Medical Statistics: Definitions

RATE

An expression of the frequency with which an event occurs in a defined population. Rates are used so that comparisons can be made between different populations at different times, in different places and among different groups of people. Usually expressed per 100, 1000 or 10,000 population.

INCIDENCE RATE

The number of newly discovered cases per unit of population during a stated period of time; usually used in conditions which are acute and of short duration.

$$\frac{\text{Number of new cases in a defined period of time}}{\text{Population at risk}} \times 1000$$

PREVALENCE RATE

The number of cases per unit population at some specific moment. More useful for disorders with an insidious onset and a long average duration.

$$\frac{\text{Number of persons sick at any given instant} \times 1000}{\text{Number of persons at risk}}$$

Example: Infant Mortality Rate = $\frac{\text{No of deaths 1 year} \times 1000}{\text{Total live births}}$

STANDARDISED MORTALITY RATIO

Allows comparison of the mortality rates in different populations, by taking into account different age & sex structures of the population

$$\text{SMR} = \frac{\text{Observed deaths} \times 100}{\text{Expected deaths}}$$

Eg. ***Dr Foster Hospital SMRs*** are used to compare death rates in hospitals in the UK. The expected number of deaths is taken from the number of deaths in a larger reference population. For example, if the analysis is looking at death rates in wards, the reference population could be the North West or England and Wales. The SMR of the reference population is always 100, a value of lower than 100 means that fewer deaths than expected occurred in the local population after adjusting for differences in age and sex; more than 100 means that there have been more deaths than expected.

drfosterhealth.co.uk

Example: Hip replacements – 1 year SMRs at different hospitals in the region

Hospital A	112.12 (higher number of deaths than expected)
Hospital B	82.59 (much lower number of deaths)
Hospital C	99.68 (equal to expected number of deaths)
Hospital D	117.79 (much higher number of deaths)
Hospital E	97.8 (close to expected)

TYPES OF RESEARCH STUDY

Case Series – a study looking at a series of identifies cases of a particular condition. By definition may look at rare conditions, does not give insight or useful statistical data about the condition in relation to other conditions

Cross sectional or prevalence study – identifies the level of a particular condition in a population at a point in time, can be used to identify risk groups. Usefulness of data depends on the size of the study.

Correlational study – correlates the frequency of a risk factor with incidence of disease.

- Case series, cross sectional studies and correlational studies are described as descriptive studies. They are useful for measuring disease frequency and studying particular disease populations, but can use a lot of resources (time and money) and interpretation of data can be difficult.

ANALYTICAL STUDIES

These studies are useful for testing a particular hypothesis.

Case control Study - A comparison between representative samples of people who get a disease and people who do not. Frequency of pathological factor determined in subjects and also in controls. Control group may include the whole population.

Advantages:

- Cheap
- Good for screening a wide range of factors
- Good for rare diseases with just a few cases
- Good for expensive or time consuming tests

Disadvantages:

- Bias in selection of cases or controls
- Difficult to interpret results
- If small numbers may miss big effects by chance
- Recall bias
- Attributable risk not usually obtainable – only the odds ratio (see below)

Cohort/Longitudinal Studies - Compare people exposed to a suspected factor and those not. Can be prospective or retrospective. Leads to a defined attributable risk (see below) of developing the disease following exposure to the cause. Usually reserved for testing precise hypotheses.

Advantages:

- Provides a direct estimate of risk of developing a disease – attributable risk
- Decreases the possibility of subjective bias; obtains info before outcome is known
- Information of people whose disease status has changed can be obtained
- Information of relationship of characteristics to other diseases can be obtained
- Deaths can be taken into account, so as not to underestimate the degree of association between characteristic and disease.

Disadvantages:

- Most difficult and more expensive; large populations & long observation periods
- Participation in the study may influence development of disease
- Sampling selection can result in biased estimates of relationships
- Inefficient or impossible for rare diseases
- Slow accumulation of results
- Difficult in maintaining constant techniques over long times (staff changes)

INTERVENTION STUDIES

Controlled Clinical Trials = Longitudinal and Intervention

Think of these as an Experiment. Apply treatment or preventive measure to one (randomly allocated) group and compare the effects of this measure with the effects of another treatment/no treatment/placebo to a comparable group of subjects. Alters the patient's life. Results assessed by rigorous comparison of rates of disease, death, recovery etc. Seen as the most rigorous method of hypothesis testing. Requires ethical consideration.

SYSTEMATIC REVIEWS AND META-ANALYSIS

A Systematic Review is the application of scientific strategies to limit bias (see below) so that systematic critical appraisal can be applied to all relevant studies on a particular topic. The main potential bias is in selection of studies for inclusion. This includes publication bias (journals more likely to publish positive results).

A Meta-Analysis is a systematic review that employs statistical methods to combine and summarize the results of several studies. Follow this link for a comprehensive explanation of meta-analysis: http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/Meta-An.pdf

Determining the sample size in a clinical trial

The minimum information needed to calculate sample size for a randomised controlled trial in which a specific event is being counted includes the ***power***, the ***level of significance,*** the ***underlying event rate in the population under investigation.*** The size of the treatment effect sought is also a factor. The calculated sample size should then be adjusted for other factors, including expected compliance rates and, less commonly, an unequal allocation ratio.

1. ***Power:*** The power of a study is its ability to detect a true difference in outcome between the standard or control arm and the intervention arm. This is usually chosen to be 80%. By definition, a study power set at 80% accepts a likelihood of one in five (that is, 20%) of missing such a real difference. Thus, the power for large trials is occasionally set at 90% to reduce to 10% the possibility of a so-called "false-negative" result.
2. ***Level of significance:*** The chosen level of significance sets the likelihood of detecting a treatment effect when no effect exists (leading to a so-called "false-positive" result) and defines the threshold "*P*-value" (see below). Results with a *P*-value above the threshold lead to the conclusion that an observed difference may be due to chance alone, while those with a *P*-value below the threshold lead to rejecting chance and concluding that the intervention has a real effect. The level of significance is most commonly set at 5% (that is, $P = 0.05$) or 1% ($P = 0.01$). This means the investigator is prepared to accept a 5% (or 1%) chance of erroneously reporting a significant effect.
3. ***Underlying population event rate:*** Unlike the statistical power and level of significance, which are generally chosen by convention, the underlying expected event rate (in the standard or control group) must be established by other means, usually from previous studies, including observational cohorts. These often provide the best information available, but may overestimate event rates, as they can be from a different time or place, and thus subject to changing and differing background practices. Additionally, trial participants are often "healthy volunteers", or at least people with stable conditions without other comorbidities, which may further erode the study event rate compared with observed rates in the population. Great care is required in specifying the event rate and, even then, during ongoing trials it is wise to have allowed for sample size adjustment, which may become necessary if the overall event rate proves to be unexpectedly low

P-Value

In statistical significance testing, the **p-value** is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis (see below) is true. The *lower* the p-value, the *less* probable the result is (assuming the null hypothesis is true) and, consequently, the *more* statistically significant the result is. One often rejects the null hypothesis

when the p-value is less than 0.05 or 0.01, corresponding respectively to a 5% or 1% chance of rejecting the null hypothesis when it is true (Type I error).

Null Hypothesis (H^0) - A type of hypothesis used in statistics that proposes that no statistical significance exists in a set of given observations. The null hypothesis attempts to show that no variation exists between variables, or that a single variable is no different than zero. It is presumed to be true until statistical evidence nullifies it for an alternative hypothesis. ie. The null hypothesis assumes that any kind of difference or significance you see in a set of data is due to chance.

Confidence Interval - A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.

The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter (see precision). A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.

Confidence intervals are more informative than the simple results of hypothesis tests (where we decide "reject H^0" or "don't reject H^0") since they provide a range of plausible values for the unknown parameter.

Will this research produce valid findings? - BIAS & CONFOUNDING FACTORS

This is probably the most important question to address when designing a research project. But how can you ensure that your research will be valid? Techniques such as blinding and randomisation can enhance validity, but they do not guarantee validity and they may be inappropriate or impractical for your study. So there is no substitute for making sure that you understand how validity may be compromised and design your study accordingly.

Threats to validity!

There are broadly three reasons why findings may not be valid- **1) Chance 2) Bias 3) Confounding**

Chance: The measurements we make while doing research are nearly always subject to random variation. Determining whether findings are due to chance is a key feature of statistical analysis. Check our statistics links to find out more about hypothesis testing and estimation. The best way to avoid error due to random variation is to ensure your sample size is adequate.

Bias: Whereas chance is caused by *random* variation, bias is caused by *systematic* variation. A systematic error in the way we select our patients, measure our outcomes, or analyse our data will lead to results that are inaccurate. There are numerous types of bias that may effect a study. Understanding how bias occurs is more important than remember the names of different types of bias.

Types of bias: *These can broadly be divided into three categories -*

1) Selection bias- The selection of subjects into your sample or their allocation to treatment group produces a sample that is not representative of the population, or treatment groups that are systematically different. Random selection and random allocation are the keys to avoiding this bias.

2) Measurement bias- Measurement of outcomes is inaccurate. This may be due to inaccuracy in the measurement instrument or bias in the expectations of study participants, carers or researchers. The latter may be addressed by blinding participants, carers or researchers.

3) Analysis bias- The protection against bias created by randomisation will only be maintained if all participants remain in the group to which they were allocated and complete follow up. Participant who change groups, withdraw from the study or are lost to follow up may be systematically different from those who complete the study. Analysis bias can be reduced by maximising follow up and carrying out an intention to treat analysis.

Accuracy and precision: These two terms are often used in an inaccurate or imprecise way! Random variation (chance) leads to results being imprecise. Systematic variation (bias) leads to results being inaccurate. For example, a huge observational study of 1000's of patients may produce results that are precise, but not accurate. Whereas a small, high quality randomised controlled trial may produce results that are accurate but not precise.

Confounding: This is similar to bias and is often confused. However, whereas bias involves error in the measurement of a variable, confounding involves error in the interpretation of what may be an accurate measurement. A classic example of confounding is to interpret the finding that people who carry matches are more likely to develop lung cancer as evidence of an association between carrying matches and lung cancer. Smoking is the confounding factor in this relationship- smokers are more likely to carry matches and they are also more likely to develop lung cancer.

What is a confounder?: A confounder is a factor that is prognostically linked to the outcome of interest and is unevenly distributed between the study groups. A factor is NOT a confounder if it lies on the causal pathway between the variables of interest. For example, the relationship between diet and coronary heart disease may be explained by measuring serum cholesterol level. Cholesterol is not a confounder because it may be the causal link between diet and coronary heart disease.

Known confounders -Dealing with confounding is relatively easy if, as in this case, you know what the likely confounders are. You could stratify your results- i.e. analyse smokers and non smokers separately, or you could use statistical techniques to adjust for confounding.

Unknown confounders - Dealing with unknown confounders is obviously much trickier. There is always a risk that an apparent association between a risk factor, or an intervention, and an outcome is being mediated by an unknown confounder. This is particularly true of observational studies where patients may be selected to one treatment group or another, not according to any explicit criteria, but by some unknown process, such as a care providers 'gut feeling'. The best defence against unknown confounders is randomisation. This ensures that both known and unknown confounders are randomly distributed between treatment groups.

ASSESSING AND COMMUNICATING RISK

What are absolute and relative risks?

Absolute risk of a disease is your risk of developing a disease over a time-period. We all have absolute risks of developing various diseases such as heart disease, cancer, stroke, etc. The same absolute risk can be expressed in different ways. For example, say you have a 1 in 10 risk of developing a certain disease in your life. This can also be said a 10% risk, or a 0.1 risk - depending if you use percentages or decimals.

Relative risk is used to compare the risk in two different groups of people. For example, the groups could be smokers and non-smokers. All sorts of groups are compared to others in medical research

to see if belonging to a group increases or decreases your risk of developing certain diseases. For example, research has shown that smokers have a higher risk of developing heart disease compared to (relative to) non-smokers.

A couple of examples may illustrate this better:

An example when talking about risks of disease

Say the absolute risk of developing a disease is 4 in 100 in non-smokers. Say the relative risk of the disease is increased by 50% in smokers. The 50% relates to the 4 - so the absolute increase in the risk is 50% of 4, which is 2. So, the absolute risk of smokers developing this disease is 6 in 100.

An example when talking about treatments:

Say men have a 2 in 20 risk of developing a certain disease by the time they reach the age of 60. Then, say research shows that a new treatment reduces the relative risk of getting this disease by 50%. The 50% is the relative risk reduction, and is referring to the effect on the 2. 50% of 2 is 1. So this means that the absolute risk is reduced from from 2 in 20, to 1 in 20

Attributable Risk: Attributable risk (AR) or risk difference is the difference between the incidence rates in exposed and non-exposed groups. In a cohort study, AR is calculated as the difference in cumulative incidences (risk difference) or incidence densities (rate difference). This reflects the absolute risk of the exposure or the excess risk of the outcome (e.g. disease) in the exposed group compared with the non-exposed group. AR is sometimes referred to as attributable risk in the exposed because it is used to quantify risk in the exposed group that is attributable to the exposure.

Odds Ratio: An odds ratio is calculated by dividing the odds in the treated or exposed group by the odds in the control group. Clinical trials typically look for treatments which reduce event rates, and which have odds ratios of less than one. In these cases a percentage reduction in the odds ratios is often quoted instead of the odds ratio. For example, the ISIS-4 trial reported a 7% reduction in the odds of mortality with captopril, rather than reporting an odds ratio of 0.93

Number needed to treat (NNT)

A figure which is often quoted in medical research is the number needed to treat - NNT. This is the number of people who need to take the treatment for one person to benefit from the treatment.

For example, say a drug company reported that drug x reduced the relative risk of developing a certain disease by 25%. If the absolute risk of developing the disease was 4 in 100 then this 25% reduction in relative risk would reduce the absolute risk to 3 in 100.

But this can be looked at another way. If 100 people do not take the drug, then 4 in that 100 people will get the disease. If 100 people do take the drug, then only 3 in that 100 people will get the disease. Therefore, 100 people need to take the treatment for one person to benefit and not get the disease. So, in this example, the NNT is 100.

A quick way of obtaining the NNT for a treatment is to divide 100 by the absolute percentage reduction in risk when taking the drug. So, another quick example. Say the absolute risk of developing complications from a certain disease is 4 in 20. Say a drug reduces the relative risk of getting these complications by 50%. This reduces the absolute risk from 4 in 20, to 2 in 20. In percentage terms, 4 in 20 is 20%, and, 2 in 20 is 10%. Therefore, the reduction in absolute risk in taking this drug is from 20% to 10% - a 10% reduction. The NNT would be 100 divided by 10. That is, 10 people would need treatment for one to benefit.

Numbers Needed to Harm (NNH): This is calculated in the same way as for NNT, but used to describe adverse events. For NNH, large numbers are good, because they mean that adverse events are rare. Small values for NNH are bad, because they mean adverse events are common.

ASSESSING THE RELIABILITY AND RISKS OF HAVING A TEST

Sensitivity: Proportion of people **with the target disorder who have a positive test**

Specificity: Proportion of people **without the target disorder who have a negative test**

Positive predictive value (PPV): Proportion of people **with a positive test who have the target disorder**

Negative predictive value (NPV): Proportion of people **with a negative test who do not have the target disorder.**

This table summarises the results of a test:

	Disease is Present	**Disease is Absent**	**TOTALS**
Test is Positive	a	b	a + b
Test is Negative	c	d	c + d
TOTALS	a + c	b + d	a + b + c + d

ie. In this table, **b** = the number of people who had a positive test result but they did NOT have the disease, **c** = the number of people who had a negative test result, but actually had the disease.

- Sensitivity = a / (a+c)
- Specificity = d / (b+d)
- PPV = a / (a+b)
- NPV = d / (c+d)

Notes:

Prepared by Nick Pendleton February 2011

Thank you credits for material from: Tunbridge Wells VTS, Bandolier, Dr Foster Health, Wikipedia, PatientUK, College of Emergency Medicine (CEM).