Statistics Notes

Interaction revisited: the difference between two estimates

Douglas G Altman, J Martin Bland

We often want to compare two estimates of the same quantity derived from separate analyses. Thus we might want to compare the treatment effect in subgroups in a randomised trial, such as two age groups. The term for such a comparison is a test of interaction. In earlier Statistics Notes we discussed interaction in terms of heterogeneity of treatment effect.[1–3] Here we revisit interaction and consider the concept more generally.

The comparison of two estimated quantities, such as means or proportions, each with its standard error, is a general method that can be applied widely. The two estimates should be independent, not obtained from the same individuals—examples are the results from subgroups in a randomised trial or from two independent studies. The samples should be large. If the estimates are E_1 and E_2 with standard errors SE(E_1) and SE(E_2), then the difference $d=E_1-E_2$ has standard error $\mathrm{SE}(d)=\sqrt{[\mathrm{SE}(E_1)^2+\mathrm{SE}(E_2)^2]}$ (that is, the square root of the sum of the squares of the separate standard errors). This formula is an example of a well known relation that the variance of the difference between two estimates is the sum of the separate variances (here the variance is the square of the standard error). Then the ratio $z=d/\mathrm{SE}(d)$ gives a test of the null hypothesis that in the population the difference d is zero, by comparing the value of z to the standard normal distribution. The 95% confidence interval for the difference is d–1.96SE(d) to d+1.96SE(d).

We illustrated this for means and proportions,[3] although we did not show how to get the standard error of the difference. Here we consider comparing relative risks or odds ratios. These measures are always analysed on the log scale because the distributions of the log ratios tend to be those closer to normal than of the ratios themselves.

In a meta-analysis of non-vertebral fractures in randomised trials of hormone replacement therapy the estimated relative risk from 22 trials was 0.73 (P=0.02) in favour of hormone replacement therapy.[4] From 14 trials of women aged on average <60 years the relative risk was 0.67 (95% confidence interval 0.46 to 0.98; P=0.03). From eight trials of women aged ⩾60 the relative risk was 0.88 (0.71 to 1.08; P=0.22). In other words, in younger women the estimated treatment benefit was a 33% reduction in risk of fracture, which was statistically significant, compared with a 12% reduction in older women, which was not significant. But are the relative risks from the subgroups significantly different from each other? We show how to answer this question using just the summary data quoted.

Because the calculations were made on the log scale, comparing the two estimates is complex (see table). We need to obtain the logs of the relative risks and their confidence intervals (rows 2 and 4).[5] As 95% confidence intervals are obtained as 1.96 standard errors either side of the estimate, the SE of each log relative risk is obtained by dividing the width of its confidence interval by 2×1.96 (row 6). The estimated difference in log relative risks is $d=E_1-E_2=-0.2726$ and its standard error 0.2206 (row 8). From these two values we can test the interaction and estimate the ratio of the relative risks (with confidence interval). The test of interaction is the ratio of d to its standard error: $z=-0.2726/0.2206=-1.24$, which gives P=0.2 when we refer it to a table of the normal distribution. The estimated interaction effect is exp(−0.2726)=0.76. (This value can also be obtained directly as 0.67/0.88=0.76.) The confidence interval for this effect is −0.7050 to 0.1598 on the log scale (row 9). Transforming back to the relative risk scale, we get 0.49 to 1.17 (row 12). There is thus no good evidence to support a different treatment effect in younger and older women.

The same approach is used for comparing odds ratios. Comparing means or regression coefficients is simpler as there is no log transformation. The two estimates must be independent: the method should not be used to compare a subset with the whole group, or two estimates from the same patients.

There is limited power to detect interactions, even in a meta-analysis combining the results from several studies. As this example illustrates, even when the two estimates and P values seem very different the test of interaction may not be significant. It is not sufficient for the relative risk to be significant in one subgroup and not in another. Conversely, it is not correct to assume that when two confidence intervals overlap the two estimates are *not* significantly different.[6] Statistical analysis should be targeted on the question in hand, and not based on comparing P values from separate analyses.[2]

Cancer Research UK Medical Statistics Group, Centre for Statistics in Medicine, Institute for Health Sciences, Oxford OX3 7LF

Douglas G Altman *professor of statistics in medicine*

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE

J Martin Bland *professor of medical statistics*

Correspondence to: D G Altman doug.altman@cancer.org.uk

BMJ 2003;326:219

1 Altman DG, Matthews JNS. Interaction 1: Heterogeneity of effects. *BMJ* 1996;313:486.
2 Matthews JNS, Altman DG. Interaction 2: Compare effect sizes not P values. *BMJ* 1996;313:808.
3 Matthews JNS, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ* 1996;313:862.
4 Torgerson DJ, Bell-Syer SEM. Hormone replacement therapy and prevention of nonvertebral fractures. A meta-analysis of randomized trials. *JAMA* 2001;285:2891-7.
5 Bland JM, Altman DG. Logarithms. *BMJ* 1996;312:700.
6 Bland M, Peacock J. Interpreting statistics with confidence. *Obstetrician and Gynaecologist* (in press).

Calculations for comparing two estimated relative risks

		Group 1	Group 2
1	RR	0.67	0.88
2	*log RR	−0.4005 (E_1)	−0.1278 (E_2)
3	95% CI for RR	0.46 to 0.98	0.71 to 1.08
4	*95% CI for log RR	−0.7765 to −0.0202	−0.3425 to 0.0770
5	Width of CI	0.7563	0.4195
6	SE[=width/(2×1.96)]	0.1929	0.1070
Difference between log relative risks			
7	$d[=E_1-E_2]$	−0.4005−(−0.1278)=−0.2726	
8	SE(d)	$\sqrt{(0.1929^2+0.1070^2)}$=0.2206	
9	CI(d)	−0.2726 ±1.96×0.2206 or −0.7050 to 0.1598	
10	Test of interaction	z=−0.2726/0.2206=−1.24 (P=0.2)	
Ratio of relative risks (RRR)			
11	RRR=exp(d)	exp(−0.2726)=0.76	
12	CI(RRR)	exp(−0.7050) to exp(0.1598), or 0.49 to 1.17	

*Values obtained by taking natural logarithms of values on preceding row.

BMJ: first published as 10.1136/bmj.326.7382.219 on 25 January 2003. Downloaded from http://www.bmj.com/ on 10 December 2023 by guest. Protected by copyright.