Chapter 21   Methods of analysis

1. Introduction to methods of analysis

This chapter describes simple statistical methods that are likely to be most useful for the basic analysis of intervention trials. Usually, a statistician will be closely involved in the design and analysis of a trial, and the more advanced analytical techniques that they might employ are not covered in this chapter. For more information on such techniques, the reader is referred to statistical texts such as Armitage and Berry (1987), Kirkwood and Sterne (2003), and Rothman et al. (2008). However, the methods presented in this chapter should enable the analysis of the main results of a trial. More advanced statistical techniques usually result in relatively small changes in the estimates of effect sizes through multivariate and associated analyses. Also, armed with the methods in this chapter, the reader should be in a good position to interpret and check the analyses reported in published studies.

The methods that are going to be used to analyse a trial should be considered at the time the trial is set up, so all of the appropriate data are collected and are assembled in a form suitable for the planned analyses. It is a common requirement nowadays for the *statistical analysis plan* to be fully developed, before any blinding in a trial is broken and in advance of a 'frozen' data set being prepared for analysis. Such plans are discussed in Section 3.

The choice of an appropriate method of analysis of a trial depends on the type of outcome measure which is of interest. The different types of outcome measure are discussed in Section 2, which also includes a brief review of the concepts of confidence intervals (CIs) and statistical tests. In Sections 4, 5, and 6, methods are described which are appropriate for the analyses of data in the form of proportions, rates, and means, respectively. RCTs have been recommended as the method of choice for determining the effects of an intervention, because such trials generally avoid the problem of confounding. Sometimes, however, particularly in small trials, there may be differences between the randomized groups, with respect to factors that might affect the outcome of interest, but which are unrelated to the intervention under test. If there has been a proper randomization process, any such differences should rise by chance only. If the trial is large, it is unlikely that there will be any important imbalance in this respect between the randomized groups. In small trials, such chance differences may have a larger effect, and, in such circumstances, it may be important to adjust for any potential confounding due to these chance differences. In addition, where randomization is not feasible, any attempt to draw conclusions about the effects of an intervention must make allowance for possible confounding factors, and simple methods for doing this are described in Section 7. The analysis of trials in which interventions are allocated to groups, rather than individuals, is discussed in Section 8. How the results of a trial may be used to assess the possible public health impact of an intervention is considered in Section 9.

2. Basics of statistical inference

2.1. Types of outcome measure

The appropriate method of statistical analysis depends on the type of outcome measure that is of interest. An outcome in an intervention study can usually be expressed as a proportion, rate, or mean. For example, in a trial of a modified vaccine, an outcome measure of interest may be the *proportion* of vaccinated subjects who develop a protective level of antibodies. In a trial of multi-drug therapy for tuberculosis, the incidence *rates* of relapse, following treatment, may be compared in the different study groups under consideration. In a trial of an anti-malarial intervention, it may be of interest to compare the *mean* packed cell volume (PCV) at the end of the malaria season in those in the intervention group and those in the comparison group.

2.2. Confidence intervals

An estimate of an outcome measure calculated in an intervention study is subject to *sampling error,* because it is based on only a sample of individuals and not on the whole population of interest. The term *sampling error* does not mean that the sampling procedure or method of randomization was applied incorrectly, but that, when random sampling is used to decide which individuals are in which group, there will be an element of random variation in the results. The methods of statistical inference allow the investigator to draw conclusions about the true value of the

outcome measure on the basis of the information in the sample. In general, the observed value of the outcome measure gives the *best estimate* of the true value. In addition, it is useful to have some indication of the *precision* of this estimate, and this is done by calculating a *confidence interval* for the estimate. The CI is a range of plausible values for the true value of the outcome measure, based on the observations in the trial. It is conventional to quote the 95% *confidence interval* (also called 95% *confidence limits)*. This is calculated in such a way that there is a 95% probability that the CI includes the *true* value of the outcome measure.

Suppose the true value of the outcome measure is ø and that this is estimated from the sample data as $ø^\wedge$. The 95% CIs to be presented here are generally of the form $ø^\wedge \pm 1.96 \times SE(ø^\wedge)$ where $SEø^\wedge$ denotes the *standard error* of the estimate. This is a measure of the amount of sampling error to which the estimate is susceptible. One of the factors influencing the magnitude of the standard error, and hence the width of the CI, is the sample size; the larger the sample, the narrower the CI.

The multiplying factor 1.96, used when calculating the 95% CI, is derived from tables of the *Normal* distribution. In this distribution, 95% of values are expected to fall within 1.96 standard deviations of the mean. In some circumstances, CIs, other than 95% limits, may be required, and then different values of the multiplying factor are appropriate, as indicated in Table 21.1.

When analysing means, the multiplying factor sometimes has to be increased to allow for additional errors in estimating the standard error (see Section 5).

2.3. Statistical tests

As well as calculating a CI to indicate a range of plausible values for the outcome measure of interest, it may be appropriate to test a specific *hypothesis* about the outcome measure. In the context of an intervention trial, this will often be the hypothesis that there is no true difference between the outcomes in the groups under comparison. (For this reason, the hypothesis is often referred to as the *null hypothesis*.) The objective is thus to assess whether any observed difference in outcomes between the study groups may have occurred just by chance, due to sampling error.

A statistical test is used to evaluate the plausibility of the null hypothesis. The sample data are used to calculate a quantity (called a statistic) which gives a measure of the difference between the groups, with respect to the outcome(s) of interest. The details of how the statistic is calculated vary, according to the type of outcome measure being examined, and are given in Sections 4 to 6. Once the statistic has been calculated, its value is referred to an appropriate set of statistical tables, in order to determine the p-value (probability value) or *statistical significance* of the results. The p-value measures the probability of obtaining a value for the statistic as extreme as the one actually observed if the null hypothesis were true. Thus, a very low p-value indicates that the null hypothesis is likely to be false.

For example, suppose, in a trial of a vaccine against malaria, an estimate of the efficacy is obtained of 20%, with an associated p-value of 0.03. This indicates that, if the vaccine had a true efficacy of zero, there would only be a 3% chance of obtaining an observed efficacy of 20% or greater.

The smaller the p-value, the less plausible the null hypothesis is as an explanation of the observed data. For example, on the one hand, a p-value of 0.001 implies that the null hypothesis is highly implausible, and this can be interpreted as very strong evidence of a real difference between the groups. On the other hand, a p-value of 0.20 implies that a difference of the observed magnitude could quite easily have occurred by chance, even if there were no real difference between the groups. Conventionally, p-values of 0.05 and below have been regarded as sufficiently low to be taken as reasonable evidence against the null hypothesis and have been referred to as indicating a *statistically significant difference*, but it is preferable to specify the actual size of the p-value attained, so that readers can draw their own conclusions about the strength of the evidence.

While a small p-value can be interpreted as evidence for a real difference between the groups, a larger *non-significant* p-value must not be interpreted as indicating that there is no difference. It merely indicates that there is insufficient evidence to reject the null hypothesis, so that there *may* be no true difference between the groups. It is never possible to *prove* the null hypothesis. Depending on the size of the study and the observed difference between the groups under comparison, the CI on the difference provides a range of plausible values in which the true difference might lie, which may include a zero difference.

Too much reliance should not be placed on the use of statistical tests. Usually, it is more important to *estimate* the effect of the intervention and to specify a CI around the estimate to indicate the plausible range of effect than it is to test a specific hypothesis. In any case, a null hypothesis of zero difference is often of no practical interest, as there may be strong grounds for believing the intervention has some effect, and the main objective should be to estimate that effect.

The statistical tests presented here are *two-sided* tests. This means that, when the p-value is computed, it measures the probability (if the null hypothesis is true) of observing a difference as great as that actually observed *in either direction* (i.e. positive or negative). It is usual to assume that tests are two-sided, unless otherwise stated, though not all authors adhere to this convention. A full discussion of the relative merits of one-sided and two-sided tests is given in Armitage and Berry (1987) and Kirkwood and Sterne (2003).

3. Statistical analysis plan

A common mistake in the planning of a trial is to delay consideration of the analyses until the data become available. It is essential that the main analyses that will be undertaken are planned at the design stage, as this provides several major benefits. First, it encourages a clearer understanding of the basic questions to be answered and thus assists with the formulation of clear and specific objectives. For example, in a vaccine trial, a simple comparison of the numbers of cases of the disease occurring over a 5-year period in the vaccinated and unvaccinated groups may answer the question of the magnitude of any protective effect. A comparison of the incidence rates of disease in vaccinated and unvaccinated individuals in the first, second, third, fourth, and fifth years after vaccination can be used to answer a rather different question, namely, whether the protective effect is constant over the 5-year period.

A second benefit of considering the analyses at the design stage is that it necessitates specification of what data need to be recorded. The investigator can check that arrangements have been made to measure and record all variables that will be needed in the analyses. Also, and perhaps as importantly, it may become clear that some variables will not be needed, and these can then be omitted from the study.

The process of planning the analyses may identify also the importance of subgroup analyses. In a vaccine trial, for example, it may reveal a need to assess the efficacy of the vaccine in children vaccinated at different ages. This may have major implications for the choice of sample size, as the need for age-specific estimates of efficacy requires a much larger sample in each age group than would be needed if only an overall estimate of efficacy was wanted.

Finally, advanced planning of the analyses is desirable to ensure that adequate arrangements have been made for data handling, the necessary computer software is available, and sufficient time for data cleaning and analysis has been allowed for in the study schedule.

Prior to any formal statistical analyses of the kinds discussed from Section 4 onwards, it is essential to perform simple tabulations of data and to construct simple diagrams to summarize the information that has been collected. Simple statistical *package* computer programs, such as Epi-Info (<http://wwwn.cdc.gov/epiinfo>) or STATA (<http://www.stata.com>), greatly facilitate doing this. The investigator should use these simple approaches to gain a good understanding of the data collected, before embarking on more complex analyses. These simple analysis methods are not described further in this manual, but they are discussed in most good textbooks on medical statistics (for example, Armitage and Berry, 1987; Kirkwood and Sterne, 2003).

If the results of a trial are to be used for submission to an appropriate authority to grant a licence for a new drug or vaccine, the licensing authorities will require that a *statistical analysis plan* (SAP) is developed as a separate document, to be completed after finalizing the protocol and before the code is broken for who is in the intervention and control groups (if it is a blinded trial). The SAP should contain a technical and detailed description of the principal analyses to be conducted on the trial data, which has more detail than would typically be included in the trial protocol. The plan should include detailed procedures for conducting the statistical analysis of the primary and secondary outcome variables and of other relevant data. Often, the licensing authority will require a copy of the SAP for them to examine and approve in advance of a trial being analysed.

It is good practice to prepare a SAP for any trial, even if the results are not to be used for product licensing. In addition to any necessary review by licensing authorities, the SAP should be reviewed and approved by the trial steering committee and also often by the trial data safety and monitoring committee (DSMC). A formal record should be kept of when the statistical analysis plan was finalized, as well as when the final data set was 'frozen' and when the trial was unblinded.

It is common to develop the computer programs for conducting the SAP in advance of breaking the treatment code. To check that these are working properly, some analysts assign study participants at random to intervention or control groups (irrespective of which group they were actually in) and run the programs on these 'test' data. In this way, they are able to check that the final tables are in an appropriate format to be interpreted, once the code is broken. Conducting such a 'dummy run' analysis generally greatly speeds the analysis and interpretation of the trial, once the data are finalized.

Often, when an analysis is conducted, further analyses will be appropriate and prompted by an initial examination of the study results, rather than being pre-planned in the SAP. Such analyses are often called 'exploratory'. They were not specifically planned in advance but were prompted by examination of the trial findings. Such exploratory analyses are sometimes informative and may suggest new hypotheses, but it is important to distinguish them from the analyses that were included in the SAP, as they were suggested by the data, rather than being planned in advance of the code being broken. It is generally wise to interpret the results of such exploratory analyses with caution.

4. Analysis of proportions

4.1. Confidence interval for a single proportion

Methods appropriate for the analysis of proportions are used when the outcome of interest is a *binary* ('yes/no') variable (for example, the proportion of individuals who develop a disease). The standard error of a proportion p, calculated from a sample of n subjects, is estimated as $\sqrt{[p(1-p)/n]}$ For example, if the prevalence of splenomegaly in a random sample of 200 children from a population is found to be 0.40 (40%) (i.e. 80 had splenomegaly), the standard error (SE) is given by:

$$SE(p)=\sqrt{(0.40\times0.60/200)}=0.035(3.5\%).$$

The 95% CI for a proportion is given by $p\pm1.96\times SE(p)$ In the example, the 95% CI is $\times0.035$ or (0.33, 0.47), i.e. 33–47%. There is a 95% chance that the true prevalence of splenomegaly in the population from which the sample of 200 was taken was between 33% and 47%.

4.2. Difference between two proportions

Suppose now that the objective is to compare the proportions observed in two groups of individuals, as is typically the case in a trial, comparing outcomes in an intervention and control group. The standard error of the *difference* between two proportions p^1 and p^2, based on n^1 and n^2 observations, respectively, is estimated approximately as:

$$\sqrt{\{\bar{p}(1-\bar{p})[(1/n1)+(1/n2)]\}}$$

where $\bar{p}=(n1p1+n2p2)/(n1+n2)$.

For example, if the proportions to be compared are 90/300 (30%) and 135/300 (45%), the observed difference between the two proportions is −0.15, $\bar{p}=0.375$, and the standard error of the difference is given by:

$$\sqrt{\{0.375\times0.625[(1/300)+(1/300)]\}}=0.040.$$

The 95% CI for the difference between the proportions is given by $(p1-p2)\pm1.96\times SE$. In the example, this gives $(-0.15)\pm1.96(0.040)$i.e. (−0.23, −0.07), or −23% to −7%.

To test the null hypothesis that there is no true difference between the two proportions, the data are first arranged in a 2 × 2 table, as in Table 21.2.

In the table, a is the number in group 1 who experiences the outcome of interest. The expected value of a, $E(a)$, and the variance of a, $V(a)$, are calculated under the hypothesis of no difference between the two groups:

$$E(a)=m1n1/N \tag{21.1},$$

$$V(a)=n1n2m1m2/[N2(N-1)] \tag{21.2}$$

The chi-squared($\chi 2$)statistic is then calculated. This gives a measure of the extent to which the observed data differ from those expected if the two proportions were truly equal.

$$\chi 2=(|a-E(a)|-0.5)2/V(a) \qquad (21.3)$$

where $|a-E(a)|$ indicates the *absolute* value of $[a-E(a)]$.

The calculated value of $\chi 2$is compared with tables of the chi-squared distribution with one degree of freedom (df). If it exceeds 3.84, then p < 0.05, indicating some evidence of a real difference in the proportions. If it exceeds 6.63, then p < 0.01, and there is strong evidence of a difference.

In the example, *a* = 90; *E(a)* = (225)(300)/600 = 112.50; and *V(a)* = (300 × 300 × 225 × 375)/(600 × 600 × 599) = 35.215. Thus $\chi 2=(|90-112.50|-0.5)2/35.212=13.74$ From tables of the chi-squared distribution, a p-value of 0.0002 is obtained, indicating a difference as large as that observed would be very unlikely to arise by chance if there really was no difference between the two groups.

If any of the quantities *E(a)*, *E(b)*, *E(c)*, or *E(d)* (for example, E(b)=m2n2/N) are less than 5.0 and *N* is less than 40, the$\chi 2$test is invalid, and a test called 'Fisher's exact test' should be used instead (Kirkwood and Sterne, 2003).

4.3. Ratio of two proportions

The ratio of two proportions is sometimes referred to as the *relative risk* (*R*). To construct a CI for a relative risk, the natural logarithm of the estimate of the relative risk is computed (Table 21.2):

$$loge(R)=loge(p1/p2)=loge[(a/n1)/(c/n2)].$$

Its standard error is estimated by:

$$SE[loge(R)]=\surd\{[b/(an1)]+[d/(cn2)]\} \qquad (21.4).$$

The 95% CI for $\log^e(R)$ is given by $loge(R)\pm1.96$and the 95% CI for the relative risk is obtained by taking anti-logarithms.

In the example given in Table 21.2, the relative risk is estimated as$0.30/0.45=0.667$ and$loge(R)=-0.405$The SE[$\log^e(R)$] is estimated as $\surd\{[210/(90 × 300)] + [165/(135 × 300)]\} = 0.109$, and the 95% CI for $\log^e(R)$ is given by $-0.405 + 1.96(0.109)$, i.e. $(-0.619, -0.191)$. Taking anti-logarithms, the 95% CI for the relative risk is (0.538, 0.826).

4.4. Trend test for proportions

Sometimes, it is of interest to examine whether there is a trend in a series of proportions associated with different levels of some underlying characteristic. For example, consider the proportion of leprosy patients who report regularly to collect their monthly drug supply from a clinic when the accessibility of the clinic is rated as very poor, poor, fair, or good (Table 21.3).

A 'score' (x^i) is assigned for each kind of clinic, of which the value relates to the level of accessibility. For example, '0' has been assigned to those with 'very poor' accessibility and '3' to those with 'good' accessibility. A test for the trend in the proportions a1/n1, a2/n2, a3/n3, and a4/n4is provided by testing, as a chi-squared with one df, the expression:

$$\chi 2=N[(N\Sigma aixi)-(A\Sigma nixi)]2/\{A(N-A)[(N\Sigma nixi2)-(\Sigma nixi)2]\} \qquad (21.5).$$

For example, suppose the data are as shown in Table 21.3 (the respective percentages of regular attenders in the four rows are 20%, 30%, 50%, and 60%). The value of $\chi 2$is:

$$150[(150×125)-(63×245)]2/\{63×87[(150×555)-2452]\}=12.95$$

which is highly significant (p = 0.0003), based on a $\chi 2$test with one df. It may be concluded therefore that there is strong evidence that the regularity of drug collection increases with the accessibility of the clinic.

5. Analysis of rates

5.1. Risks, rates, and person-time-at-risk

The terms 'risk' and 'rate' are often used rather loosely and interchangeably to describe the frequencies of events in epidemiological studies. Usually, this is of no great consequence, but, in some circumstances, the distinction is important and, in particular, may affect the way in which a study is analysed. A *risk* is essentially a proportion, or equivalently a probability. The numerator consists of the number of individuals who experience the event of interest (say, develop the disease) in a defined period. The denominator consists of the total number of individuals who were followed for the defined period, some of whom experienced the event of interest (for example, developed the disease) and the remainder of whom did not (ignoring, for the moment, complications that might arise if some individuals are lost to follow-up). A *rate* takes into account both the number of persons at risk and also the duration of observation for each person. In the simplest case, the numerator is the number of individuals who experience the event of interest during the study period (i.e. the same as the numerator for a risk), but the denominator is expressed as the person-time (for example, person-years or person-days) at risk for the individuals in the study.

For example, if 120 persons are observed for 3 years and 40 of them die at some time during the period, and none are otherwise lost to follow-up, the *risk* of death over the 3 years is estimated as 40/120 = 0.33, whereas the death *rate* is estimated as 40/(the number of person-years-at-risk). The denominator for the rate calculation is (80 × 3) + (40 × 1.5) = 300 years, as 80 persons were 'at risk' for the full 3-year period, and 40 were at risk until they died (which, on average, is likely to have been about halfway through the follow-up period if deaths occurred uniformly over the period). Thus, the death rate is 40/300 = 0.133 per person-year-at-risk (which is not the same as the risk of death during the 3 years of 0.33 divided by 3).

Mathematically, it is straightforward to convert rates to risks, and vice versa, if it may be assumed that the rates are constant over time (see, for example, Breslow and Day, 1980). The reason for discussing the distinction in this chapter is that different methods of statistical analysis are appropriate for risks and rates. As mentioned in Section 4, risks are proportions, and thus the methods described in that section are applicable. Modifications of these methods are necessary for the analysis of rates.

Rates are useful if different individuals in a study have been followed for different periods. This may arise if recruitment to the study population is staggered over time, but follow-up is to a common date, or if individuals are lost to follow-up at different times (for example, because of death, migration, or non-co-operation).

An example of the computation of person-years-at-risk in a large study is given in Table 21.4. In this study, a census was done of the study population on the 1 November each year, and the number of persons remaining at risk was ascertained.

Alternatively, the exact period of follow-up may be known for each subject in the study (if the dates of entry and exit are available for each person), in which case these periods would be summed to derive the total person-years-at-risk.

Another situation in which rates, rather than risks, may be more appropriate is when each individual may be at risk of experiencing the event of interest more than once during the study period (for example, an episode of diarrhoea). The incidence rate in the study population would be calculated as the total number of events (for example, episodes of diarrhoea) for those in the study divided by the total person-time-at-risk (which, in this case, would not end at the first episode). Responses such as this can always be converted to a risk by expressing the outcome as the proportion of individuals who experience more than a specified number of events (for example, one or more episodes of diarrhoea), but, in doing this, some information is lost, with a consequent reduction in the power of the study to detect a difference between groups being compared. The analysis of rate data of this kind (where one individual may experience more than one episode of disease) is not straightforward, as the approach depends upon whether it is reasonable to assume that, once an individual has experienced one event, he or she is no more or less likely to experience another event than anyone else in the same intervention group (say, of the same age and sex). Usually, it is not reasonable to make this assumption, as it is frequently found that susceptibility and exposure to disease vary considerably between individuals in ways that cannot be predicted. A simple way out of the analytical problem is to classify individuals, according to whether or not they experienced any events or not. If this is done, the data can either be analysed as a proportion (using the methods given in Section 4) or the individual can be excluded from follow-up for purposes of analysis, from the time the *first* event occurs (i.e. they are not counted as 'at risk' after the first event), and the methods given in Sections 5.2 to 5.5 can be used.

5.2. Confidence interval for a rate

Suppose e is the number of events that occurred during the study period, and the total person-years-at-risk during the period was y. (Note that the period does not have to be measured in 'years'; it could be in, for example, days, weeks, or months.) The event rate (r) is estimated by e/y. For example, suppose 5000 patients who have received a new tuberculosis (TB) vaccine have been followed for 5 years, but, due to losses in follow-up, the total person-years-at-risk is 20 000 (instead of the nearly 25 000 that would have been appropriate if every patient—except the cases whose follow-up period would be counted up to the time they developed TB—had been followed up throughout the 5 years). If the number of new cases of TB that were detected during the follow-up was 80, the estimated incidence rate of TB would be 80/20 000 = 0.0040/person-year, i.e. four per thousand person-years.

The standard error of a rate (r) is, $\sqrt{(r/y)}$ and the approximate 95% CI for the rate is given by r $1.96\sqrt{(r/y)}$. Thus, in the TB example, the 95% CI for the TB incidence rate is:

$$0.0040 \pm 1.96\sqrt{(0.0040/20\ 000)} = 0.0040 \pm 0.0009,$$

i.e. 3.1–4.9 per thousand person-years.

5.3. Difference between two rates

Suppose it is required to compare event rates in two groups, and the number of events and the person-years-at-risk in the two groups are as in Table 21.5.

The standard error of the difference between two rates is given by: $\sqrt{(r_1/y_1 + r_2 y_2)}$ and the 95% CI on the difference is given by $(r_1 - r_2) \pm 1.96 SE$.

Thus, for the example, the 95% CI on the rate difference of the vaccinated, compared to the unvaccinated, group in Table 21.5 is:

$$(0.0041 - 0.0084) \pm 1.96\sqrt{[(0.0041/19\ 470) + (0.0084/19\ 030)]} = -0.0043 \pm 0.0016 \qquad = -0.0059 \text{ to } -0.0027$$

i.e. -5.9 to $-2.7/1000/\text{year}$.

To perform a statistical test, it is necessary to calculate a test statistic, which may be done along similar lines to those described in Section 4.2. If e^1 is the observed number of events among those in group 1 (say, those vaccinated), then:

Expected value of

$$e_1 = E(e_1) = e y_1/y \qquad\qquad (21.7).$$

Variance of

$$e_1 = V(e_1) = e y_1 y_2/y^2 \qquad\qquad (21.8).$$

Then:

$$\chi^2 = (|e_1 - E(e_1)| - 0.5)^2/V(e_1) \qquad\qquad (21.9).$$

And the value of χ^2 is looked up in tables of the χ^2 distribution, with one df, to assess the p-value.

In the example shown in Table 21.5, $e_1 = 80 E(e_1) = 240 \times 19\ 470/38\ 500 = 121.37$ and $V(e_1) = (240 \times 19\ 470)/(38\ 500 \times 38\ 500) = 59.99$

Thus, $\chi^2 = (|80 - 121.37| - 0.5)^2/59.99 = 27.84$ and $p < 0.000001$, indicating that the difference is highly unlikely to have arisen by chance.

5.4. Ratio of two rates

In some situations, the *ratio* of two rates will be of greater interest than their *difference*. For example, vaccine efficacy is usually calculated from a ratio. The test of the null hypothesis is identical in the two situations (i.e. the difference is zero, or the ratio is unity), but the CIs are calculated in a different way.

The ratio of two rates, sometimes called the relative risk, but more correctly called the relative rate, is $(e_1/y_1)/(e_2/y_2)$, and the standard error of the logarithm of this ratio is approximated by $\sqrt{[(1/e_1)+(1/e_2)]}$. In the example given in Table 21.5, the ratio of the rates is 0.489 (corresponding to a vaccine efficacy of 51.1%), and the standard error of the logarithm of the ratio is $\sqrt{[(1/80)+(1/160)]}=0.1369$ The 95% CI of the logarithm of the ratio is given by $-0.715\pm1.96(0.1369)$ i.e. -0.983 to -0.447. Thus, the 95% CI for the ratio of the two rates is 0.37–0.64 (or the 95% CI on the estimate of vaccine efficacy is from 36% to 63%, i.e. $100(1-0.64)$ to $100(1-0.37)$.

5.5. Trend test for rates

Directly analogous to the trend test for proportions described in Section 4.4, there is a similar test for a trend in rates. Suppose data have been collected from the time since the start of a study to the first attack of malaria among children of different ages, and it is of interest to test whether the attack rate declines with age. The data may be summarized, as in Table 21.6.

A 'score' has been assigned to each group. In the example, the scores have been taken as the mid points of the different age groups (for example, those aged 1–2 years range in age from 1.00 to 2.99 years).

A test for trend in the attack rates in the three age groups e_1/y_1 e_2/y_2 and e_3/y_3 is provided by testing the expression, as χ^2 with one df:

$$\chi^2=\{[\Sigma e_i x_i-[(e/y)\ \Sigma y_i x_i]\}^2/\{(e/y^2)[y\Sigma y_i x_i^2-(\Sigma y_i x_i)^2]\}}\qquad(21.10).$$

For example, suppose the malaria attack rates (attacks/weeks-at-risk) were as in Table 21.6, then the value of χ^2 is:

$$\{205-[(60/500)1975]\}^2/\{(60\times500^2)\ [(500\times9537.5)-1975^2]\}=4.91$$

which has an associated p-value of 0.03, and thus there are grounds for believing that, in the study area, the risk of a malaria attack declined with increasing age.

6. Analysis of mean values

6.1. Confidence interval for a mean

If the outcome measure is taken as the mean $\bar{x}$ of a sample of *n* observations, for example, the weights of a sample of newborn infants, the standard error of the mean is given by $\sigma/\sqrt{n}$ where σ is the standard deviation of the variable measured (for example, weights of newborn infants) in the population from which the sample of *n* observations was taken. The 95% CI on the mean is given by $\bar{x}\pm1.96(\sigma/\sqrt{n})$.

In general, σ (the standard deviation in the population) will not be known but must be estimated, based on the *n* observations in the sample. Thus, the estimate of σ is subject to sampling error also, and this must be taken into account in the computation of the CI on the mean. This is done by using a multiplying factor in the CI calculation taken from tables of the t-distribution, rather than from tables of the 'Normal' distribution, on which Table 21.1 was based. The value of the multiplying factor will depend on the size of the sample from which the standard deviation was estimated. For example, for 95% CIs, appropriate multiplying factors for sample sizes of 10, 20, 50, and 100 are 2.26, 2.09, 2.01, and 1.98, respectively. (Note that, in using the tables, the values of *t* are given for different 'degrees of freedom'. In the situation considered here, the degrees of freedom correspond to the sample size minus one, i.e. *n* − 1.) If the sample size is 30 or more, little error is introduced by using the value of 1.96 derived from the normal distribution when calculating 95% CI, rather than the appropriate t-value.

If the estimate of the standard deviation, based on the sample, is *s*, the 95% CI on the mean is given by $\bar{x}\pm t(s/\sqrt{n})$ For example, if the mean birthweight of 25 infants was 3.10 kg and the standard deviation of the weights in the sample was 0.90 kg, the 95% CI would be given by $3.10\pm2.06(0.90/\sqrt{25})$ i.e. $2.73-3.47$ kg, where the multiplying factor 2.06 is taken from a table of the t-distribution corresponding to 24 df.

6.2. Difference between two means

In a trial, it is very common to want to compare the means of observations in different groups, for example, to compare observations from an intervention group with those from a control group. Suppose that two groups are to be compared and the means are $\bar{x}_1$ and $\bar{x}_2$ respectively, and the corresponding standard deviations observed in the groups are s_1 and s_2. The standard error of the difference between the means is given by $\sqrt{\{s[(1/n_1)+(1/n_2)]\}}$, where s is the *pooled* estimate of the standard deviation, based on the observations from the two groups. s is estimated as:

$$s=\sqrt{\{[(n_1-1)s_1^2+(n_2-1)s_2^2]/(n_1+n_2-2)\}}.$$

The 95% CI for the difference between the means is given by:

$$(\bar{x}_1-\bar{x}_2)\pm ts\sqrt{[(1/n_1)+(1/n_2)]}$$

where t is taken from a table of the t-distribution with (n_1+n_2-2) df.

For example, suppose erythrocyte sedimentation rates (ESRs) were measured in an intervention group and in a control group, as shown in Table 21.7. The standard deviation s may be calculated as $\sqrt{\{[(9\times2.41^2)+(11\times2.54^2)]/(10+12-2)\}}=2.48$ and the 95% CI on the difference is given by:

$$(9.7-6.5)\pm\{(2.09\times2.48)\sqrt{[1/10)+(1/12)]}\}=3.2\pm2.2=1.0-5.4.$$

To test the null hypothesis that there is no true difference in the mean ESRs between the two groups, a statistical test must be performed. A test statistic is calculated to assess the probability of the observed results (or more extreme) if there really is no difference between the two groups. The difference of the means divided by the standard error of the difference gives a value of a test statistic that may be looked up in tables of the t-distribution with (n_1+n_2-2) df.

For the example in Table 21.7, the test statistic$=(\bar{x}_1-\bar{x}_2)/\{s\sqrt{[(1/n_1)+(1/n_2)]}\}=3.01$ The associated p-value is 0.0035, i.e. if there really is no effect of the intervention on ESRs, the chance of observing a difference in the means as large or larger than that in the study is 0.35% (i.e. not impossible, but rather unlikely!).

6.3. Analysis of more than two groups

If a study involves the comparison of observations in more than two groups, it is necessary to generalize the methods given in Section 6.2. This is straightforward but is beyond the scope of this book, and the reader is referred to standard statistical texts, such as that by Armitage and Berry (1987) or Kirkwood and Sterne (2003), for details. The relevant sections to which to refer are those on 'one-way analysis of variance'.

Of course, it is always possible to use the methods given in Section 6.2 to compare groups, just two at a time. This is a reasonable approach, but some caution must be exercised when interpreting the findings, as the chances of finding at least one pair to be significantly different (for example, $p < 0.05$) may be substantial, even if there are, in truth, no differences between the groups. To illustrate this, suppose six groups are being compared. In an analysis of variance, the question is asked: 'Considered as a whole, is the variation between the means observed in the six groups more than might be expected to arise by chance if there were no differences in the true means?'. This question may be answered with one statistical test in an analysis of variance, and the null hypothesis may, or may not, be rejected on the basis of this one test. Suppose, however, it was decided to examine all possible pairs of comparisons of the groups. There are 15 possible pairs, and, if a *t-test* was done on each pair, there is a reasonable chance that at least one comparison would be found to be '$p < 0.05$' by chance alone, because of the number of different tests that had been performed. There are ways of adjusting the significance levels to allow for this effect, and the reader is referred to standard texts again for a discussion of 'the multiple comparison problem'.

7. Controlling for confounding variables

7.1. The nature of confounding variables

A risk factor for the disease under study that is differentially distributed among the groups receiving different interventions in which the disease incidence is being compared is called a *confounding* factor. Unless the trial is very small, confounding factors are not likely to bias the comparisons between intervention and control groups in randomized trials, as the process of randomization ensures that any such factors, whether known or unknown, will be

equally distributed in the different groups (apart from random variation). In studies in which those in the different groups have not been allocated at random, the control of confounding factors is a critical component in the analysis. For example, consider a comparative study of TB incidence in persons who received BCG in a routine vaccination programme and those who were not vaccinated. BCG coverage is often higher in urban areas and, independently of any effect of BCG, those living in urban areas also tend to have a higher incidence of TB because of overcrowding and other environmental factors. In this instance, residential status (rural/urban) could be a confounding factor, and, if it is not taken into account in the analyses, any protective effect of BCG against TB might be underestimated. Consider the hypothetical situation depicted in Table 21.8, which shows the incidence of TB over a 10-year period in BCG-vaccinated and unvaccinated individuals in urban and rural areas.

BCG coverage is appreciably higher in the urban population (80%) than in the rural population (50%). Also, in unvaccinated persons, the incidence of TB is higher in the urban population (20 per thousand over 10 years) than in the rural population (10 per thousand). In consequence, although BCG vaccine efficacy is 50% in both urban and rural areas, the estimate obtained from a comparative study, in which the place of residence is ignored, is only 41%. This difference is due to the confounding effect of the place of residence on the estimate of efficacy (the place of residence being related to both the disease incidence and, independently, to the prevalence of vaccination).

7.2. Adjusting for confounding variables

A powerful way of removing the effect of a confounding variable is to restrict comparative analyses to individuals who share a common level of the confounding variable and then to combine the results across the different levels in such a way so as to avoid bias. Thus, in the example in Table 21.8, if the vaccine efficacy was first estimated separately for rural and urban dwellers, and then the two estimates were to be combined, the estimate of efficacy obtained (50%) would be free of the confounding bias of the place of residence. In general, to control for confounding, the study population is divided into a number of *strata.* Within each stratum, individuals share a common level of the confounding variable. Estimates of risk, rate or mean differences, or ratios are made within each stratum, and the resulting estimates are then *pooled* in some way across strata, in order to obtain an overall measure of the effect which is free of any confounding due to the variable on which the stratification was made. Such stratification may be carried out on several confounding variables simultaneously (for example, age and sex).

If it is known, when a study is planned, that it will be necessary to allow for confounding variables in the analysis, it is desirable to give consideration to this at the design stage, both in terms of the information which must be collected and because it will require an increase in the required sample sizes (to achieve the desired statistical power, see Chapter 5). Usually, the necessary increase in sample size to allow for confounding variables is not great (for example, less than 20%), and often the information needed for these sample size calculations is not available before the study starts anyway. Formal methods for calculating sample sizes, allowing for adjustment for confounding variables, are given in Breslow and Day (1987).

7.3. Adjusting risks

7.3.1. Overall test of significance

After stratifying on the basis of the confounding variable(s), the analysis is conducted one stratum at a time, and then the results are pooled. In the ith stratum, the data may be depicted, as shown in Table 21.9.

To test the hypothesis that the relative risk is 1 in all strata or equivalently that the risk difference is zero in each stratum, a generalization of the method given in Section 4.2 may be used. The statistical test is known as the Mantel–Haenszel test.

In the ith stratum:

Expected value of

$$a_i = E(a_i) = m_{1i}n_{1i}/N_i \qquad (21.11).$$

.

Variance of

$$a_i = V(a_i) = n_{1i}n_{2i}m_{1i}m_{2i}/[N_i^2(N_i-1)] \qquad (21.12).$$

.

An overall test of the null hypothesis that the relative risk is unity is given by calculating $\chi^2 = (|\Sigma a_i - \Sigma E(a_i)| - 0.5)^2 / \Sigma V(a_i)$ where the summation is over all strata, which may be tested for statistical significance using tables of the chi-squared distribution with one df.

The calculations are illustrated in Table 21.10, with data on disease incidence rates in vaccinated and unvaccinated individuals in three areas—urban, semi-urban, and rural.

$$\Sigma a_i = 530; \ \Sigma E(a_i) = 738; \ \Sigma V(a_i) = 284.21.$$

Thus:

$$\chi^2 = (|530 - 738] - 0.5)^2 / 284.21 = 151.49.$$

Thus, there is very strong evidence against the null hypothesis, as p < 0.000001 (from tables of the chi-squared distribution).

7.3.2. Pooled estimate of risk difference

If it is considered that the risk *difference* (rather than the risk ratio) is likely to be constant across different strata, a pooled estimate of the common risk difference may be required. This is obtained by taking a weighted average of the risk differences in each stratum, weighting each by the inverse of its variance (as this may be shown to give the 'best' estimate of the common risk difference).

In the ith stratum, the risk difference is:

$$d_i = p_{1i} - p_{2i} = (a_i/n_{1i}) - (c_i/n_{2i})$$

and the variance of the risk difference is:

$$V(d_i) = \{p_i(1-p_i)[(1/n_{1i}) + (1/n_{2i})]\}$$

(as given also in Section 4.2), where:

$$p_i = [n_{1i}p_{1i} + n_{2i}p_{2i}]/(n_{1i} + n_{2i}) = (a_i + c_i)/(n_{1i} + n_{2i}).$$

Now, let $w_i = 1/V(d_i)$.

The pooled estimate of the common risk difference is given by $d = \Sigma w_i d_i / \Sigma w_i$ For the data in the example given in Table 21.10, the computations for the common risk difference are shown in Table 21.11.

Pooled estimate of the common risk difference $d = \Sigma w_i d_i / \Sigma w_i = -0.0083$

7.3.3. Pooled estimate of risk ratio

A pooled estimate of the common risk ratio R across strata may be obtained, using the following formulae.

In the ith stratum, the risk ratio is given by:

$$R_i = (a_i/n_{1i})/(c_i/n_{2i}) = a_i n_{2i}/(c_i n_{1i}) \qquad (21.13).$$

A pooled estimate across all strata is given by:

$$R = \Sigma(a_i n_{2i}/N_i)/\Sigma(c_i n_{1i}/N_i).$$

Thus, for the example in Table 21.10:

$$R=\{[(160)(4000)/20\ 000]+(170)(16\ 000)/40\ 000]+[(200)(40\ 000)/80\ 000]\}/\{[(80)(16\ 000)/20\ 000]+[(240)(24\ 000)/40\ 000]+[(400)(40\ 000)/80\ 000]\}=200/408=0.49.$$

7.3.4. Confidence intervals

The easiest way of obtaining CIs on the estimates of the common risk difference or the common risk ratio is to use the 'test-based' method (Miettinen, 1976).

The approximate 95% CI on the risk difference is given by:

$$d(1\pm1.96/\sqrt{\chi2}) \tag{21.15}$$

Thus, in the example in Tables 21.10 and 21.11, the confidence limits are:

$$-0.0083(1\pm1.96/\sqrt{151.49})=-0.0096\ \text{to}\ -0.0070.$$

The 95% CI on the logarithm of the relative risk is given by:

$$\log_eR(1\pm1.96/\sqrt{\chi2}) \tag{21.16}$$

In the example, the confidence limits are:

$$\log_e(0.49)(1\pm1.96/\sqrt{151.49})=-0.8269\ \text{to}\ -0.5998.$$

And thus the confidence limits on the relative risk are 0.44 to 0.55.

7.4. Adjusting rates

The computations for adjusting rates are very similar to those for adjusting risks and involve only some changes to the formulae given in Section 7.3.

Suppose the results observed in the ith stratum are as shown in Table 21.12.

7.4.1. Overall test of significance

In the ith stratum, e_{1i} is the number of individuals who developed disease in group 1.

Expected value of

$$e_{1i}=E(e_{1i})=e_iy_{1i}/y_i \tag{21.17}$$

Variance of

$$e_{1i}=V(e_{1i})=e_iy_{1i}y_{2i}/y_i^2 \tag{21.18}$$

An overall test of significance (that the common rate ratio is unity or the common rate difference is zero) is given by:

$$\chi2=(|\Sigma e_{1i}-\Sigma E(e_{1i})|-0.5)^2/\Sigma V(e_{1i}) \tag{21.19}$$

where the summation is over all strata.

The value calculated should be looked up in tables of the chi-squared distribution with one df.

7.4.2. Pooled estimate of rate difference

In the ith stratum, the rate difference is:

$d_i = r_{1i} - r_{2i} = (e_{ii}/y_{1i}) - (e_{2i}/y_{2i})$.

Its estimated variance is:

$V(d_i) = (r_{1i}/y_{1i} + r_{2i}/y_{2i})$.

Let $w_i = 1/V(d_i)$ then the estimate of the common rate difference across all strata is given by:

$$d = \sum w_i d_i / \sum w_i.$$

7.4.3. Pooled estimate of rate ratio

In the ith stratum, the rate ratio is:

$$R_i = r_{1i}/r_{2i} = (e_{ii}/y_{1i})/(e_{2i}/y_{2i}) = e_{1i}y_{2i}/(e_{2i}y_{1i}).$$

A pooled estimate of the common rate ratio is given by:

$$R = \Sigma(e_{1i}y_{2i}/y_i)/\Sigma(e_{2i}y_{1i}/y_i) \tag{21.20}.$$

7.4.4. Confidence intervals

The 95% CI on the common rate difference is given by:

$$d(1 \pm 1.96/\sqrt{\chi^2}) \tag{21.21}.$$

The 95% CI on the logarithm of the common rate ratio is given by:

$$\log_e R(1 \pm 1.96/\sqrt{\chi^2}) \tag{21.22}.$$

Example: In Table 21.13, the numerical computations are illustrated, as before, with data on the disease incidence in vaccinated and unvaccinated individuals in three areas: urban, semi-urban, and rural.

Overall test of significance:

$$\chi^2 = (|\Sigma e_{1i} - \Sigma E(e_{1i})| - 0.5)^2/\Sigma V(e_{1i}) \tag{21.23}$$

$$= (|245 - 359| - 0.5)^2/138.40 = 93.08(p < 0.000001).$$

The estimation of the common rate difference is shown in Table 21.14

$$d = \sum w_i d_i / \sum w_i = -0.0066.$$

The 95% CI on the common rate difference is:

$$d(1 \pm 1.96/\sqrt{\chi^2}) = -0.0066(1 \pm 1.96/\sqrt{93.08}) \tag{21.24}$$

$$= -0.0079 \text{ to } -0.0053.$$

The estimate of the common rate ratio is:

$$R = \Sigma(e_{1i}y_{2i}/y_i)/\Sigma(e_{2i}y_{1i}y_i) = \{[(80)(2000)/10\,000] + [(85)(8000)/20\,000] + [(80)(20\,000)/40\,000]\}/\{[(40)(8000)/10\,000] + [(120)(12\,000)/20\,000] + [(200)(20\,000)/40\,000]\} = 90/204 = 0.44.$$

The 95% confidence limits on the logarithm of the common rate ratio is:

$$\log_e R(1 \pm 1.96/\sqrt{\chi^2}) = \log_e(0.44)(1 \pm 1.96/\sqrt{93.08}) = -0.988 \text{ to } -0.654.$$

Taking the anti-logarithm, the 95% confidence limits on the common rate ratio are 0.37 to 0.52.

7.5. Adjusting means

If the outcome variable is a quantitative measure, other than a risk or rate, adjustment for the effects of a confounding variable involves performing a stratified t-test.

A numerical example is given in Table 21.15 where the comparison is between subjects using mosquito nets (intervention group) and those not using them (control group), and the outcome measure (x) is the number of episodes of malaria over a period of 1 year. In this example, age is considered as the confounding variable, and the stratification has been made by dividing the study subjects into three age groups. The size of each subgroup has been made small to simplify the computations for illustrative purposes.

The data may be represented algebraically for those in the ith stratum, as shown in Table 21.16.

An estimate of the common difference in response between the intervention and control groups is obtained by calculating a weighted average of the differences within each stratum:

$$d = \Sigma w_i d_i / \Sigma w_i$$

where $w_i = [n_{1i} n_{2i}/(n_{1i}+n_{2i})]$.

Difference $d_i = \bar{x}_1 - \bar{x}_2$.

Thus, in the example:

$$d = \{[(-0.75)32/12]+[(-0.45)132/23]+[(-0.20)50/15]\}/\{(32/12)+(132/23)+(50/15)\} = -5.25/11.75 = -0.45.$$

An overall test of significance is obtained by calculating a test statistic as:

$$\Sigma w_i d_i / [s\sqrt{\Sigma w_i}]$$

where $s = \sqrt{\{[\Sigma(n_{1i}-1)s_{1i}^2 + \Sigma(n_{2i}-1)s_{2i}^2]/\Sigma(n_{1i}+n_{2i}-2)\}}$ and the value of the test statistic can be compared with tables of the t-distribution with $\Sigma(n_{1i}+n_{2i}-2)$ df.

In the example:

$$s = \sqrt{(23.0265/44)} = 0.7234.$$

The test statistic (44 df) is:

$$-5.25[0.7234 \times \sqrt{(11.74)}] = -2.12.$$

The absolute value is larger than 2.02, which is the tabulated 5% value for t with 44 df. Thus, there is statistically significant evidence regarding the efficacy of intervention; the reduction in the average number of episodes of malaria is estimated as 0.45 per child per year. The 95% confidence limits on the difference are given by:

$$(\Sigma w_i d_i / \Sigma w_i) \pm ts/\Sigma w_i \qquad (21.26)$$

where t is taken from tables of the t-distribution for 95% confidence limits with $\Sigma(n_{1i}+n_{2i}-2)$ df.

Thus, the 95% confidence limits are:

$$-0.45 \pm 2.02/(0.723)/\sqrt{(11.74)} = 0.88 \text{ to} -0.02 \text{ episodes/year/child.}$$

If it is thought that the intervention is likely to affect the response measured in a *relative*, rather than an *absolute*, fashion (i.e. a constant percentage reduction in the number of malaria attacks, rather than a constant absolute reduction in the number of malaria attacks), then it would be appropriate to *transform* the data initially by taking logarithms of the number of attacks (or, say, $\log_e$(number of attacks + 0.1) to avoid zero numbers) and to perform the calculation on the transformed values.

8. Analyses when communities have been randomized

In some intervention studies, communities, rather than individuals, are used as the unit of randomization. If this has been done, it is inappropriate to base analyses on responses of individuals, ignoring the fact that randomization was over larger units. An appropriate method of analysis would be to summarize the response in each sampling unit by a single value and analyse these summary values as though they were individual values.

The analysis of such trials, often called 'cluster randomized trials', is not straightforward, and only simple methods for performing statistical tests are given here. A comprehensive discussion of the design and analysis of such trials is given in Hayes and Moulton (2009).

8.1. Calculation of standardized responses

Often, trials in which communities have been randomized suffer from problems with confounding variables. If the number of units randomized is large, confounding variables are likely to balance out between groups, but, if the number of units is small (as may be the case when communities have been randomized, even though the number of individuals in each community is large), confounding may be a potentially serious problem, and some adjustment should be made in the analysis. One method of doing this is by *standardization.*

Within each community, the sampled population is divided into strata on the basis of the confounding variable(s) (for example, age and sex groups). The average value of the outcome measure is computed for those in each stratum (for example, a disease incidence rate). A weighted average of the rates in the different strata is then computed to give a single 'standardized' measure for the community, the weights being based on some 'standard' population. The same standard population is used for each community, and thus the standardized measures for each community are not biased by the differential composition of each community, with respect to the confounding variable that is being standardized for.

This method is called the 'direct' method of standardization. If the number of individuals in some strata is small, it may be better to use the 'indirect' method, and details of both are given (see Armitage and Berry (1987) for a more detailed discussion of these methods).

Consider a community in which disease risks p^i have been measured for individuals in k strata (for example, age groups). This may be represented in Table 21.17. Also shown are the corresponding data for a 'standard' population. For example, this might be chosen as the combined data for all communities in the study.

The *directly standardized* disease risk for the community (standardized to the standard population) is given by: $(\sum p_i N_i)/\sum N_i$

The *indirectly standardized* disease risk for the community is given by: $[\sum a_i/(\sum n_i P_i)]/(A/N)$

Having calculated standardized values for each community, the means of the standardized values for the intervention communities may be compared with those for the control communities, using a simple *t-test* (see Section 6.2).

It is usually safer, however, to perform a non-parametric test if the assumptions underlying the *t-test* are in any doubt (Armitage and Berry, 1987)), as it may be impossible to verify the assumptions if the study involves a small number of communities.

8.2. Non-parametric rank sum test

Suppose there are n^1 communities in one group and n^2 in the other $(n1 \leq n2)$ and a summary response has been derived for each community. To perform a non-parametric test, consider all the $(n1+n2)$ observations together, and rank them, giving a rank of 1 to the smallest value and $(n1+n2)$ to the highest. Tied ranks are allotted the mid rank of the group. Let T^1 = sum of the ranks in group 1 with n^1 observations. Under the null hypothesis, the expectation of $T1 = n1(n1+n2+1)/2$. Then calculate:

$$T1 = T1,$$

if T^1 is less than or equal to the expected value

$$T1 = n1(n1+n2+1) - T1,$$

if T^1 is more than its expected value.

T^1 may be compared with tabulated critical values (see Table A8 of Armitage and Berry, 1987) to determine the statistical significance.

Consider the example shown in Table 21.18 in which age-standardized leprosy prevalence rates are compared in 12 'intervention' villages and ten 'control' villages.

The expected value of $T1=n1(n1+n2+1)/2=10(10+12+1)/2=115$. As T^1 is greater than its expectation, T^1 is $10(10+12+1)−142.5=87.5$ The critical value of T^1 at the 5% level of significance is 84 (from tables in Armitage and Berry, 1987). As T^1 is greater than the critical value, it is concluded that the intervention has not had a statistically significant effect (the average prevalence was 6.5 per thousand in intervention villages, and 8.7 per thousand in control villages).

8.3. Tests on paired data

In some study designs, communities may be 'paired' on the basis of similarity, with respect to confounding variables and baseline disease prevalence or incidence rates. Within each pair of communities, one receives the intervention and the other serves as the control. If this has been done, the analysis should take the pairing into account.

First, standardized response rates are computed for each community (as discussed in Section 8.1), and then the standardized response rates are compared using a paired t-test (Armitage and Berry, 1987) or a non-parametric test.

To perform a paired t-test for n pairs of communities, suppose d^i is the difference in outcome measured between the intervention and control unit for the ith pair. Calculate a test statistic $(\Sigma di/n)/(s/\sqrt{n})$ where s is the standard deviation of the n differences. This value of the test statistic may be compared to tabulated values of the t-distribution with $(n − 1)$ df.

Consider the data shown in Table 21.19, which shows leprosy prevalence rates in ten pairs of communities.

The mean difference $d=−19/10=−1.9$ and the standard deviation of the difference (s) is 2.23. Thus, the test statistic $=−1.9/(2.23/\sqrt{10})=−2.69$ with 9 df. From tables of the t-distribution, p is <0.05, and it may be concluded that the prevalence of leprosy is significantly lower in the intervention villages.

Alternatively, a non-parametric test may be preferred. In this instance, the appropriate such test is Wilcoxon's signed rank test.

The differences between each pair of villages are arranged in ascending order of magnitude of the absolute value of the differences (i.e. ignoring the sign) and given ranks 1 to n; zero values are excluded from analysis. Any group of tied ranks is allotted the mid rank of the group. Let:

   T+ = sum of ranks of positive differences

   T− = sum of ranks of negative differences.

The smaller of the two (T+ and T−) is compared with the tabulated critical value (see Table A9 of Armitage and Berry, 1987). If it is lower than the tabulated value, it is concluded that there is a significant difference. For the data in the table, $T+=4.5$ and $T−=40.5 n=9$ (excluding one zero difference). The tabulated critical 5% value is 5. Since $T+=4.5$ is less than 5, it is concluded that the difference is significant at the 5% level.

9. Prevented fraction of disease

The objective of most field trials is to measure the effect of an intervention in reducing disease rates. The results of such studies may be used to estimate the impact that an intervention might have on disease rates if it was introduced into a public health programme. In such circumstances, the overall effect is much influenced by the *coverage* achieved by the programme.

The *prevented fraction* among individuals exposed to an intervention measure is defined as the percentage of the disease incidence in such individuals that has been prevented due to having received the intervention. For example, if the efficacy of BCG vaccination against TB is 60%, among persons who receive BCG vaccination, 60% of the TB cases that would have developed otherwise have been prevented by the vaccination. For vaccine studies, the

prevented fraction is directly equivalent to the *vaccine efficacy,* but the former term may be used for interventions other than vaccines.

The prevented fraction is computed by subtracting the disease risk in individuals with the intervention measure (for example, an anti-leprosy vaccine) from the disease risk in individuals without the intervention, and expressing the difference as a proportion of the latter. For example, if the annual incidence of leprosy is 2.8 per thousand in the vaccinated and 4.2 in the unvaccinated, the prevented fraction is equal to $[(4.2-2.8)/4.2]=0.33$(or 33%).

If the relative risk (R) (of disease in those who receive the intervention, compared to those who do not) is known, the prevented fraction may be obtained by calculating $(1-R)$For example, if the relative risk of developing malaria in homes where mosquito-nets are used is a quarter of that in homes where they are not used, the prevented fraction is equal to$1-0.25$, i.e. 75%.

The *population prevented fraction* is defined as the proportion of cases of the disease in the *total population* that have been prevented by the intervention. If the relative risk (R) and the proportion of individuals in the population who receive the intervention measure (P) are known, the population prevented fraction is obtained by calculating $P(1-R)$Thus, the extent of reduction possible in disease incidence in the total population, if all individuals were to receive the intervention measure$(P=1)$is $(1-R)$

Consider a situation in which the annual incidence of TB is 2.0 per thousand in those who do not receive BCG vaccination and 0.8 per thousand in those who do, i.e. the relative risk in those vaccinated is$0.8/2.0=0.4$Table 21.20 shows the fraction of all cases prevented by the intervention, according to the disease incidence in the total population and the vaccination coverage.

References

Armitage, P. and Berry, G. 1987. *Statistical methods in medical research*. Oxford: Blackwell Scientific.

Breslow, N. E. and Day, N. E. 1980. *Statistical methods in cancer research. Volume 1. The analysis of case-control studies*. Lyon: International Agency for Research on Cancer.

Breslow, N. E. and Day, N. E. 1987. *Statistical methods in cancer research. Volume 2. The design and analysis of cohort studies*. Lyon: International Agency for Research on Cancer.

Hayes, R. J. and Moulton, L. H. 2009. *Cluster randomized trials*. Boca Raton, FL: Chapman & Hall/CRC.10.1201/9781584888178 [CrossRef]

Kirkwood, B. R. and Sterne, J. A. C. 2003. *Essential medical statistics*. Malden, MA: Blackwell Science.

Miettinen, O. 1976. Estimability and estimation in case-referent studies. *American Journal of Epidemiology*, **103**, 226–35. [PubMed: 1251836]

Rothman, K. J., Greenland, S., and Lash, T. L. 2008. *Chapter 10: Precision and statistics in epidemiologic studies. Modern epidemiology*, 3rd ed. Philadelphia: Lippincott Williams & Wilkins.

Tables

Table 21.1   Multiplying factors for calculating CIs, based on the Normal distribution

CI (%)	Multiplying factor
90	1.64
95	1.96
99	2.58
99.9	3.29

Tables

Table 21.1   Multiplying factors for calculating CIs, based on the Normal distribution

CI (%)	Multiplying factor

Table 21.2   Comparison of two proportions

Group	Outcome		Total	Proportion with outcome
	Yes	**No**		
1	a (90)	b (210)	n^1 (300)	$p^1 = a/n^1$ (0.30)
2	c (135)	d (165)	n^2 (300)	$p^2 = c/n^2$ (0.45)
Total	m^1 (225)	m^2 (375)	N (600)	

Table 21.3   Regularity of collection of drugs by leprosy patients, according to accessibility of clinic

Accessibility of clinic	Collection of drugs		Total	'Score' x^i
	Regular	Not regular		
Very poor	a^1 (5)	$n^1 - a^1$ (20)	n^1 (25)	0
Poor	a^2 (12)	$n^2 - a^2$ (28)	n^2 (40)	1
Fair	a^3 (25)	$n^3 - a^3$ (25)	n^3 (50)	2
Good	a^4 (21)	$n^4 - a^4$ (14)	n^4 (35)	3
Total	A (63)	$N - A$ (87)	N (150)	

Table 21.3   Regularity of collection of drugs by leprosy patients, according to accessibility of clinic

Accessibility of clinic	Collection of drugs		Total	'Score' x^i
	Regular	Not regular		
Very poor	a^1 (5)	$n^1 - a^1$ (20)	n^1 (25)	

Table 21.4　Example of the computation of person-years-at-risk in a large study

	Date	No. of persons under observation	Average of successive numbers	Years of observation	Person-years
	(a)	(b)	(c)	(d)	(c × d)
Start date	1 November 2004	10140			
	1 November 2005	9145	9642.5[*]	1	9642.5
	1 November 2006	8232	8688.5	1	8688.5
	1 November 2007	7389	7810.5	1	7810.5
	1 November 2008	6281	6835.0	1	6835.0
End date	1 April 2009	5779	6030.0	5/12	2512.5
				Total	35489.0

[*]　If 10140 persons were alive on 1 November 2004, and 9145 of them were known to be alive on 1 November 2005, and if losses to follow-up occurred evenly throughout the year, there would have been, on average, (10140 + 9145)/2 = 9642.5 persons at risk on each day during the first year, hence a total of 9642.5 person-years.

Table 21.5   TB incidence rates in vaccinated and unvaccinated groups

	Number of events (new TB cases)	Person-years-at-risk (pyar)	Event rate (TB cases per 1000 pyar)
Vaccinated	e^1 (80)	y^1 (19470)	r^1 (4.1)
Not vaccinated	e^2 (160)	y^2 (19030)	r^2 (8.4)
Total	e (240)	y (38500)	r

	Number of events (new TB cases)	Person-years-at-risk (pyar)	Event rate (TB cases per 1000 pyar)
Vaccinated	e^1 (80)	y^1 (19470)	r^1 (4.1)
Not vaccinated	e^2 (160)	y^2 (19030)	r^2 (8.4)

Table 21.6   Malaria attack rates in children of different ages

Age of children (years)	'Score' x^i	No. with malaria attack (e^i)	Child-weeks-at-risk (y^i)	Attack rate e^i/y^i
1–2	x^1 (= 2.0)	e^1 (30)	y^1 (200)	0.150
3–4	x^2 (= 4.0)	e^2 (20)	y^2 (150)	0.133
5–7	x^3 (= 6.5)	e^3 (10)	y^3 (150)	0.067
Total		e (60)	y (500)	0.120

Table 21.6   Malaria attack rates in children of different ages

Age of children (years)	'Score' x^i	No. with malaria attack (e^i)	Child-weeks-at-risk (y^i)	Attack rate e^i/y^i
1–2	x^1 (= 2.0)	e^1 (30)	y^1 (200)	
3–4	x^2 (= 4.0)	e^2 (20)	y^2 (150)	

Table 21.7   ESR in an intervention and a control group

	Intervention group (i = 1)	Control group (i = 2)
Number of subjects (n^i)	10	12
Mean ESR ($\bar{x}^i$)	9.7	6.5
Standard deviation (s^i)	2.41	2.54

Table 21.8   TB incidence rates by BCG vaccination status and urban or rural residence

BCG vaccination status	Urban			Rural			Both groups		
	Total popn.	TB cases[*]		Total popn.	TB cases[*]		Total popn.	TB cases[*]	
		No.	/1000		No.	/1000		No.	/1000
Vaccinated	16000	160	10	40000	200	5	56000	360	6.4
Unvaccinated	4000	80	20	40000	400	10	44000	480	10.9
Vaccine efficacy		50%			50%				41%

*   Over a period of 10 years.

Table 21.9   Comparison of proportions developing disease in two intervention groups for individuals in the ith stratum

Intervention group	Developed disease	Did not develop disease	Total
1	a^i	b^i	$a^i + b^i = n^{1i}$
2	c^i	d^i	$c^i + d^i = n^{2i}$
Total	$a^i + c^i = m^{1i}$	$b^i + d^i = m^{2i}$	N^i

Table 21.10 Disease incidence rates in urban, semi-urban, and rural areas, according to vaccination status

Area (i)	Vaccinated			Unvaccinated			Grand total (N^i)	E (a^i)	V (a^i)
	Cases (a^i)	Non-cases (b^i)	Total (n^{1i})	Cases (c^i)	Non-cases (d^i)	Total (n^{2i})			
Urban (1)	160	15840	16000	80	3920	4000	20000	192	37.94
Semi-urban (2)	170	23830	24000	240	15760	16000	40000	246	97.39
Rural (3)	200	39800	40000	400	39600	40000	80000	300	148.88
Total	530	79470	80000	720	59280	60000	140000	738	284.21

Table 21.11   Computation of the common risk difference for the data in Table 21.10

Area	p^{1i}	p^{2i}	d^i	$V(d^i)$	w^i
Urban	0.0100	0.0200	−0.0100	3.705×10^{-6}	270×10^3
Semi-urban	0.0071	0.0150	−0.0079	1.057×10^{-6}	946×10^3
Rural	0.0050	0.0100	−0.0050	37.219×10^{-6}	27×10^3

Table 21.11   Computation of the common risk difference for the data in Table 21.10

Area	p^{1i}	p^{2i}	d^i	$V(d^i)$	w^i
Urban	0.0100	0.0200	−0.0100	3.705×10^{-6}	270×10^3
Semi-urban	0.0071	0.0150	−0.0079	1.057×10^{-6}	946×10^3

Table 21.12   Disease rates in two intervention groups for individuals in the *i*th stratum

Intervention group	Developed disease	Person-years-at-risk
1	e_{1i}	y_{1i}
2	e_{2i}	y_{2i}
Total	e_i	y_i

Table 21.13   Disease incidence rates in urban, semi-urban, and rural areas, according to vaccination status

Area (i)	Vaccinated		Unvaccinated		Both groups		$E(e^{1i})$	$V(e^{1i})$
	Cases (e^{1i})	Person-years (y^{1i})	Cases (e^{2i})	Person-years (y^{2i})	Cases (e^i)	Person-years (y^i)		
Urban (1)	80	8000	40	2000	120	10000	96	19.2
Semi-urban (2)	85	12000	120	8000	205	20000	123	49.2
Rural (3)	80	20000	200	20000	280	40000	140	70.0
Total	245	40000	360	30000	605	70000	359	138.4

Table 21.14　Computation of the common rate difference for the data in Table 21.13

Area (i)	r^{1i}	r^{2i}	Difference d^i	$V(d^i)$	$w^i = 1/V(d^i)$
Urban (1)	0.0100	0.0200	−0.0100	11.250×10^{-6}	89×10^3
Semi-urban (2)	0.0071	0.0150	−0.0079	2.465×10^{-6}	406×10^3
Rural (3)	0.0040	0.0100	−0.0060	0.700×10^{-6}	1429×10^3

Table 21.15   Attacks of malaria in children of different ages in those using (intervention) and not using (control) mosquito-nets

(Stratum) Age group		Attacks of malaria/child (x)	No. of children (n)	Mean ($\bar{x}$)	Standard deviation (s)
(1) <2y	I*	1,0,2,3,1,2,1,0	8	1.25	1.0351
	C*	2,3,1,2	4	2.00	0.8165
(2) 2–3y	I	0,1,1,2,1,1,0,2,2,1,1,0	12	1.00	0.7385
	C	2,2,1,1,1,2,1,1,2,2,1	11	1.45	0.5222
(3) 4–5y	I	1,0,1,1,1	5	0.80	0.4472
	C	1,1,2,0,1,1,0,2,1,1	10	1.00	0.6667

*   I, intervention group; C, control group.

Table 21.16   Algebraic representation of data in Table 21.15 for those in the *i*th stratum

Intervention group	No. in group	Mean	Standard deviation
1 (I)	n^{1i}	$\bar{X}^{1i}$	s^{1i}
2(C)	n^{2i}	$\bar{X}^{2i}$	s^{2i}

Intervention group	No. in group	Mean	Standard deviation
1 (I)	n^{1i}	$\bar{X}^{1i}$	s^{1i}
2(C)	n^{2i}	$\bar{X}^{2i}$	s^{2i}

Table 21.17   Disease risks in study community and in standard population in each of _k_ strata

Stratum	Study community			Standard population		
	Total	Cases		Total	Cases	
		No.	Risk		No.	Risk
1	n^1	a^1	p^1	N^1	A^1	P^1
2	n^2	a^2	p^2	N^2	A^2	P^2
_i_	n^i	a^i	p^i	N^i	A^i	P^i
.	.	.	.	.	.	.
_k_	n^k	a^k	p^k	N^k	A^k	P^k
Total	_n_	_a_	_p_	_N_	_A_	_P_

Table 21.18   Age-standardized leprosy prevalence rates and ranks in 12 'intervention', and ten 'control', villages

Intervention villages		Control villages	
Prevalence rate/1000	Rank	Prevalence rate/1000	Rank
3	1.5	10	18.5
9	15.5	13	22
8	12.5	6	7.5
6	7.5	11	21
5	5	10	18.5
5	5	7	9.5
7	9.5	8	12.5
3	1.5	8	12.5
10	18.5	5	5
8	12.5	9	15.5
10	18.5		
4	3		
Sum of ranks	$T^2 = 110.5$		$T^1 = 142.5$

Table 21.19   Leprosy prevalence rates in ten pairs of communities

Village pair no.	Prevalence of leprosy (per thousand)		Difference (d^i)	Rank (ignoring sign)
	Intervention	**No Intervention**		
1	6	10	−4	8
2	9	13	−4	8
3	3	6	−3	5
4	12	11	+1	1.5
5	10	10	0	
6	4	7	−3	5
7	7	8	−1	1.5
8	5	8	−3	5
9	7	5	+2	3
10	5	9	−4	8

Table 21.20   Population prevented fraction, according to vaccination coverage and disease incidence rate

Vaccination coverage (%) (P × 100)	Disease incidence in total population (per thousand) [0.8 P + 2.0(1 − P)]	Population prevented fraction (%) [P (1 − R) × 100]
0	2.00	0
20	1.76	12
40	1.52	24
60	1.28	36
80	1.04	48
100	0.80	60

Bookshelf ID: NBK305516