QUESTIONS WITH MODELS ANSWERS–For students

Session 6: Tutorial 1 - Tools of the trade: understanding and interpreting the findings commonly reported in papers

Dr Alex Bottle, Dr Mireille Toledano

Instructions:

This tutorial is designed to help you to understand the commonly reported findings you see in papers published in medical journals.

Based on feedback from previous years' students, we have changed the way we teach medical and epidemiological statistics. The focus is now on the interpretation of the statistics rather than their calculation and teaching will be done via a tutorial session rather than in a lecture theatre.

The introductory text describes two worked examples. These examples have been provided to teach you core concepts and to help put into context what you have already learnt; there is also a glossary at the end of the tutorial to define the key terms you will need to know, these terms are italicised in the text. You should read these worked examples before the tutorial, so that you have sufficient time to work through the questions provided during the timetabled session with your tutor.

Learning outcomes:

- Be able to understand the concept of sampling and sampling variation
- Be able to understand that from a *sample*, estimates of the true underlying *risk* in a population can be calculated.
- Be able to define and interpret a *P value* and a *confidence interval*
- Be able to explain the role of statistical hypothesis testing and *confidence intervals* when dealing with chance
- To know the difference between probability and odds and be able to interpret and explain measures of association (*relative risk*, *attributable risk*, *odds ratio*) from simple examples
- Define *confounding* and understand the problems associated with it. Be able to list some methods for dealing with *confounding* (including *stratification*, *standardisation* and *regression*).

Suggested further reading:

Martin Bland (2000) ***An introduction to medical statistics***. Oxford University Press.

Worked example 1 (sampling, P values and confidence intervals)

What is the role of statistics in medicine? Discuss!

1.1 Sampling – estimating prevalence of disease or risk factors

A Primary Care Trust (PCT) wants to estimate the *prevalence* of smoking among their 100,000 residents. What does prevalence mean? How would they do this?
Suppose they surveyed a random *sample* of people – why take a random sample?
Suppose they asked 100 people if they smoked and found that 28 did. If they then asked another 100, would they also find that 28 of them smoked? Why might they not?

If they kept sampling sets of 100 people and plotted the percentage of smokers (prevalence of smoking) in each sample, we would expect to see a *normal distribution* (see glossary), with most sample estimates centred around the true population percentage.

1.2 Confidence intervals and P values – assessing the role of chance

The PCT's estimate of their population's smoking prevalence is 28% from their sample, but there will be some uncertainty around this estimate. We express this uncertainty using a 95% *confidence interval* (95% CI) around the estimate, e.g. 19% to 37%. This means that if we repeated the sampling 100 times, we would expect the true prevalence of smoking in the PCT to fall within the CI in 95 of the 100 samples.

Suppose the PCT wanted to lower this prevalence; they could implement a smoking reduction campaign and then see if it worked by comparing their first estimated prevalence with an estimate after the campaign. They took two random samples, the first finding that 28% smoked as above, and the second finding that 21% smoked. Can we therefore say for certain that the campaign has worked and cut the prevalence by 28-21=7%?
Why not?
We want to know whether the difference of 7% could simply be due to chance (sampling error) or is a real difference in prevalence. This is done statistically by setting up a *null hypothesis* of no difference and looking for evidence to disprove it: what is the likelihood that our two samples were 28% and 21% if the two true underlying prevalences were the same? We then choose the appropriate statistical test (e.g. chi-squared test to compare the two proportions) to get this likelihood, which is the *P value*. The lower the P value, the less likely that our estimated difference is a chance finding. Suppose the P value was 0.014. Convention has it that if $P<0.05$ (and this is an arbitrary cut-off!) then we can reject the null hypothesis and conclude that the smoking prevalence fell after the campaign. Such a result is called statistically significant.

THE PROCEDURE:

1. Set up a null hypothesis (e.g. difference in prevalence between the two groups is zero)
2. Choose an appropriate statistical test
3. Inspect the results (estimated measure of association – or, in this case, estimated difference in prevalences – plus its CI and P value) for evidence of real difference: can we reject the null hypothesis?

Are statistically significant results more or less likely with small sample size than with large sample sizes? Why (the answer is to do with the nature of the sample rather than statistics)?

Worked example 2 (measures of association)

The main aim of epidemiological research is to investigate the association between exposure to a risk factor (e.g. smoking) and the occurrence of disease (e.g. lung cancer). We compare the incidence in a group of people exposed to the risk factor with a group who were not exposed. Suppose the incidence in one group is higher than in the other – what are the two different ways of stating this? If Joe is 36 and John is 18, how could we say by how much Joe is the elder?

2.1 Ratio measures: relative risk and odds ratio

Two key concepts: risk and odds. What is the difference?

Suppose you wanted to look at possible risk factors for lung cancer: smoking and occupational exposure. How might you select your population sample to do this?

COHORT STUDY: Malarcher et al (2000) took a large group of US males, some smoke and some never have done, and followed them up over time. They measured the *rates* of lung cancer in the two groups. They set up the null hypothesis of equal rates and calculated the *relative risk*: 27 for smokers compared with those who have never smoked (95% CI 19 to 38). Can we reject the null hypothesis?

The interpretation of a relative risk is straightforward: if you smoke, you are 27 times more likely to die from lung cancer than if you don't smoke.

CASE-CONTROL STUDY: Richiardi 2005 took a group of people with lung cancer (the *cases*) and another without lung cancer (the *controls*) and asked each about their occupation (whether they were dockers or freight handlers). Their occupation is the exposure here. Richiardi measured the odds of exposure (odds of working as a docker or freight handler) in the cases and then in the controls. They set up the null hypothesis of equal odds and calculated the *odds ratio*: 1.5 for those with lung cancer compared with those without (95% CI 1.1 to 2.1). Can we reject the null hypothesis?

This odds ratio means that someone with lung cancer is 1.5 times more likely to have worked as a docker or freight handler than someone who doesn't have lung cancer. Notice that it compares the exposure in the two groups – it does not compare the disease rates in the two groups, which the relative risk does. The odds ratio is an estimate of the relative risk, and it is usually more useful (and easier!) to interpret an odds ratio to mean that if you work as a docker or freight handler you are 1.5 times more likely to get lung cancer than if you work in a different occupation. See the glossary for an explanation of the relationship between relative risk and odds ratio and on why case-control studies can only provide us with the latter.

2.2 Difference measure: attributable risk (or attributable fraction)

The attributable risk for lung cancer in smokers is the rate of lung cancer amongst smokers minus the rate of lung cancer amongst non-smokers (i.e. the risk difference). It gives an indication of how many extra cases for which the exposure is responsible, making the important assumption that the relation between the exposure and the disease is causal (i.e. not explained by other confounding factors – see below). The attributable risk and related measures are typically used to help guide policymakers in planning public health interventions.

2.3 Confounding – and controlling for it

How can we prove that an exposure causes a disease, rather than is merely associated with higher rates of that disease? We try to eliminate (i.e. control or adjust for) the effects of confounders. Confounders are associated with both the exposure of interest and the outcome of interest (e.g. developing a disease or dying).

Confounding can be dealt with at the design stage of a study by *randomisation* (in a randomised controlled trial), *restriction*, or *matching* (in a case-control study). Alternatively, confounding variables can be controlled for at the analysis stage, by *stratification* (splitting the analysis e.g. by age group), *standardisation*, or *regression* (building a statistical model).

In Richiardi's case-control study, regression was used to control for the effect of smoking on lung cancer risk. The lung cancer risk associated with working as a docker or freight handler after controlling for the effect of smoking was reduced to 1.3 (95% confidence interval 0.9 to 1.9). Although the odds ratio is still higher (by 30%) for dockers or freight handlers, the confidence interval now spans 1 and so we can accept the null hypothesis that working as a docker or freight handler has no effect on lung cancer risk. This is because the higher odds reported for dockers or freight handlers could just have been found by chance. Smoking is therefore a confounder here, as it's associated with both the exposure (being a docker or freight handler) and the disease (lung cancer).

We can only adjust (control) for confounding factors if we have measured them. How often have you watched a TV news piece about an association between some potential risk factor and a disease and wondered, 'but could that be due to X instead?' Journalists rarely bother to talk about confounders.

Tutorial questions

The following questions will be undertaken in small groups, facilitated by a tutor. All the questions are designed to test your understanding of, and help you apply, the knowledge you will have learnt by reading the above worked examples, from listening to your tutor briefly explain the core concepts in the worked examples, and from the material covered in your lectures on the course so far. The questions should be worked though in groups; if you get stuck at any point please refer to the glossary at the end of this tutorial and ask your tutor for help.

Question 1 – Sampling distribution and confidence intervals

A study was conducted to assess whether hormone replacement therapy (HRT) conferred a protective effect on acute myocardial infarction *risk*. 1013 women with an acute myocardial infarction and 5000 women of a similar age range without acute myocardial infarction were asked whether or not they currently used HRT. 13.1% of the women who had had an MI used HRT, whereas 17.1% of the women who had not had an MI had used HRT. This study reported an *odds ratio* of 0.72 (95% *confidence interval* 0.59-0.88) for current or recent HRT use on acute myocardial infarction *risk* (Varas Lorenzo, 2000).

a) What type of study is this?
Case-control – participants were selected on the basis of whether they were cases or not, and then their exposure status was assessed.

b) Why were 1013 women with an MI recruited instead of, say, 50? Why not 50,000?
50 would give an unreliable estimation of the proportion taking HRT, whilst 50,000 would cost a fortune and might be overkill in terms of reliability of estimates.

c) Why were the 5000 "*controls*" (women without the *outcome* of interest, i.e. MI) chosen to have a similar age range as the "*cases*" (women with MI)?
Age is clearly related to a woman's chances of taking HRT. Imagine if the controls had been taken from women aged 65+ or 18-44: would this have been a fair comparison?

d) What is the *null hypothesis* that this study is trying to disprove? Always be specific – don't just say "that there is no difference".
That the odds of taking HRT in women who had had an MI are the same as the odds of taking HRT in women who had not had an MI, i.e. the odds ratio equals 1. This would mean that taking HRT does not affect your chances of getting an MI (at least in the age range of those studied here)

e) The 95% *confidence interval* for the *odds ratio* was 0.59-0.88. What does this mean?
In general, if we repeated this study at a 100 other hospitals with 100 other sets of controls, each time generating an odds ratio and a 95% confidence interval for that odds ratio, 95% of these confidence intervals would contain the true (population-level) odds ratio. Although not strictly correct, it is best to think of the 95% CI as meaning that the real odds ratio for the effect of HRT on MI is somewhere between 0.59 and 0.88.

f) For us to accept the *null hypothesis*, what would the 95% *confidence interval* look like? Give an example of its values.
It would include the value expected under the null hypothesis, i.e. 1.00, e.g. 0.35-1.09.

g) What does the *odds ratio* of 0.72 mean in words, and how would you explain this *odds ratio* to someone taking HRT?
The odds of taking HRT in women who have had an MI is 0.72 times the odds of taking HRT in women without an MI.
The risk of acute myocardial infarction is reduced by ~30% in women who currently or recently used HRT.

Question 2 – Dealing with confounding (in study design and analysis)

In a randomised controlled trial of patient self-monitoring of blood pressure in Birmingham general practices (McManus et al, 2005), 441 hypertensives were randomly allocated to either the usual monitoring by the practice (*control* group) or self-monitoring (intervention group). After six months, the intervention group reduced their systolic BP by an average of 4.3 mmHg (95% CI 0.8-7.9) more than the *control* group.

a) What is an appropriate distribution for a group of patients' BP?
The normal distribution, defined by its mean and SD.

b) What was the main *null hypothesis* for this study? Be specific, rather than just saying that "there is no difference".
That the mean systolic BP of the two groups would be the same after six months of follow up: i.e. after six months, the mean BP of the intervention group minus the mean BP of the control group would be 0.

c) Do we have evidence to reject the *null hypothesis*? What does this mean?
Yes – the CI does not include 0, the value expected under the null hypothesis. The mean difference of 4.3 mmHg is unlikely to be due to chance, suggesting a genuine difference between the two groups. However, this difference might be explained by other factors (bias and confounding).

d) Why did the investigators randomly allocate patients to the two groups?
To try to ensure that possible confounders were equally distributed across the groups. This randomisation is one method for controlling for unknown confounding variables at the design stage of a study.

e) *Randomisation* was "stratified by diabetic status". What does this mean and why was it done?
All the people with diabetes were divided evenly (with each person allocated at random to a particular group) between the groups so that any treatment effect would not be due to a disproportionate number of people with diabetes, who have a higher BP on average than people without, in one group or the other. This is an example of controlling for confounding in the design stage.

f) Other than diabetes, what other confounders might we want to control for?
You would want to control for any factors known to be associated with both the exposure (monitoring blood pressure) and outcome (blood pressure), and also for any factors thought to be potentially associated with exposure and outcome. The authors also adjusted for GP practice, sex and deprivation.

g) The authors found that the intervention group had lost more weight and cut down their alcohol at the six-month follow-up stage. What is the relevance of this finding?
It gives a mechanism for the observed fall in mean BP for the intervention group, i.e. they improved their lifestyle and BMI and therefore reduced their BP. A useful point when assessing whether the study's finding is real or due to chance, bias and confounding is whether there is a plausible biological explanation for it, as here. Recall the Bradford-Hill criteria for causation.

Question 3 – Understanding measures of association (and confounding)

To estimate the *incidence* of breast cancer in the UK population, records from the NHS breast screening programme (which screens women aged between 50 and 70) were explored. These data indicated that the *incidence* of breast cancer was 289 per 100,000 population.

a) What does "*incidence*" mean?
The number of new cases of breast cancer over a defined period of time divided by the number of people in the population over the same period (usually we would restrict the population to those at risk of breast cancer).

b) What can the *incidence* in this *sample* of the population tell us about the *incidence* in the whole UK female population?
Because breast cancer risk varies with age, these data cannot tell us anything about the incidence of breast cancer in the whole female population of the UK. However, assuming that a fairly random selection of women in this age group attend the screening, these data can provide a pretty good estimate of the incidence in women aged between 50 and 70.

c) A *null hypothesis* that the *incidence* of breast cancer in the UK female population aged 50-70 (289 per 100,000) is that same as the *incidence* in the UK female population aged 30-50 (90 per 100,000) gives a *p-value* of <0.0001. How would you interpret this *p-value*?
This p-value suggests that the null hypothesis can be rejected. This observed difference in incidence of breast cancer in these two populations of different ages would only be expected to occur by chance less than once every 10,000 times, meaning we can be quite happy in accepting that the incidences by age group are different.

d) The *risk* of getting breast cancer if you are a woman aged 50-70 relative to the *risk* of getting breast cancer if you are a woman aged 30-50 is 3.20 (95% *confidence interval* 3.11-3.29). How would you interpret this *relative risk*?
Women aged 50-70 are 3.2 times more likely to get breast cancer than women aged 30-50. The confidence intervals give us an indication as where the true (population-level) relative risk is likely to be, and in this case women aged 50-70 are most likely to be between 3.11 and 3.29 times more likely to get breast cancer than women aged 30-50.

e) The *odds* of being aged 50-70 if you have breast cancer compared with the *odds* of being aged 30-50 if you have breast cancer is also 3.2 (95% *confidence interval* 3.11-3.29). When the *odds ratio* and *relative risk* are calculated differently, why is it that they are the same in this study?
The odds ratio is an estimate of the relative risk, and for rare outcomes (such as cancers), these measures of effect will give similar estimates of risk. However, this is not the case for common outcomes, or for rare outcome studied over very long periods of time.

f) The crude *relative risk* of breast cancer in women who are current users of HRT is 1.83 (95% CI 1.72-1.93), compared with the age-adjusted *relative risk* of 2.00 (1.91-2.09) (Beral, 2003). Which of these *risk* estimates would you consider to best reflect the *risk* of breast cancer associated with HRT use?
The incidence of breast cancer is very dependant on the age of the women (as shown above); additionally HRT use is likely to be related to age (pre versus post menopausal). Here age is likely to act as a confounder in the relationship between HRT use and breast cancer, so the age adjusted risk is likely to be most informative.
Because many disease incidences vary with age, and because most populations do not have identical age structures it is common to see 'age standardised' relative risks reported. This means that if the cases and controls, exposed and non exposed, or study and comparison populations have different age structures, a useful risk measure can still be derived.

Question 4 – Relative risk vs attributable risk

An occupational study was carried out to investigate the effect of *exposure* to aromatic amines on bladder cancer *risk*. 6667 workers with potential *exposure* to aromatic amines were followed over 30 years to see what effect this *exposure* had on bladder cancer *risk*.

a) What type of study is this?
A cohort study. A group of people were selected for the study on the basis of their exposure status, and were followed over time to see who developed bladder cancer.

b) One quarter of the study population were exposed to aromatic amines, and the *risk* associated with this *exposure* on bladder cancer was found to be 296.94 (95% CI 41.45-2127.34). What does this *risk* measure tells us?
That exposure to aromatic amines is strongly (and statistically significantly) associated with bladder cancer risk.

c) How would you explain this *risk* to someone with occupational *exposure* to aromatic amines?
If you are exposed to aromatic amines at work your risk of getting bladder cancer is 297 times higher than if you were not exposed to aromatic amines.

d) The *population excess fraction* (excess fraction of bladder cancer due to aromatic amine *exposure* in the whole study population) is 98.7 percent. How would you interpret this figure?
Assuming causality, 98.7 percent of the bladder cancer cases in the study population can be attributed to occupational exposure to aromatic amines.

e) One quarter of this study population are cigarette smokers. Cigarettes contain low doses of aromatic amines and have also been found to be associated with an excess *risk* of bladder cancer, with a *relative risk* of 5.11 (95% CI 3.42-7.64), and a *population excess fraction* of 50.7%. To reduce bladder cancer *incidence* in this cohort, would it be better to reduce work place *exposure* to aromatic amines, or to encourage the workers to stop smoking?
Reducing work place exposures to aromatic amines would lead to the greatest reduction in bladder cancer risk. Whilst the prevalence of occupational and smoking related exposure in this cohort is the same (25%), the risk associated with occupational exposure is much greater that that associated with smoking.

f) Assume the same *risks* associated with occupational aromatic amine *exposure* (*relative risk* of ~297) and smoking (*relative risk* of ~5) in the occupational cohort apply to the whole population of England. In this England 'cohort', only 0.001% of the population has occupational exposure to aromatic amines, whilst 25% smoke. The *population excess fraction* is now 22.8% for aromatic amines, but remains at ~50% for smoking. Which *exposure* should be minimised to reduce *incidence* of bladder cancer in this population?
Although the relative risk associated with smoking is much lower than the relative risk associated with exposure to aromatic amines, the prevalence of exposure to aromatic amines is very low in this population. Assuming these exposures are causally related to bladder cancer, then of the ~10,000 new cases of bladder cancer in England each year, half of them (~5,000) could be attributed to smoking, and 22.8% of them (~2,280) could be attributed to occupational exposure to aromatic amines. It would therefore be more effective to reduce the number of smokers in the population than to minimise aromatic amine exposure (although obviously it would be best to reduce both exposures!).

g) Assuming the *relative risk* of smoking on coronary heart disease mortality is ~2 (*population excess fraction* ~20%), and again taking the *relative risk* of smoking on bladder cancer in the population of England to be 5 (*population excess fraction* ~50%), and how is it that more deaths from coronary heart disease are attributed to smoking than bladder cancer *cases*?
Coronary heart disease mortality is a much more common in the population than bladder cancer incidence, so although 50% of bladder cancer cases can be attributed to smoking (~5000 cases

per year in England), 20% of coronary heart disease deaths in England amounts to the much larger figure of ~20,000 per year.
The population excess fraction is not just influenced by the relative risk associated with exposure, it is also dependent on the prevalence of exposure in the population being studied, as well as on the underlying incidence of the disease in the population.

h) What is the most useful measure of *risk* – the relative or the absolute (excess fraction) *risk*?
It depends on what you want to know. If you want to identify the risk factors for a disease, the relative measure tells you what you need to know (i.e. how many times more likely are the exposed compared with the non exposed people to develop the outcome of interest). If you can establish that an exposure is causally associated with a disease, you can then think about the impact of exposure on the incidence of the disease in the population (attributable or excess risk).

There are several limitations to reporting population attributable risks and excess fractions. You need to know that the exposure is causally related to the outcome of interest; you need to know that there is no bias or confounding that might influence the risk measure; and if you want to extrapolate the findings from a specific cohort, you also need to be sure that the study population is generalisable to the wider population. It is difficult to satisfy all these criteria, and as a result attributable risks should be looked upon as a best guess of the impact of exposure. You should also be aware that by calculating population excess fractions individually (e.g. for smoking, occupational exposure, diet), and ignoring the fact that many risk factors interact with each other, the percentages can add up to more than 100%. In the occupational cohort above, 98.7% of the bladder cancer cases were attributed to aromatic amine exposure; however 50.7% of the same cases were attributed to smoking!

References

Beral, V. & Million Women, S. C. 2003, "Breast cancer and hormone-replacement therapy in the Million Women Study.[see comment][erratum appears in Lancet. 2003 Oct 4;362(9390):1160]", *Lancet*, vol. 362, no. 9382, pp. 419-427.

Malarcher, A. M., Schulman, J., Epstein, L. A., Thun, M. J., Mowery, P., Pierce, B., Escobedo, L., & Giovino, G. A. 2000, "Methodological issues in estimating smoking-attributable mortality in the United States", *American Journal of Epidemiology*, vol. 152, no. 6, pp. 573-584.

McManus RJ, Mant J, Roalfe A, Oakes RA, Bryan S, Pattison HM, Hobbs FDR. Targets and self monitoring in hypertension: randomised controlled trial and cost effectiveness analysis. Br Med J 2005 (Sep); 331: 493-498

Richiardi, L., Forastiere, F., Boffetta, P., Simonato, L., & Merletti, F. 2005, "Effect of different approaches to treatment of smoking as a potential confounder in a case-control study on occupational exposures", *Occupational & Environmental Medicine*, vol. 62, no. 2, pp. 101-104.

Varas-Lorenzo, C., Garcia-Rodriguez, L. A., Perez-Gutthann, S., & Duque-Oliart, A. 2000, "Hormone replacement therapy and incidence of acute myocardial infarction. A population-based nested case-control study", *Circulation*, vol. 101, no. 22, pp. 2572-2578.

Glossary

Words *in italics* are defined elsewhere in the glossary

Attributable risk – the *attributable risk* is a measure of *exposure* effect that indicates, on an absolute scale, how much greater the frequency of disease in the exposed group is compared with the unexposed, assuming the relationship between *exposure* and disease is causal (an important assumption). It is the difference between the *incidence rate* in the exposed and non exposed groups, i.e. it represents the *risk* attributable to the *exposure* of interest.

Attributable risk = Incidence in the exposed - Incidence in the unexposed

For example, if 20 out of 100 smokers got lung cancer (in a given period of time) compared with 5 out of 100 non-smokers, the *relative risk* (see below) would be 20/5 = 4, but the *attributable risk* would be (20 - 5)/100 = 15 per 100. This may also be expressed as an excess fraction; 15 per 100/20 per 100 = 75%. Of the 20 *cases* of lung cancer in the smoking population, 15 of them (75%) could be attributed to smoking. The *attributable risk* is especially useful in evaluating the impact of introduction or removal of risk factors. Its value indicates the number of *cases* of the disease among the exposed group that could be prevented if the *exposure* were completely eliminated.

Case – an individual with the *outcome* under study (in a *case-control study*). Epidemiological research is based on the ability to quantify the occurrence of disease in populations. This requires a clear definition of what is meant by a case. This could be a person who has the disease, health disorder, or suffers the event of interest (by "event" we mean a change in health status, e.g. death in studies of mortality or becoming pregnant in fertility studies). The epidemiological definition of a case is not necessarily the same as the clinical definition.

Case-control study – study in which individuals are selected on the basis of whether or not they have the *outcome* of interest; usually some relatively rare *outcome*. *Exposure* (risk factor) status is explored to establish whether the *exposure* is more common in the *case* (those that have the *outcome*) or *control* (those that do not have the *outcome*) group. This type of study always results in an *odds ratio*, for example comparing the *odds* of being exposed (e.g. a smoker) in those who had the *outcome* (e.g. pancreatic cancer), with the *odds* of being a smoker in those who did not have pancreatic cancer.

Cause – the key question in most medical research. Did *exposure* to electromagnetic radiation *cause* the leukaemia in children living near mobile phone masts? Did HRT *cause* the higher DVT *rates* in women taking it? Research works by trying to disprove alternative explanations (e.g. chance, *confounding*). If this can be done, then the relationship between the *exposure* and the *outcome* will be one of causation.

Count - the most basic measure of disease frequency is a simple count of affected individuals. The number (count) of *cases* that occurred in a particular population is of little use in comparing populations and groups. For instance, knowing that there were 100 *cases* of lung cancer in city A and 50 in city B does not tell us that lung cancer is more frequent in city A than B. There may simply be more people in city A. The number of *cases* may, however, be useful in planning services. For instance, if you wanted to set up an incontinence clinic, you would want to know the number of people with incontinence in your population.

Chi squared test – a statistical procedure for testing whether two proportions are similar (e.g. whether the proportion of lung cancer *cases* in males who smoke is significantly different to the proportion of lung cancer *cases* in males who do not smoke).

Cohort study – study in which individuals are selected on the basis of *exposure* status and are followed over a period of time to allow the frequency of occurrence of the *outcome* of interest in the exposed and non exposed groups to be compared. Take a group of people, note whether they've been exposed or not, observed them over time and wait for them to get ill, to die etc. This type of study typically produces a *relative risk.*

(95%) Confidence interval – an estimated range of values calculated from a given set of *sample* data which are likely to contain the 'true' population value. E.g. a range of values around a *relative risk* measure which would, in 95% of such studies, contain the 'true' *risk* (the true *risk* being the *relative risk* that would be obtained if the study had included the entire population of patients). By "contain (or 'span') the true value", we mean that the true value lies above the lower value of the *confidence interval* but below the upper values of the *confidence interval*. For example, for a *95% confidence interval* of 1.2 – 3.4, we can say that we are 95% confident that the true value of risk will not be lower than 1.2 and will not be higher than 3.4.
If we find that our confidence interval for the *relative risk* or *odds ratio* for group A compared with group B does not include 1, then we typically reject the null hypothesis of no difference. However, if our study is not on *rates* of disease or on proportions of patients *exposed* but is on a measure such as blood pressure or weight, we would typically reject the *null hypothesis* if the confidence interval for the average difference in blood pressure or weight between group A and group B does not include 0, not 1. Why is this? See entry for *null hypothesis*.

Confounding – a possible explanation for the study finding if *confounding variables* have not been taken into account in the study.

Confounding variable – a factor that is associated with both the *exposure* and *outcome* of interest. Common confounders include age, smoking, socio-economic deprivation. Smoking is a confounder because smoking tends to be more prevalent in people exposed to non-tobacco-related toxins and carcinogens, and also more prevalent in people with a range of diseases.

Control (as opposed to a *case*) – a person without the *outcome* under study (in a *case-control study*), or a person not receiving the intervention (in a clinical trial). The choice of an appropriate group of *controls* requires care, as we need to be able to draw useful comparisons between these *controls* and the *cases*/intervention group.

Exposure – when people have been 'exposed', they have been in contact with something that is hypothesised to have an effect on health e.g. tobacco, nuclear radiation, pesticides in food, HRT. Contact may be via any route: oral, inhalation, through the skin etc. These are typically called 'risk factors' of disease. We are interested in whether the *exposure* results in higher (or sometimes lower) *outcome rates*.

Incidence – the number of new *cases* of the *outcome* of interest occurring in a defined population over a define period of time. Note that this is not the same as *prevalence*, which includes new and old *cases*. *Incidence* measures events (a change from a healthy state to a diseased state).

$$\text{Incidence} = \frac{\text{Number of new cases of disease in a given time period}}{\text{Number of disease-free persons at the beginning of that time period}}$$

This measure of *incidence* can be interpreted as the probability, or risk, that an individual will develop the disease during a specific time period.

Matching - a method for "controlling for" (i.e. effectively removing) the effect of *confounding* at the design stage of a *case-control study*; *controls* are selected to have a similar distribution of

potentially *confounding variables* to the *cases,* e.g. they are said to be "matched" for sex if there are similar proportions of men and women in both groups.

Normal distribution – a set of values and frequencies that describe many things in nature, at least approximately, e.g. height, weight, blood pressure. This symmetrical distribution (see Figure 1) is the basis of many statistical tests because, if you know the average value (usually called the mean) and the standard deviation, then you can draw every point of a *normal distribution* and you know what proportion of values are greater than (or less than) any given point, e.g. the % of men more than two metres tall. Some things are not normally distributed (e.g. proportions of anything, serum concentrations of electrolytes) but can be made to fit quite well after some simple mathematical trickery.

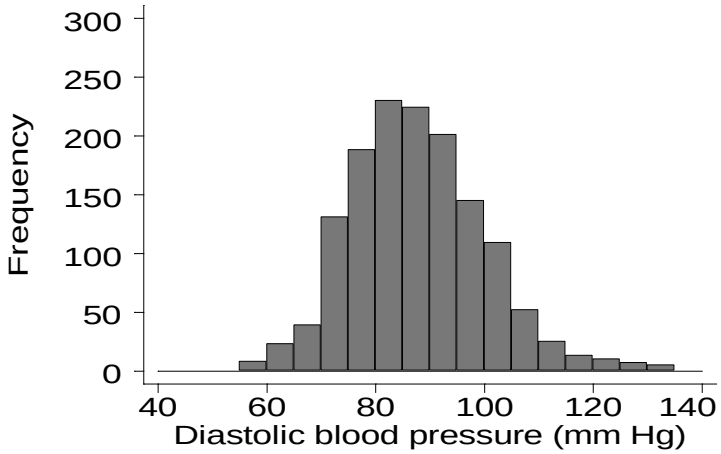

Figure 1. A *normal distribution* - from a study of diastolic blood pressure among men (*British Medical Journal*·1974; **3**: 600-3).

Null hypothesis – formulating a *null hypothesis* is the first stage in performing any statistical test. Typically, when two groups (A and B) are being compared, the *null hypothesis* that the statistical test tries to disprove is that there is no difference between the two groups in the measure being tested. If we are comparing rates, then the null hypothesis would be that *rate* A equals *rate* B, which means that the *relative risk* (*rate* A divided by *rate* B) equals 1. For *case-control studies*, the null hypothesis would be that the odds of *exposure* for group A equal the odds of *exposure* for group B, i.e. the *odds ratio* (*odds* of exposure for A divided by the *odds* of exposure for B) equals 1. A *statistical test* is then performed on the *relative risk* or the *odds ratio* and a *confidence interval* for it is derived. We can reject the null hypothesis if the *confidence interval* does not include the value expected under the null. In this case, the null has RR=1 or OR=1, so we would reject it if the *confidence interval* does not include 1.
However, for *normally distributed* variables such as blood pressure (BP) in Question 4, the null hypothesis would be that the average BP for group A equals the average BP for group B, i.e. the difference between the two average BPs equals 0. The *statistical test* would then be performed on this difference in average BPs and the resulting *confidence interval* would also relate to the difference in average BPs. We therefore would reject the null hypothesis if the *confidence interval* did not include 0, which is the value expected under the null.
If, when faced with a *confidence interval* around some measure and wondering whether to reject the null hypothesis or not, you can't remember whether it should include 1 or 0, always think in terms of what value the null hypothesis expects your measure to have and then see if that value falls within the range of values covered by the *confidence interval*.

Odds – the *odds* is another way to express probability, e.g. the *odds* of *exposure* is the number of people who have been exposed divided by the number of people who have not been exposed. The mathematical relationship between odds and probability is:
Odds = probability / (1 – probability)

Odds ratio – the *relative risk* can be calculated from *cohort studies*, since the *incidence* of disease in the exposed and non-exposed is known. In *case-control studies*, however, the subjects are selected on the basis of their disease status (sample of subjects with a particular disease (*cases*) and sample of subjects without that disease (*controls*)), not on the basis of exposure. Therefore, it is not possible to calculate the incidence of disease in the exposed and non-exposed individuals. It is, however, possible to calculate the *odds of exposure*. The *odds ratio* (of exposure) is the ratio between two odds, e.g. the *odds* of *exposure* in the *case* s divided by the *odds* of *exposure* in the *control*s.

$$\text{Odds ratio} = \frac{\text{Odds of exposure in the diseased group (cases)}}{\text{Odds of exposure in the disease-free group (controls)}}$$

This ratio is the measure reported in *case-control studies* instead of the *relative risk*. It can be mathematically shown that the *odds ratio* of exposure is generally a good estimate of the *relative risk*. An *odds ratio* of 1 tells us that *exposure* is no more likely in the *cases* than *controls* (which implies that *exposure* has no effect on *case*/*control* status); an *odds ratio* greater than 1 tells us that *exposure* is more likely in the *case* group (which implies that *exposure* might increase the *risk* of the disease). An *odds ratio* less than 1 tells us that *exposure* is less likely in the *case* group (which implies that *exposure* might have a protective effect).

Outcome – the event or main quantity of interest in a particular study, e.g. death, contracting a disease, blood pressure.

Population attributable risk (also known as the population excess risk) – a measure of the *risk* of *outcome* in the study population which is attributable to the *exposure* of interest.

Population excess fraction (also known as the population attributable fraction) – a measure of the proportion (fraction) of the *cases* observed in the study population attributable to the *exposure* of interest.

Prevalence – the number of *cases* of an *outcome* of interest in a defined population at a particular point of time, hence it is often called *point prevalence.* This includes both new (also called "incident") *cases* and existing *cases*.

$$\text{Point prevalence} = \frac{\text{Number of cases in a defined population at one point in time}}{\text{Number of persons in a defined population at the same point in time}}$$

p-value – the probability of obtaining the study result (*relative risk*, *odds ratio* etc) if the *null hypothesis* is true. The smaller the p-value, the easier it is for us to reject the *null hypothesis* and accept that the result was not just due to chance. A *p-value* of <0.05 means that there is only a very small chance of obtaining the study result if the *null hypothesis* is true, and so we would usually reject the null. Such as result is commonly called "statistically significant". A *p-value* of >0.05 is usually seen as providing insufficient evidence against the *null hypothesis*, so we accept the null.

Randomisation – a method for ensuring that both groups in a clinical trial (i.e. those receiving the intervention and those not receiving the intervention (*controls*)), have similar proportions of *confounding variables*, such as age.

Rate and risk – these words are often taken to mean the same thing (though to some epidemiological purists they are not always the same). We talk of someone's *risk*/chance/probability of getting a disease (or getting pregnant or dying etc.) and a population

having a disease *rate*. Both terms imply a proportion, i.e. the number of people with the *outcome* of interest divided by the total number of people at *risk* of the *outcome*.

Regression - a method for controlling the effect of *confounding* at the analysis stage of a study - statistical modelling is used to control for one or many *confounding variables*.

Relative risk – the *relative risk* is used as a measure of association between an exposure and disease. It is the ratio of the *incidence rate* in the *exposed* group and the *incidence rate* in the non-exposed group.

$$\text{Relative risk} = \frac{\text{Incidence in the exposed group}}{\text{Incidence in the unexposed group}}$$

For example, the proportion of people with high cholesterol who developed ischaemic heart disease divided by the proportion of people with normal cholesterol who developed ischaemic heart disease. A value of 1.0 indicates that the incidence of disease in the exposed and the unexposed are identical and thus the data shows no association between the exposure and the disease. A value greater than 1.0 indicates a positive association or an increased risk among those exposed to a factor. Similarly, a relative risk less than 1.0 means there is an inverse association or a decreased risk among those exposed, i.e. the exposure is protective.

Restriction – a method for controlling the effect of *confounding* at the design stage of a study, e.g. by including patients in a clinical trial only between the ages of 18 and 65 without pre-existing illness so that the results of the trial are not confused ('confounded') by different levels of age or morbidity in the two treatment groups.

Sample – a relatively small number of observations (or patients) from which we try to describe the whole population from which the *sample* has been taken. Typically, we calculate the mean for the *sample* and use the *confidence interval* to describe the range within which we think the population mean lies. This is one of the absolutely key concepts behind all medical research (and much non-medical research too).

Standardisation - a method for controlling the effect of *confounding* at the analysis stage of a study. Used to produce a Standardised Mortality Ratio, a commonly used measure in epidemiology.

Statistical test – the only way to decide whether the results of your analysis, e.g. your measure for group A compared with your measure for group B, are likely to be due to chance or could be real. The procedure for doing a statistical test is to take one value representing the observed difference in your study between groups A and B and compare that value against tables of an appropriate mathematical distribution such as the *normal distribution* to see how extreme it is (we use computers instead of printed tables, thankfully, these days). For example, to see if someone is unusually tall, we would need to compare their height with a normal distribution with the mean and standard distribution taken from members of the population of the same age and sex. This would be done by subtracting the population mean from the person's height and dividing by the population standard deviation and looking up the result (called the "test statistic") in a table of the standard normal distribution (so-called because it has a mean of 0 and standard deviation of 1) to find out what proportion of values are greater than this. This proportion is therefore the proportion of the population who are taller than the person. Something similar is routinely done on infants to monitor their growth.

Stratification - a method for controlling the effect of *confounding* at the analysis stage of a study - *risks* are calculated separately for each category of *confounding variable*, e.g. each age group and each sex separately.