

The relation between treatment benefit and underlying risk in meta-analysis

Stephen J Sharp, Simon G Thompson, Douglas G Altman

In meta-analyses of clinical trials comparing a treated group with a control group it has been common to ask whether the treatment benefit varies according to the underlying risk of the patients in the different trials, with the hope of defining which patients would benefit most and which least from medical interventions. The usual analysis used to investigate this issue, however, which uses the observed proportions of events in the control groups of the trials as a measure of the underlying risk, is flawed and produces seriously misleading results. This arises through a bias due to regression to the mean and will be particularly acute in meta-analyses which include some small trials or in which the variability in the true underlying risks across trials is small. Approaches which previously have been thought to be more appropriate are to substitute the average proportion of events in the control and treated groups as the measure of underlying risk or to plot the proportion of events in the treated group against that in the control group (L'Abbé plot). However, these are still subject to bias in most circumstances. Because of the potentially seriously flawed conclusions that can result from such analyses, they should be replaced either by statistically appropriate (but more complex) approaches or, preferably, by analyses which investigate the dependence of the treatment effect on measured baseline characteristics of the patients in each trial.

Where there are substantial clinical differences between the different trials of a meta-analysis and their patients, or substantial quantitative differences in the results from the different trials, a single overall summary estimate of treatment benefit has little practical applicability.¹ An analysis which ignores this heterogeneity is clinically misleading and scientifically naive.² Many authors have now emphasised the clinical and scientific importance of investigating potential sources of heterogeneity when conducting a meta-analysis.³⁻⁵

If the results of a meta-analysis of clinical trials are to affect future clinical practice the clinician needs to know how the expected net treatment benefit varies according to certain measurable characteristics of a patient. A frequently investigated question is whether there is variation in the treatment benefit according to a patient's underlying risk of the event that the treatment is designed to prevent or delay. Underlying risk is used as a convenient summary of a number of characteristics which may be measurable risk factors but for which individual patient data are not available from some or all of the trials. Whereas it is often expected that the absolute risk reduction attributable to treatment will vary, possibly almost proportionately, with the underlying risk of patients in each trial, it is usually assumed that the relative risk (or odds ratio) does not vary in this way. Indeed, a recently proposed model to assess net benefit

is based on this assumption.⁶ The existence of a relation between relative treatment benefit and underlying risk would have crucial implications for the interpretation of the results of a meta-analysis, in terms of both assessment of net treatment benefit and economic considerations.⁷

We address two issues. Firstly, while the intention to investigate possible variation in treatment benefit is laudable, we show that common approaches to this problem contain serious statistical pitfalls. Secondly, we argue that these same approaches fail to indicate the potential treatment benefit for individual patients, which is of more practical use to clinicians, and advocate instead the direct use of measurable patient characteristics.

The data

To demonstrate the statistical pitfalls in an analysis of underlying risk as a source of heterogeneity we use data from a meta-analysis of randomised trials to assess the effectiveness of endoscopic sclerotherapy in reducing mortality in patients with cirrhosis and oesophago-gastric varices.⁸ Nineteen such trials comparing sclerotherapy with a control were reviewed; the data relevant to our discussion are shown in table 1. In the following, we use proportion of deaths (expressed as a percentage) on a log odds scale as the measure of "underlying risk" and the log odds ratio as the measure of treatment effect. However, the principles apply more generally to other measures of treatment effect, such as the risk ratio or mean difference for a quantitative outcome, and corresponding measures of "underlying risk." There was substantial evidence of heterogeneity ($P < 0.0001$) in the observed odds ratios across the 19 sclerotherapy trials, and therefore an investigation of whether underlying risk was part of the explanation for different observed treatment effects is scientifically relevant.²

Relating treatment effect to underlying risk: three conventional approaches

(1) GRAPH OF TREATMENT EFFECT AGAINST PROPORTION OF EVENTS IN CONTROL GROUP

A natural measure of underlying risk in a trial population is the observed proportion of events in the control group. Figure 1 shows a graph of odds ratio of death (log scale) against proportion of deaths in the control group (log odds scale) for the data from the sclerotherapy trials. Each trial on the graph is represented by a circle, the area of which indicates the study size. The graph includes the line of predicted values obtained from a weighted regression. The estimated slope is -0.61 (95% confidence interval -0.99 to -0.23), giving strong evidence of a negative association—that is, an increase in treatment benefit (lower odds ratio) with increasing proportion of events in the control group. The conclusion from this analysis would be that underlying risk is a significant source of heterogeneity. Furthermore, there is a temptation to use point T in the figure to define a cut off value of risk in

Medical Statistics Unit,
London School of Hygiene
and Tropical Medicine,
London WC1E 7HT
Stephen J Sharp, *research
fellow in medical statistics*

Department of Medical
Statistics and Evaluation,
Royal Postgraduate
Medical School, London
W12 0NN
Simon G Thompson,
*professor of medical statistics
and epidemiology*

Imperial Cancer Research
Fund Medical Statistics
Group, Centre for
Statistics in Medicine,
Institute of Health
Sciences, PO Box 777,
Oxford OX3 7LF
Douglas G Altman, *head*

Correspondence to:
Mr Sharp.

BMJ 1996;313:735-8

Table 1—Mortality results from 19 trials of sclerotherapy taken from a published meta-analysis^a

Trial*	Control group		Treated group		Odds ratio
	Deaths/total	%	Deaths/total	%	
1	14/36	38.9	2/35	5.7	0.10
2	2/20	10.0	0/21	0	0.17†
3	29/53	54.7	12/56	21.4	0.23
4	19/41	46.3	10/41	24.4	0.37
5	27/60	45.0	15/53	28.3	0.48
6	26/69	37.7	16/71	22.5	0.48
7	4/16	25.0	2/13	15.4	0.55
8	6/22	27.3	4/23	17.4	0.56
9	34/46	73.9	30/49	61.2	0.56
10	24/51	47.1	19/55	34.5	0.59
11	18/41	43.9	18/42	42.9	0.96
12	8/28	28.6	6/21	28.9	1.00
13	21/35	60.0	20/33	60.6	1.03
14	14/60	23.3	13/53	24.5	1.07
15	5/24	20.8	5/22	22.7	1.12
16	6/18	33.3	6/16	37.5	1.20
17	14/72	19.4	18/73	24.7	1.36
18	6/19	31.6	7/18	38.9	1.38
19	23/138	16.7	46/143	32.2	2.37

*Original reference to trials given in Pagliaro *et al.*⁸

†To avoid degeneracy, odds ratio is estimated by $(0.5/21.5)/(2.5/18.5)$.²⁹

the control group and conclude that the treatment is effective (odds ratio below 1) only in patients with an underlying risk higher than this value.

The problem with this interpretation is due to regression towards the mean.^{9, 10} Because the outcome in the control group is being related to the treatment effect, an expression that also includes the control group outcome, a relation is expected. In the case where the treatment reduces the risk a high observed proportion of events in the control group will tend to lead to a larger observed treatment effect—and the converse when the observed proportion is low. In other words, the bias will lead to the potentially incorrect inference that the treatment is most beneficial among high risk patients and least among low risk patients. The size of the bias can be surprisingly large. In the extreme case when the treated and control group outcomes are unrelated the expected correlation can be -0.71 .¹¹ Underlying risk may indeed be a source of heterogeneity, but such a graph and regression will misrepresent any true effect.

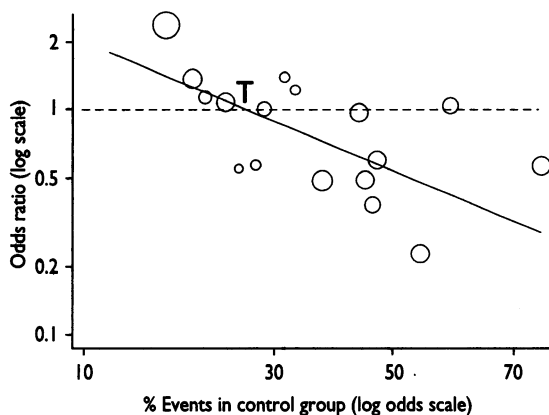


Fig 1—Treatment effect versus percentage of events in control group for sclerotherapy trials. The area of each circle is inversely proportional to the variance of the estimated treatment effect in the trial

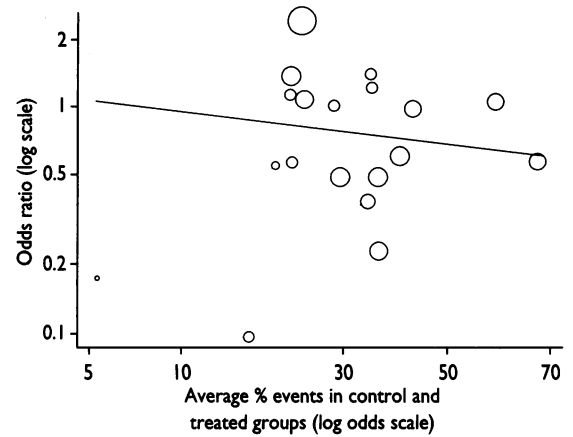


Fig 2—Treatment effect versus average percentage events in control and treated groups for sclerotherapy trials

(2) GRAPH OF TREATMENT EFFECT AGAINST AVERAGE PROPORTION OF EVENTS IN THE CONTROL AND TREATED GROUPS

Figure 2 is a graph of odds ratio of death (log scale) against the average proportion of events in the control and treated groups (log odds scale) for the data from the sclerotherapy trials; there is a slight increase in treatment effect (reduction in odds ratio) as the average proportion increases, but the evidence is unconvincing: the slope of the fitted line is -0.16 (-0.73 to 0.42). Use of the average has led to a different conclusion: underlying risk is not a significant source of heterogeneity.

Unfortunately the validity of the approach using the average relies on the assumption that the true treatment effect does not vary between trials¹²; departures from this assumption will lead to bias in the size and direction of any observed association. To take an extreme example, consider a set of very large trials (where errors of measurement are negligible) which have the same underlying risk (as measured by proportion of events in the control group) but some of which have larger treatment benefits than others. A graph of treatment effect against average proportion will show a positive relation, whereas in truth there is no relation with underlying risk. Furthermore, there is also a conceptual difficulty with using a method which depends on an assumption of no variation in true treatment effect to estimate how the treatment effect varies with underlying risk.

To illustrate further the danger of a simple analysis using some observed measure of "risk," we consider briefly an approach using a graph of treatment effect against proportion of events in the treated group. The data from the sclerotherapy trials produce a slope of the fitted line of 0.51 (0.02 to 1.00), leading to the opposite conclusion from that obtained using the proportion of events in the control group. In other words, the danger is that any observed association between treatment effect and underlying risk is strongly dependent on the choice of measure of underlying risk, even to the extent of determining the direction of the association.

(3) L'ABBÉ PLOT: PROPORTION OF EVENTS IN THE TREATED GROUP AGAINST PROPORTION OF EVENTS IN THE CONTROL GROUP

The L'Abbé plot of proportion of events in the treated group against that in the control group was proposed as a graphical means of exploring possible heterogeneity.¹³ If the trials are fairly homogeneous the points would lie around a line corresponding to the pooled treatment effect parallel to the line of identity: large deviations would indicate possible heterogeneity.

Figure 3 is a L'Abbé plot for the data from the sclerotherapy trials, including the line of identity and weighted least squares regression line, estimated slope

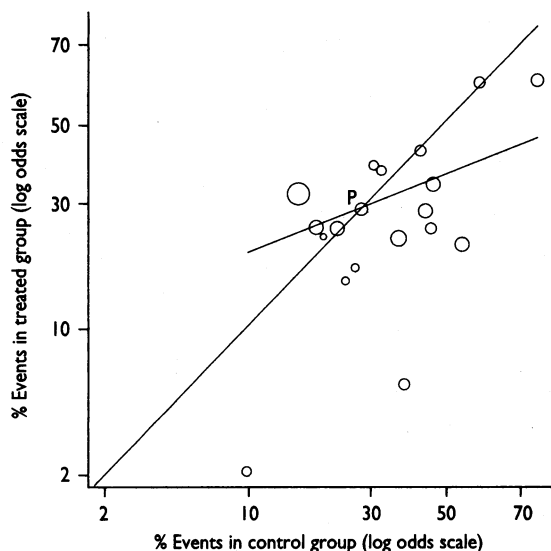


Fig 3—L'Abbé plot for sclerotherapy trials. Log scales used because treatment effects are estimated odds ratios, not risk differences

0.38 (0.07 to 0.69). This illustrates how L'Abbé plots may be used to draw two inappropriate conclusions. Firstly, treatment effect is associated with underlying risk due to the slope of the regression line being different from the slope of the line of identity, and, secondly, point P defines a cut off value of proportion of events in the control group, above which the treatment is effective (odds ratio below 1) and below which it is not (odds ratio above 1), analogous to point T in figure 1. These interpretations are misleading because the proportion of events in the control group is measured with error, leading to underestimation of the true regression slope,¹⁴ another manifestation of regression towards the mean.^{9 10} L'Abbé plots may be useful as an exploratory graphical device—for example, to identify an unusual or outlying trial—but they cannot be used in conjunction with a regression analysis to define regions in which a treatment is or is not effective.

Interpretation of conventional graphs: practical guidelines

The extent to which these conventional approaches yield misleading conclusions depends on a number of factors, and below we summarise guidelines based on algebraic results (see appendix) for the interpretation of such plots.

Use of proportion of events in the control group (1):

- is not an appropriate method, and will always be biased,
- will be less misleading—that is, less biased—if the trials are mostly large, or the variation in true underlying risks is large.

Use of average proportion of events in control and treated groups (2):

- is appropriate only if the true treatment effect is constant across trials,
- will be less misleading if the variation in true underlying risks is large.

A L'Abbé plot (3):

- is a useful exploratory graphical method as an adjunct to a standard meta-analysis plot,
- is not appropriate for defining groups in which treatment is or is not effective.

In all of these approaches it is the use of ordinary linear regression to interpret the graph which leads to misleading conclusions.

A clinically more useful alternative

Given that a patient's "underlying risk" is known only to the clinician through certain measured characteristics, a clinically more useful alternative to the problematic analyses we have described is to relate treatment benefit to measurable baseline characteristics. These characteristics, or some combination of them, would act as a surrogate measure of the patient's risk. In a meta-analysis such an analysis would ideally be based on individual patient data, but it would also be possible to use group data. To take a simple example, we might use age either for each patient or, much more crudely, the mean age of all patients in a trial. Such an analysis would allow age specific estimates of the treatment benefit to be derived.

An extension of this idea would be to combine several prognostic variables into a risk score.¹⁵ The relation of treatment benefit to risk score for individual patients could be evaluated using a regression analysis.⁵ Such a combination would avoid the problem of post hoc dredging which arises when many variables are considered separately and would best be based on data from sources other than the trials which form the meta-analysis for treatment effects, such as prospective studies. Rothwell has discussed such prognostic modelling in the context of a single clinical trial, where the aim is to identify those patients most likely to benefit from treatment.¹⁶

Applying this approach to meta-analysis would avoid the biases inherent in the "underlying risk" analyses. Moreover, the benefit from treatment, for a particular patient, could then be estimated on the basis of specific measurable patient characteristics, rather than the unknown quantity "underlying risk."

Discussion

The question of whether treatment effects are related to underlying risk has been considered in several conditions—for example, in trials of cholesterol reduction and mortality,¹⁷ of tocolysis using β mimetics in preterm delivery,¹⁸ and of antiarrhythmic drugs after acute myocardial infarction.^{19 20} Differences in the underlying risk of patients have also been proposed as an explanation for the differences between the results of meta-analysis and a subsequent megatrial of magnesium therapy after acute myocardial infarction. Unfortunately, while the question posed is simple, it turns out that there is no easy way to answer it validly. A similar problem occurs in several guises in medicine, perhaps most commonly the issue of the possible relation between change and initial value.^{10 12 21}

As we have shown, the conventional approaches to answering this question are flawed and liable to produce misleading results, yet there are several examples of such methods in the medical literature. The problem of using the observed proportion of events in the control group was first discussed¹¹ in response to a meta-analysis of 14 placebo controlled clinical trials to evaluate the effect of tocolysis with β mimetics on the risk of preterm birth.¹⁸ An example where the outcome was a continuous measure appeared in a review of 18 randomised controlled trials of prophylactic desmopressin to reduce perioperative blood loss during cardiac surgery,²² where one conclusion was that the efficacy of desmopressin was a function of blood loss in the control group. In a meta-analysis of 17 controlled clinical trials to test whether dietary supplementation with fish oil (ω -3 polyunsaturated fatty acids) reduces systolic blood pressure²³ the conclusion that the magnitude of blood pressure reduction increased progressively with the level of blood pressure must also be in doubt, since the

Key messages

- Variations in the effect of treatment found in a meta-analysis might be explained by differences in the patients' underlying risk of adverse outcome
- Such an association would have important implications in the evaluation of the treatment
- Conventional analysis methods are based on the observed proportion of events in the control group or the average proportion in the control and treated groups
- Such methods are flawed in most situations and can lead to seriously misleading conclusions
- A preferable alternative approach is to relate treatment benefit to *measurable* patient characteristics

level used was the average of the levels at pretreatment and at the end of treatment. Finally, exactly the misleading interpretation of a L'Abbé plot described earlier was given after a meta-analysis of drug trials in mild hypertension^{24, 25}: deaths per 1000 person years in the intervention group were plotted against the same quantity in the control group, and the point of intersection of the regression line and line of identity was used to estimate an underlying risk of mortality below which drugs were harmful and above which they were beneficial.

The biases described will not apply equally to all meta-analyses. As we have noted, the effect will be particularly acute in meta-analyses which include some small trials (a common case) or in which the true variability in underlying risk across trials is small. Although we have concentrated on binary outcomes, the same issues apply to all types of outcome. An approach based on the average may work for other problems,²⁶ but the assumptions made¹² are untenable in the context of meta-analysis, where each estimate has a different precision, being from a different trial. While the use of the average proportion of events is likely to be less biased than the use of the proportion of events in the control group, it would be clearly preferable to use an unbiased method. A statistically valid approach requires a correction of the observed relation between treatment effect and proportion of events in the control group to allow for the bias. A solution to this problem is difficult because each trial result has a different precision. Complex solutions have been proposed,^{27, 28} but based on strong assumptions: we do not consider them further here. Even were a correct solution to this question available, it is unclear how this information could be used by a clinician, as there is no direct way of assessing the risk for an individual patient.

In conclusion, the statistical and philosophical difficulties with the conventional simple approaches (except in the extreme situation where all the trials in a meta-analysis are very large) lead us to recommend that these should be avoided. Attention instead should be given to the more practically applicable question of how measurable patient characteristics impact on treatment efficacy.

Funding: Medical Research Council (SJS), University of London (SGT), Imperial Cancer Research Fund (DGA).

Conflict of interest: None.

Appendix

Let x_j and y_j be the observed measures of "risk" in the control and treated groups of the j th trial—these could be log odds, absolute risks, log risks, or means of a quantitative outcome, whichever is most appropriate.

Write

$$\begin{aligned}x_j &= \mu_j + \epsilon_{1j} \\y_j &= \mu_j + \delta_j + \epsilon_{2j}\end{aligned}$$

where μ_j is the underlying risk in trial j , δ_j the true treatment effect, ϵ_{1j} and ϵ_{2j} random errors of measurement, and for all j

$$\begin{aligned}\text{Var}(\mu_j) &= \sigma_\mu^2 & \text{Var}(\delta_j) &= \sigma_\delta^2 \\E(\epsilon_{1j}) &= E(\epsilon_{2j}) = 0, & \text{Var}(\epsilon_{1j}) &= \text{Var}(\epsilon_{2j}) = \sigma^2\end{aligned}$$

Then σ_μ^2 represents the variability in the true underlying risks, and σ_δ^2 the variability in the true treatment effects between trials; σ^2 is the variability of the measured outcomes *within* each trial group, here assumed the same in each trial.

The observed treatment effect—for example, log odds ratio, absolute risk difference, log relative risk, mean difference—is

$$t_j = y_j - x_j$$

Even when there is *no* relation between the true treatment effects δ_j and underlying risks μ_j across trials, it can be shown that:

$$\text{Expected slope of regression of } t_j \text{ on } x_j = \frac{-\sigma^2}{\sigma_\mu^2 + \sigma^2}$$

Expected slope of regression of t_j on

$$\frac{x_j + y_j}{2} = \frac{2\sigma_\delta^2}{2\sigma^2 + \sigma_\delta^2 + 4\sigma_\mu^2}$$

The former is necessarily negative, and the latter positive unless $\sigma_\delta^2 = 0$

- Horowitz R. Large scale randomised evidence, large simple trials and overviews of trials: discussion—a clinician's perspective on meta-analysis. *J Clin Epidemiol* 1995;48:41-4.
- Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994;309:1351-5.
- Chalmers T. Problems induced by meta-analysis. *Statistics in Medicine* 1991;10:971-80.
- Victor N. The challenge of meta-analysis: discussion. *J Clin Epidemiol* 1995;48:5-8.
- Berkey C, Hoaglin D, Mosteller F, Colditz G. A random effects regression model for meta-analysis. *Statistics in Medicine* 1995;14:395-411.
- Glasziou P, Irwig L. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356-9.
- Davey Smith G, Egger M. Who benefits from medical interventions? *BMJ* 1994;308:72-4.
- Pagliari L, D'Amico G, Sorensen T, Lebrech D, Burroughs AK, Morabito A, et al. Prevention of first bleeding in cirrhosis - a meta-analysis of randomized trials of nonsurgical treatment. *Ann Intern Med* 1992;117:59-70.
- Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994; 308:1499.
- Bland JM, Altman DG. Some examples of regression to the mean. *BMJ* 1994;309:780.
- Senn S. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine* 1994;13:293-6.
- Hayes R. Methods for assessing whether change depends on initial value. *Statistics in Medicine* 1988;7:915-27.
- L'Abbé K, Detsky A, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987;107:224-33.
- Armitage P, Berry G. *Statistical methods in medical research*. 3rd ed. Oxford: Blackwell, 1994:288-90.
- Yusuf A, Zucker D, Peduzzi P, Fisher L, Takaro T, Kennedy J, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet* 1994;344:563-70.
- Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995;345:1616-9.
- Davey Smith G, Song F, Sheldon T. Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ* 1993;206:1367-73.
- Brand R, Kragt H. Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine* 1992;11:2077-82.
- Boissel J, Collet J, Lieve M, Girard P. An effect model for the assessment of drug benefit: example of antiarrhythmic drugs in postmyocardial infarction patients. *J Cardiovasc Pharmacol* 1993;22:356-63.
- Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240-8.
- Blomqvist N. On the relation between change and initial value. *Journal of the American Statistical Association* 1977;72:746-9.
- Cattaneo M, Mannucci P. Desmopressin and blood loss after cardiac surgery. *Lancet* 1993;342:812.
- Appel L, Miller E, Seidler A, Whelton P. Does supplementation of diet with "fish oil" reduce blood pressure? A meta-analysis of controlled clinical trials. *Arch Intern Med* 1993;153:1429-38.
- Hoes AW, Grobbee DE, Lubsen J. Does drug treatment improve survival? Reconciling the trials in mild-to-moderate hypertension. *J Hypertension* 1995;13:805-11.
- Egger M, Davey Smith G. Risks and benefits of treating mild hypertension: a misleading meta-analysis? *J Hypertension* 1995;13:813-5.
- Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; 346:1085-7.
- Van Houwelingen H, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine* 1993;12:2273-84.
- McIntosh M. The population risk as an explanatory variable in research synthesis of clinical trials. *Harvard Medical School Technical Report* 1995;4:1-35.
- Cox DR, Snell EJ. *Analysis of binary data*. 2nd ed. London: Chapman and Hall, 1989.

(Accepted 21 June 1996)