agreed to participate, the overall results for the association with smoking were not seriously biased (Pandeya *et al.*, 2009).

Measuring disease and exposure: measurement bias

We also have to consider the information collected from or about the people in the study – particularly the measurement of 'outcome' and 'exposure' but also measurement of other factors that might be important confounders. Attention to unbiased measurement of *outcome* is crucial for cross-sectional, cohort and intervention studies. It is of relatively less importance in a case–control study, in which cases are selected because they have already experienced the outcome of interest (although a clear definition of what constitutes a case is still essential). Accurate measurement of *exposure* is important in every study, and in a case–control study it is critical to ensure that there are no systematic differences in measurement between cases and controls. Good measurement of *confounders* is often overlooked, but this is also essential to enable optimal control of confounding in the analysis (see comments on residual confounding in Chapter 8).

Some questions to ask when reading a paper are the following.

- Were the outcome/exposures/confounders clearly defined, and how were they measured?
- Have all relevant outcomes and/or exposures and/or confounders been included and, if not, how important are those omitted?
- Were the same definitions and methods of measurement used in all of the study groups?
- Is measurement error likely to be a problem and, if so, could there be **non-differential misclassification** (look back to Chapter 7 if you are unsure about this)?

No measurement is perfect and some measurements are very poor. The effect of the ubiquitous random error and consequent non-differential misclassification must always be considered. The practical implication of this is that effects (OR, RR) estimated in the face of equal measurement error in the compared groups will usually appear *weaker* than they truly are. Thus a finding of a positive association, despite poor measurement, should not be dismissed because of this – the true association is likely to be more impressive. On the other hand, a null finding or a very weak effect in the presence of non-differential misclassification is uninformative since it may reflect the imprecise measurement (thereby masking a true association) or there may truly be no effect.

- Is the extent of any measurement error likely to differ between groups (e.g. could there be **recall** or **interviewer bias** in exposure measurement in a case–control study) and so could there be **differential misclassification**?

Differential misclassification can bias results in either direction. It is particularly important to consider this possibility in cross-sectional and case–control studies when exposure is measured after the outcome has occurred. In analytic research it is generally easier to distinguish clearly between outcome states (diseased versus non-diseased) than it is to measure exposures precisely, but the avoidance of differential outcome assessment is central to the integrity of cohort studies and trials, and again for cross-sectional studies.

- Finally, what practical effects might any measurement bias (outcome or exposure) have had on the results of the study?

Example 3: a case–control study of body-mass index (BMI) and asthma in Mexico

A significant association between asthma and obesity (defined as BMI > 30 kg/m^2 based on self-reported weight and height) was observed among women (adjusted OR = 1.7; 95% CI 1.1–2.7), with a weaker non-significant association (adjusted OR = 1.3; 95% CI 0.6–2.9) seen among men (Santillan and Camargo, 2003); but how reliable are self-reported data on body size and could measurement error have affected the results? The authors specifically addressed this question by weighing and measuring all of the participants. They found that, on average, people tended to report that they were taller and lighter than they really were, particularly the men. As a result, the *true* prevalence of obesity based on measured BMI was higher than that based on self-reported BMI and the difference was somewhat greater for cases (40% versus 24% for men and 44% versus 38% for women) than for controls (28% versus 22% for men; 24% versus 23% for women).

Is the error in the self-reported information on body-size differential or non-differential?

Assuming that the measured BMI values are correct, is the *true* association between obesity and asthma likely to be stronger or weaker than that seen for self-reported obesity?

In this example there is *differential* error since cases, particularly men, were more likely to underestimate their weight and overestimate their height than controls. The effect of these errors would be to reduce the association seen and this is what happened. When the authors calculated the association between asthma and *measured* obesity, the OR was 2.3 (95% CI 1.5–3.8) for women and 2.5 (95% CI 1.1–5.9) for men, i.e. the associations were much stronger than those based on self-reported BMI above. (Note that although the OR based on measured BMI is likely to give a more accurate estimate than that based on self-reporting, even this may be an underestimate of the 'true' effect since there is still likely to be some *non-differential* random misclassification.) Validation studies such as this

Box 9.1 Sensitivity analysis: kangaroos and Ross River virus (RRV) infection

Authors of a case–control study found an odds ratio of 4.3 (95% CI 0.9–21) for the association between seeing kangaroos in the backyard and risk of RRV infection, possibly because kangaroos provide a host for the mosquitoes that spread RRV. However, information was missing for a number of cases and controls, so the authors performed a sensitivity analysis.

If they assumed that all cases and controls for whom exposure data were missing had seen kangaroos, the OR was 1.9 (95% CI 0.7–5.1); if they assumed that none had seen kangaroos the OR was 3.5 (0.9–14.1). If they assumed that cases were more likely to remember exposure than controls (i.e. there is recall bias) and, therefore, that cases with data missing had not seen kangaroos whereas controls had seen them, then the association disappeared completely (OR = 1.0, 95% CI 0.4–2.8). This analysis raised questions about the validity of the observed association between sighting of kangaroos and RRV infection.

(Harley *et al.*, 2005)

can provide valuable insights into the accuracy of study results, as can sensitivity analyses such as that described in Box 9.1.

Confounding

The next major issue to consider is that of confounding.

- Have the authors considered all important confounders and controlled for them in their analysis?
- Could there be residual confounding by variables that have not been considered or because of incomplete adjustment for factors that have?
- If so, what effect might this have had on the study results?

Again, the important thing is to think practically: in which direction is any residual confounding likely to operate? If when the authors adjusted for confounding the association became *stronger* (i.e. the confounding had originally biased the effect towards the null) then, if there is residual confounding, the real effect is likely to be even more extreme than that observed. Conversely, if the adjustment brought the estimate *closer* to 1.0 (i.e. the confounding had biased the estimate away from the null) then the true result may be even closer to the null than that reported. In the latter situation our confidence in the value of a positive effect estimate would decrease, unless it was very large. A large effect is less likely to be wholly due to confounding because, to explain away a very strong RR (e.g. 10.0), the confounder itself would have to be an even stronger risk factor for the

disease. If this is the case then it is likely to be known already, and hence should have been measured and controlled for.

Example 4: a cross-sectional study of risk factors for depression in the UK

Among 14,217 adults aged over 75 years, the risk of depression appeared to be somewhat higher among women than among men (crude OR = 1.3, 95% CI 1.1–1.5) (Osborn *et al.*, 2003). After adjustment for potential confounding factors including age, marital status, living alone, smoking and alcohol consumption, the adjusted OR was 1.1 (95% CI 1.0–1.3).

What do these results suggest about the association between sex and risk of depression?

The adjustment has reduced the OR, bringing it closer to 1.0. It is also likely that there is further residual confounding, which might bring the true OR even closer to 1.0, suggesting that sex is not associated with depression (at least in this study). This example also highlights the need to consider the clinical or practical significance of the results of a study. A very large study can show what appears to be a very small effect with great precision (a narrow confidence interval); even though the result might be statistically significant ($p < 0.05$) the key question is whether such a small difference is meaningful.

Example 5: a cohort study of statin use and atrial fibrillation in the USA

A cohort of patients with coronary artery disease was followed for a minimum of 12 months to document the incidence of atrial fibrillation (AF, an abnormal heart rhythm); 263 of the patients were using statins (cholesterol-lowering drugs) and 186 had never used them (Young-Xu *et al.*, 2003). (Note that this was an observational study, not a randomised trial.) Overall, the rate of AF was lower among the group taking statins, giving a crude relative risk of 0.5 (95% CI 0.3–0.8). When the authors adjusted for potential confounding factors including age, systolic blood pressure, alcohol consumption, history of heart failure and total serum cholesterol level, the RR was 0.4 (95% CI 0.2–0.8).

Assuming that there are no important selection or measurement errors, what conclusions can we draw about the association between statin use and AF?

It appears that there was some confounding by the other factors such as age since the RR dropped from 0.5 to 0.4 after adjustment, indicating that the real effect of statin use was even stronger than the crude RR suggested. However, doctors prescribe treatment partly on the basis of prognostic judgements, which are difficult to measure. There may thus be other unknown and unmeasured confounders that have not been controlled for, so we would still need to be cautious about this particular result which, as you saw in Chapter 8, could be due to **confounding by indication**. Large, well-conducted RCTs remove this potential problem.

Chance

Finally, it is important to consider the role of chance. Have the authors included confidence intervals for their estimates? How narrow (good precision) or wide (poor precision) are they? If an association is seen, how likely is it that there is really no effect (i.e. the association arose by chance)? If there is no clear association (e.g. if the confidence limits are very wide and include 1.0), is it possible that there is a real effect but the study was simply too small to detect it? Is the study useful or are the results inconclusive? As well as *statistical significance* it is also important to consider whether the results are *socially* or *clinically* significant (see *Statistical versus clinical significance* in Chapter 6). A large study may give an association that is statistically significant, for example the odds ratio of 1.1 (95% CI 1.0–1.3) seen for the association between sex and depression in Example 4 above, but we would then have to ask whether a 10% higher risk of depression in women than men was a meaningful difference.

Study validity

Once we have considered all of these aspects (summarised in Figure 9.1) we can make an overall judgement of the validity of the study results. There are two separate issues here. The first and most important, often called **internal validity**, is the extent to which the results of a study reflect the true situation *in the study sample* in the absence of any alternative explanations. These alternative explanations, namely chance, bias and confounding, have been the focus of this and the previous chapters. *The prime objective of study design, implementation, analysis and interpretation is to maximise the internal validity of a study.* The second issue is one of **generalisability** or **external validity**. Are the results of a study applicable to populations other than the study population?

Internal validity

Have the authors discussed the limitations of their study? What conclusions do they draw with respect to the research question? Are these conclusions justified? Does the study appear to be internally valid or could the results be due to chance, bias or confounding? It is important to remember that in public health we are dealing with real people and complex exposures that are often difficult to measure and/or impossible to control adequately and we are, quite rightly, constrained as to what we can do by codes of ethics. Any study is thus likely to fall short of perfection and it is important to realise this. Research should be appraised in the light of what it has been able to achieve – there will be

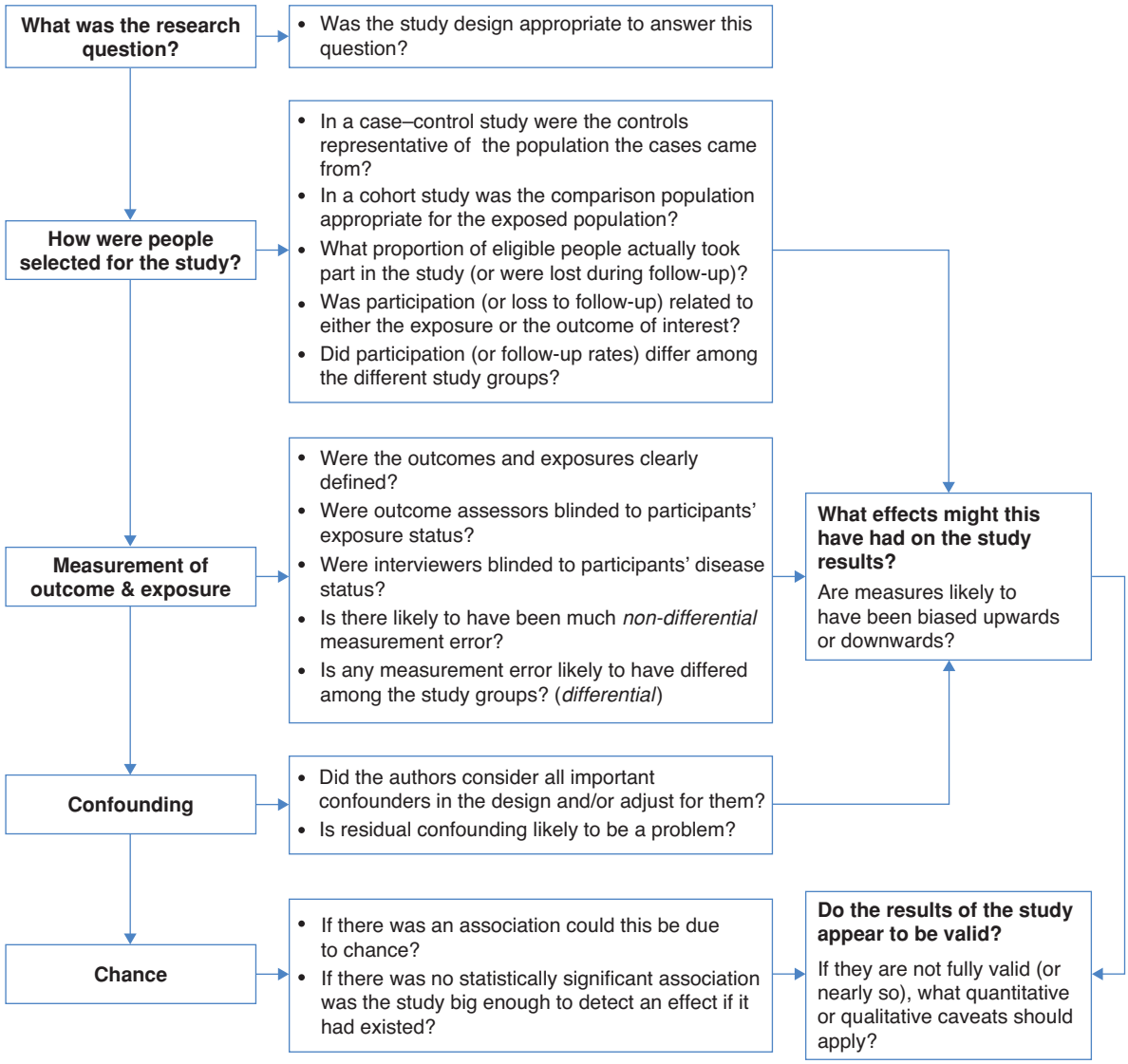


Figure 9.1 Issues to consider when reading epidemiological papers.

deficiencies but, given the particular circumstances, could things realistically have been improved? The evidence reported in a research paper might not always be strong but, if it is the best that is likely to be available, we should not discount it because of the flaws. Rather we should draw from it what information we can that bears on the question in hand.

External validity

It is important to remember that the aim of the 'causal arm' of epidemiology is to discover general scientific truths about cause and effect. Can the results of a study of, for example, American men aged 50–65 be generalised to older or younger men? Women? Non-Americans? (Note that such a question presumes the internal validity of a study: *if a study is not internally valid, then the results should not be applied to anyone.*)

There are no firm rules to help with generalising from a study to the wider population. In case–control comparisons, population-based studies are the ideal in order to reduce the possibilities of selection bias and, as a result, it might not require such a leap of faith to extrapolate the results from one population to another. However, the process is not simply a matter of statistical representativeness, but is more fundamentally one of biological insight. The question then is 'How relevant (biologically) is a result for a given population?' Can a study in a very select population (e.g. urban-living Japanese, Czech women, Brazilian men) inform us about disease causation more generally? Well, we certainly hope so. As an example, careful follow-up of the survivors of the atomic bomb blasts in Hiroshima and Nagasaki, Japan, has yielded volumes of information regarding the relation between exposure to ionising radiation and subsequent risks of mortality, cancers and other rare diseases. While this information comes only from the Japanese, no one would argue that radiation would not have similar effects in other nationalities, and we certainly do not want to see this 'unnatural experiment' repeated. While this generalisation is perhaps easier than many because of the magnitude and timing of the effects and the well-understood physical and biological properties of ionising radiation, the principle is identical for other abstract causal speculation.

Generalising from clinical and other trials raises additional issues. For practical reasons, many clinical trials are conducted on highly selected groups of people who are almost certainly not representative of the general population. This can make the results of the specific trials easier to interpret (internal validity), but means that they can be harder to generalise to other groups (see Chapter 11).

Descriptive studies

The discussion above has focused on papers evaluating associations between exposure and outcome and that, therefore, address the 'Why?' of epidemiology. It is equally important to evaluate the results of descriptive studies that provide the 'Who?', 'Where?' and 'When?' information that is essential to make a community diagnosis and, as you will see in Chapter 14, also play an important role

in evaluating the effects of public health interventions. In practice this requires us to consider exactly the same issues: selection and measurement error, confounding and chance.

- How was the survey sample selected? Is it representative of the wider population?
- How was the factor of interest measured? Is it likely to be over- or under-reported?
- If we are making comparisons, are we comparing like with like or is there a need for standardisation (to remove confounding by, e.g., differences in the age structure of populations)?
- Could any observed excesses (or deficits) of disease in different populations, in different places or at different times be due to chance? For example, it is unlikely that several cases of a rare disease would occur in the same small community (what is known as a 'cluster' – see Chapter 12), but it is not impossible for this to occur by chance. Similarly, rates of disease (particularly rare diseases) will naturally vary from year to year, so could an apparent increase or decline just be due to chance?

It is also important to note that, although *representativeness* is not the primary issue in studies of aetiology, it is crucial for applying the results of a descriptive study to a wider population. If a sample of people is surveyed to identify the health needs of an area then, if those participating do not represent the whole population, the results could be very misleading. If, for example, they were unusually healthy then the needs of the population might be greatly underestimated, and vice versa.

Writing papers

We have focused on the information that you need to look for when reading a paper and, as we suggested at the start of this chapter, it goes without saying that this is also the information that you need to provide when writing a paper. To improve the reporting of experimental research, some journals now require that authors follow the checklist of points contained in the Consolidated Standards of Reporting Trials or CONSORT statement (Moher *et al.*, 2001). This document has since been modified to give the TREND (Transparent Reporting of Non-randomised Designs) checklist for reporting results of studies of behavioural and public health interventions with non-randomised designs (Des Jarlais *et al.*, 2004). Similar guidelines have been developed for observational studies including the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement (von Elm *et al.*, 2007), and the STREGA (STrengthening the REporting of Genetic Association Studies) statement, a modification of STROBE for genetic studies (Little *et al.*, 2009), as well as a guide specific to longitudinal studies (Tooth *et al.*, 2005).

Box 9.2 The problem of multiple testing

The more hypotheses that we test, the more likely it is that some apparently statistically significant results will arise by chance. For this reason statisticians often recommend 'correcting' for this problem of multiple testing. A simple form of this is to reduce the α-level at which a result is considered to be statistically significant based on the number of tests performed. For example, if 20 separate tests are conducted within a single study then the p-value at which a result is considered statistically significant would be reduced from 0.05 to $0.05 \div 20 = 0.0025$. The net result is that fewer results, those with the strongest associations, will be deemed statistically significant and, hopefully, these are also the results that are less likely to be due to chance. However many epidemiologists have pointed out the illogicality of such an arbitrary rule (for example, should an epidemiologist adjust their results based on the number of statistical tests performed that day or for the number of tests they have ever done? (Rothman, 1990)) and prefer to take a more common-sense approach. One notable exception is in the context of modern genetic studies which may evaluate tens or hundreds of thousands of genetic markers at the same time. In this situation, increased stringency is essential to minimise the thousands of spurious results that will arise simply by chance if we accept a significance level of 5% (5% of 100,000 genes is ~5000 significant results by chance!). Results from the new 'genome-wide association studies' (GWAS) which may look at 1 million or more genetic variants in relation to disease are usually not considered statistically significant unless p is less than about 0.0000001.

Summary: one swallow doesn't make a summer

We will end with a note of caution. The ultimate aim of much public health research is to change practice or policy to improve health outcomes, but even if a well-written paper that is (largely) free from major sources of bias and confounding finds what appears to be a statistically and practically significant association between an exposure and health outcome, we cannot rush out to act on this. Despite our best efforts and those of the investigators it is still possible that statistically significant results can arise by chance. As you saw in Chapter 6 the probability of this happening is usually defined as <5%; however, in many modern studies the investigators study multiple associations so the probability that one will arise by chance is greatly increased. Some authors recommend correcting results for this problem which is known as 'multiple testing' (see Box 9.2); however, we and many others prefer to rely more on a common-sense approach that places less emphasis on the question of statistical significance and more on

the overall strength, coherence and plausibility of an observed association. We will discuss some of these issues further in Chapter 10. With the possible exception of a large randomised trial, no practical or policy decision should be made on the basis of the results of a single study, however good. As you have seen, individual studies can never be perfect so it is important to consider all of the evidence on a given subject before attempting to make policy or practical decisions. We will come back to the ways in which you can do this in Chapter 11.

Questions

We have not included any questions for this chapter, but the Epidemic Intelligence Service of the US Centers for Disease Control and Prevention has developed an excellent exercise, 'Cigarette smoking and lung cancer', that draws on many of the issues covered in this and the previous chapters. This and other similar exercises are freely available from http://www.cdc.gov/eis/casestudies/casestudies.htm.

REFERENCES

Des Jarlais, D. C., Lyles, C., Crepaz, N. and the TREND Group. (2004). Improving the reporting quality of non-randomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health*, **94**: 361–366.

Harley, D., Ritchie, S., Bain, C. and Sleigh, A. C. (2005). Risks for Ross River virus disease in tropical Australia. *International Journal of Epidemiology*, **34**: 548–555.

Little, J., Higgins, J. P. Y., Ioannidis, J. P. A. *et al.* (2009). Strengthening the Reporting of Genetic Association Studies (STREGA) – an extension of the STROBE statement. *PLoS Medicine*, **6**(2): e1000022.

Moher, D., Schultz, K. F. and Altman, D. G. for the CONSORT Group. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, **357**: 1191–1194.

Osborn, D. P. J., Fletcher, A. E., Smeeth, L. *et al.* (2003). Factors associated with depression in a representative sample of 14,217 people aged 75 and over in the United Kingdom: results from the MRC trial of assessment and management of older people in the community. *International Journal of Geriatric Psychiatry*, **18**: 623–630.

Pandeya, N., Williams, G. M., Green, A. C. *et al.* (2009). Do low control response rates always affect the findings? Assessments of smoking and obesity in two Australian case-control studies of cancer. *Australian and New Zealand Journal of Public Health*, **33**: 312–319.

Riggs, J. E., Moudgil, S. S. and Hobbs, G. R. (2001). Creutzfeld–Jakob disease and blood transfusions: a meta-analysis of case–control studies. *Military Medicine*, **166**: 1057–1058.

Rothman, K. J. (1990). No adjustments are needed for multiple testing. *Epidemiology*, **1**: 43–46.

Santillan, A. A. and Camargo Jr, C. A. (2003). Body mass index and asthma among Mexican adults: the effect of using self-reported versus measured weight and height. *International Journal of Obesity*, **27**: 1430–1433.

Tooth, L., Ware, R., Dobson, A., Purdie, D. and Bain, C. (2005). Quality of reporting of observational longitudinal research. *American Journal of Epidemiology*, **161**: 280–288.

von Elm, E., Altman, D. G., Egger, M. *et al.* (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PLoS Medicine*, **4**(10): e296.

Whiteman, D. C., Whiteman, C. A. and Green, A. C. (2001). Childhood sun exposure as a risk factor for melanoma: a systematic review of epidemiologic studies. *Cancer Causes and Control*, **12**: 69–82.

Young-Xu, Y., Jabbour, S., Goldberg, R. *et al.* (2003). Usefulness of statin drugs in protecting against atrial fibrillation in patients with coronary artery disease. *American Journal of Cardiology*, **92**: 1379–1383.