Estimation: Understanding confidence interval

A Cook, A Sheik

INTRODUCTIO

The previous paper of this series described hypothesi testing, showing how a p-value indicates the likelihood of a genuine difference existing betwee two or more groups. An alternative to using p-value alone, to show whether or not differences exist, is t report the actual size of the difference. Since a difference observed between groups is subject to random variation it becomes necessary to present no only the difference, but also a range of values aroun the observed difference within which it is believed th true value will lie. Such a range is known as a confidence interval

Confidence intervals may be calculated for means proportions, differences between means or proportions, relative risks, odds ratios and many othe summary statistics. Here we describe in detail onl one simple calculation, that of a confidence interva for a proportion. Using examples from the literatur we look at the interpretation of other confidence intervals, and show the relationship between p-value and confidence intervals

WHY P-VALUES ALONE ARE NOT ENOUG

P-values express statistical significance, but statistically significant results may have little clinica significance. This is particularly the case with larg studies that have power to detect very small differences. For example, an improvement in averag peak flow of 1 l/min, when comparing interventio and control groups, may be statistically significant bu clearly has little clinical benefit

A further drawback of p-values is the emphasis place on p=0.05, a value chosen purely by convention bu which has spawned a tendency to dismiss anythin larger and focus attention on smaller values only. B presenting instead a confidence interval, one require the reader to think about what the values actuall mean, thus interpreting the results more fully

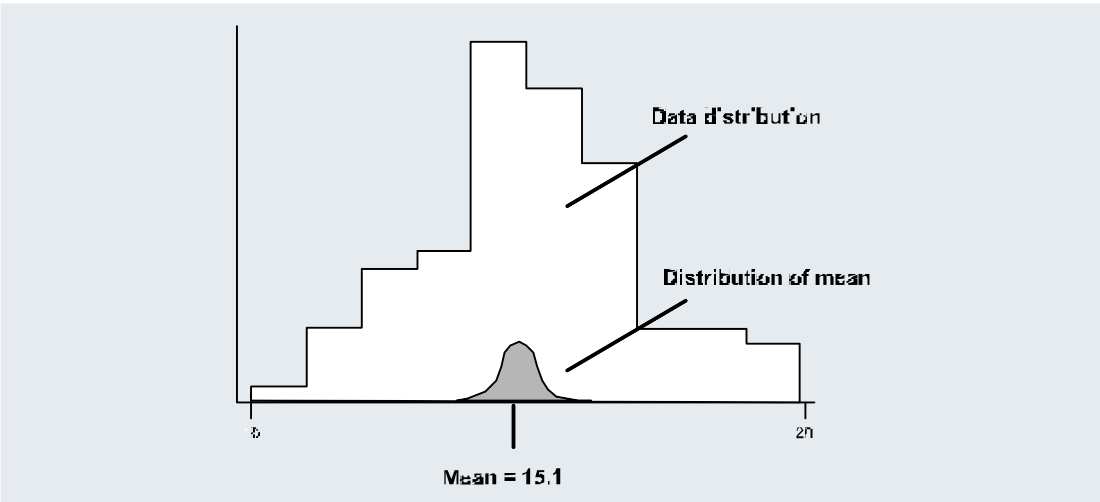
CALCULATING THE CONFIDENCE INTER FOR A PROPORTIO

Kaur *et a* carried out a prevalence study of asthm symptoms and diagnosis in British 12-14 year olds [1] From a sample of 27,507 children, 20.8% (n=5,736 reported 'ever having' asthma. Assuming the sampl to be representative, the true national prevalenc should be close to this figure, but it remains unknow and a different sample would probably yield a slightly different estimate. To calculate a range likel to contain the true figure, we need t know by how much the sample proportion is likely t vary. Put more technically, we need t know the standard deviation of the sample statistic This is known as the standard error

One way of finding the standard error would be t take several more samples, calculate the proportio 'ever having' asthma separately in each sample, the calculate the standard deviation of these proportions Fortunately, such labour is unnecessary because it ha been shown that most summary statistics follow normal distributions, particularly when sample size are large. Furthermore, the standard deviations o these distributions are directly related to the standar deviation of the original data. The situation is illustrated in Figure 1, which shows a possible distribution of data, together with the expected distribution of mean values generated by data samples

With data from a normally distributed variable, 95 of observations should lie within two standard deviations of the mean. Having said that a sample statistic is expected to be normally distributed, it

Figure 1. Expected distribution of a sample mea



Adrian Cook
Statisticia

Aziz Sheik
NHS R&D National Primar Care Training Fello
Department of Primary Healt Care & General Practic
Imperial College School o Medicin

Correspondence to
Adrian Coo
Department of Primary Healt Care & General Practic
ICSM Charing Cross Campus
Reynolds Building
St Dunstan's Road,
London, W6 8R
a.d.cook@ic.ac.u

Date submitted: 19/7/0
Date accepted: 17/10/0

Prim Care Respir J 2000 9(3); 48-5

Table 1. Formulae for calculating standard erro

Summary statisti	Formul	Symbo	Standard error (SE
Mea	$\frac{\sum x}{n}$	$\bar{x}$	$\frac{sd(x)}{\sqrt{n}}$
Proportio	$\frac{cases}{cases + noncases}$	p	$\sqrt{\frac{p(1-p)}{n}}$
Difference in mean	$\frac{\sum x_1}{n_1} - \frac{\sum x_2}{n_2}$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{sd(x_1)^2}{n_1} + \frac{sd(x_2)^2}{n_2}}$
Difference in proportion	$\frac{cases_1}{cases_1 + noncases_1} - \frac{cases_2}{cases_2 + noncases_2}$	$p_1 - p_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
Relative ris	$\frac{cases_1/(cases_1 + noncases_1)}{cases_2/(cases_2 + noncases_2)}$	RR	$\sqrt{\frac{1}{cases_1} - \frac{1}{cases_1 + noncases_1} + \frac{1}{cases_2} - \frac{1}{cases_2 + noncases_2}}$ *
Odds rati	$\frac{cases_1/noncases_1}{cases_2/noncases_2}$	OR	$\sqrt{\frac{1}{cases_1} + \frac{1}{noncases_1} + \frac{1}{cases_2} + \frac{1}{noncases_2}}$ *

*Standard error of the logged statistic

follows that on 95% of occasions it will be less tha two standard errors from its true value. The probabil - ty of the range formed by the sample statistic plus o minus two standard errors containing the true value i therefore 95%. This range is the 95% confidence interval. Formulae for the standard error of commo summary statistics are in Table 1, those for other statistics are usually readily available in textbooks. Fo the example of asthma prevalence, the standard erro can be calculated as 0.25% giving a 95% confidenc interval of 20.3% to 21.3%. The narrowness of thi confidence interval reflects the large sample size, illustrating how certainty in a result grows as the number of observations increase, resulting in smalle standard errors and narrower confidence intervals

INTERPRETING CONFIDENCE INTERVAL FOR RELATIVE RISK

Wald and Watt compared all-cause mortality amon different types of smoker with that of lifelong non smokers [2] Compared to non-smokers, the relative ris (RR) of mortality among former cigarette smokers wa 1.11 (95% confidence interval 0.92 to 1.34). The bes estimate of the effect on mortality is an increase o 11%, RR=1.11, but the possibility of no effec (RR=1.0) remains. Among current smokers the rel - tive mortality was 2.26 (95% confidence interval 1.9 to 2.58). This confidence interval does not includ RR=1.0 and so we can be confident that mortality i higher among current smokers

Among pipe and cigar smokers who had neve smoked cigarettes, mortality compared to non-smoker was 1.23 (95% confidence interval 0.99 to 1.75) Relative mortality is higher than that of ex-smokers but the confidence interval is much wider. The greate width is partly due to the pipe/cigar group bein smaller than the ex-smokers group, one more or on fewer death thus has a greater effect on mortality an the wider confidence interval reflects the less stabl result

CONFIDENCE INTERVALS AND P-VALUES

It may have become apparent that the statistical significance of differences can be gleaned from confidence intervals. A confidence interval containin 1.0 for a relative risk or an odds ratio means we ar less than 95% sure that a genuine difference exists, significance test of the difference would thus giv p>0.05. Similarly a confidence interval not includin 1.0 corresponds to p<0.05, while an interval bounde at one end by 1.0 exactly would give p=0.05. A sim - lar situation exists with confidence intervals for differences in means or proportions, the only difference being that no effect is represented by th value 0.0, rather than 1.0

The practice of reporting confidence intervals togethe with p-values is questionable, p-values adding littl information for the informed reader. An exception t this rule occurs when a large number of confidenc intervals are reported, in this instance the generall discouraged habit of replacing p-values with stars indicating p<0.05 and p<0.01 becomes useful, allowing a rapid overview of results to be made

CONCLUSION

Here we have outlined the theory and practice of calculating confidence intervals, and give pointers toward their meaningful interpretation

Table 2. Relative all-cause mortality of different smoking group

Smoking grou	n	die	RR †	95% C
Lifelong non-smoke	653	84	1.0	
Former cigarette smoke	546	126	1.1	0.92 to 1.3
Pipe/cigar smoke never smoked cigarette	930	31	1.2	0.99 to 1.7
Current cigarette smoke	218	64	2.2	1.97 to 2.5

†Adjusted for age at entry to stud

Clinical significance may be gauged both from th point estimate of the difference, and consideration o the confidence limit's upper and lower bounds Whether or not a confidence interval contains unity fo a relative difference, or zero for an absolute differenc reveals statistical significance. Because they conve both aspects of significance, confidence intervals hav become the strongly preferred way of presentin results ■

Acknowledgement
AS is supported by an NHS R&D National Primar Care Training Fellowship

Recommended readin
Gardner MJ, Altman DG. *Statistics with confidence* London: BMJ, 1993

Reference
1 Kaur B, Ross Anderson H, Austin J *et a* . Prevalence of asthm symptoms, diagnosis, and treatment in 12-14 year old childre across Great Britain (international study of asthma and allergie in childhood, ISAAC UK). *BM* 1998 **31** :118-24
2. Wald NJ, Watt HC. Prospective study of effect of switching fro cigarettes to pipes or cigars on mortality from three smokin related diseases. *BM* 1997 **31** :1860-3

Erratu

In the June 2000 issue of the *Primary Care Respiratory Journa* , reference: Cropper JA, Frank TL, Fran PI, Hannaford PC. Primary care workload and prescribing costs in children. The impact of respiratory symptoms. *Prim Care Respir* 2000; **9**(1);8-11. The table should read as follows

Table 2: Percentage of children having at least one consultation or prescription i primary care by positive response

	Positive response categor				c^2 test for linear tre d[a]	p val e[a]
	0	1–2	3	4–5		
Total in each grou	56	76	29	98		
Total surgery consultation	92.7	98.8	96.4	98.4	9.8	0.02
Lower respiratory consultation	29.1	37.7	64.0	77.3	206.0	<0.00
Upper respiratory consultation	51.5	59.8	72.4	74.1	25.2	<0.00
Non-respiratory consultation	89.1	93.4	90.1	93.7	1.2	0.25
Total home visit	34.5	46.1	52.1	55.0	53.9	0.00
Lower respiratory home visit	3.0	5.4	18.2	23.3	62.9	<0.00
Upper respiratory home visit	18.8	26.3	33.8	32.3	20.2	0.00
Non-respiratory home visit	18.2	24.5	23.4	27.5	9.8	0.16
Unknown cause home visit	4.8	4.8	10.4	10.1	6.3	0.03
Total number of prescription	81.2	91.6	92.7	97.4	25.	<0.00
Non-respiratory prescription	70.3	83.2	82.3	86.2	11.4	0.00
Respiratory prescription	67.3	72.4	86.4	93.6	60.1	<0.00
BNF 3 1[b] prescription	9.1	13.7	50	65.1	568.6	<0.00
BNF 3 2[c] prescription	1.8	3.6	20.3	39.7	412.8	<0.00
BNF 5 1[d] prescription	55.7	61.1	76.0	83.1	58.1	<0.00
BNF 6 3[e] prescription	0.6	0	8.3	9.0	25.5	<0.00

[a]Calculated using discrete positive response values; [b]bronchodilators; [c]inhaled steroids [d]antibiotics; [e]oral steroid