STATISTICS IN MEDICINE
Statist. Med. 2000; **19**:3127–3131

A simple method for converting an odds ratio to effect size for use in meta-analysis

Susan Chinn*,†

Department of Public Health Sciences, King's College, London, 5th floor, Capital House, 42 Weston Street, London SE1 3QD, U.K.

SUMMARY

A systematic review may encompass both odds ratios and mean differences in continuous outcomes. A separate meta-analysis of each type of outcome results in loss of information and may be misleading. It is shown that a ln(odds ratio) can be converted to effect size by dividing by 1.81. The validity of effect size, the estimate of interest divided by the residual standard deviation, depends on comparable variation across studies. If researchers routinely report residual standard deviation, any subsequent review can combine both odds ratios and effect sizes in a single meta-analysis when this is justified. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

Meta-analysis is now used extensively in reviews of randomized controlled trials and observational studies, but the problems include how to compare non-identical outcomes [1]. A review of guidelines for systematic reviews of randomized trials recommended 'identification of a common set of definitions of outcome'.

One form of lack of common definition is the use of a continuous outcome by some authors, and a dichotomous outcome by others. This can lead to reviewers performing two separate meta-analyses [2, 3]. This problem has already been tackled in a systematic review of the prophylactic use of oxytocics on postpartum blood loss in the third stage of labour [4]. The authors considered that postpartum haemorrhage, defined as the loss of 500 ml or more, was the outcome of interest, but some trials summarized blood loss by mean and standard deviation. A method of estimating the log-odds ratio for the latter trials was presented, under the assumption of a Normal distribution.

However in many situations there is no natural dichotomy, and reviewers will wish to retain the greater power generally provided by continuous outcome measures. Continuous outcome measures on different scales can be combined using 'effect size', the estimate of interest, which may be a difference in means or a regression coefficient, divided by the residual standard

* Correspondence to: Susan Chinn, Department of Public Health Sciences, King's College, London, 5th floor, Capital House, 42 Weston Street, London SE1 3QD, U.K.
† E-mail: sue.chinn@kcl.ac.uk

Received February 1999
Accepted June 2000

deviation [1]. The simple approximate method for converting an odds ratio to effect size presented here enables reviewers to maximize the information available.

2. THE EQUIVALENCE OF ODDS RATIOS AND EFFECT SIZE

Logistic regression with results reported as odds ratios, or a close equivalent, is unavoidable for truly dichotomous outcome variables. However some continuous outcomes, such as blood pressure [5], are frequently dichotomized. The odds of a subject being hypertensive depend on the mean and variation of the underlying distribution of blood pressure. The odds ratio for one risk factor group compared to another is invariant to choice of cut-off point if the logit of the proportion with hypertension plotted against blood pressure is parallel for the two risk factor groups. It is a sufficient, but not necessary, condition for the underlying distribution of blood pressure in the two groups to be logistic with equal variances.

Most analyses of continuous outcomes proceed on the assumption that the distribution is Normal, often after transformation of the data. However it is known that the logistic and Normal distributions differ little, except in the tails of the distributions [6]. This is illustrated in Figure 1, where the logit of a proportion is plotted against the Normal equivalent deviate (NED). The standard logistic distribution [7] has variance $\pi^2/3$, so a difference in ln(odds) can be converted to an approximate difference in NED by dividing by $\pi/\sqrt{3}$, which is 1.81 to 2 decimal places. As a difference in NED is the effect size [1], a meta-analysis of ln(odds) is equivalent, albeit with loss of power, to one of effect size except for the scaling factor of 1.81.

In order to convert a difference in NED to a difference on the underlying scale, a standard deviation is required. Hence conversion of an odds ratio to an absolute difference is possible if the standard deviation is known. It follows from this that studies reporting odds ratios are truly comparable in absolute terms if and only if the underlying standard deviation is the same for each study.

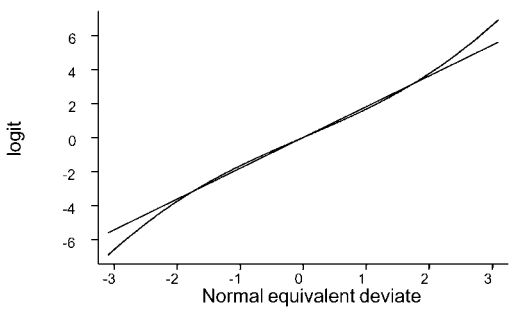

Figure 1. The relation between the logit of a proportion and the Normal equivalent deviate (curved line), and the fit of a Normal distribution with variance equal to that of the logistic distribution (straight line, gradient 1.81).

Table I. Published odds ratios for improvement in asthma symptoms following house dust mite control measures, published standardized mean difference in symptom scores [2] and re-analysis.

Study	Total *n*	Odds ratio	95 per cent confidence interval	Utilized effect size*	95 per cent confidence interval*	Utilized SE
1	30	1.32	0.31 to 5.68	0.153		0.410
2	53	1.27	0.43 to 3.76	0.132		0.306
3	42	5.33	1.07 to 26.5	0.924		0.452
4	56	0.58	0.18 to 1.89	−0.301		0.331
5	49	0.83	0.17 to 4.05	−0.103		0.447
6	23			−0.237	−1.059 to 0.585	0.395
7	46			−0.552	−1.141 to 0.038	0.292
8	52			0.387	−0.162 to 0.936	0.273
9	18			0.288	−0.632 to 1.219	0.436
10	28			−0.032	−0.773 to 0.709	0.360
11	24			1.433	0.518 to 2.348	0.691
12	35			−0.541	−1.217 to 0.135	0.331
Pooled: published		1.20	0.66 to 2.18	−0.064†	−0.408 to 0.536†	
Pooled: re-analysis		1.23	0.62 to 2.40	−0.070†	−0.402 to 0.542†	
Pooled: all studies				0.087	−0.222 to 0.395	

* Sign reversed from that published for studies 6 to 12.
† Studies 6 to 12 only.

3. AN EXAMPLE

The example is a review of house dust mite control measures in the management of asthma [2]. Twelve studies were included, of which five reported the odds ratio for an improvement in symptoms in the treated group relative to the control group, and seven reported a standardized mean difference in symptoms, that is, effect size. The estimate and associated 95 per cent confidence interval, and number of subjects, were reported for each individual study. In order to convert the odds ratios to effect size each odds ratio and associated confidence interval was ln-transformed, and the standard error calculated as the width of the confidence interval divided by 2×1.96. Each ln(odds ratio) and associated standard error were then converted to effect size and its standard error by dividing by 1.81. The standard error of each reported effect size was calculated from the width of the confidence interval divided by $2 \times t_{df,0.05}$. The two separate meta-analyses in the paper were repeated, as different software was used, and then one estimate of effect size was combined in a single random effects meta-analysis. Stata 5.0 was used [8], which provides the moment estimator of DerSimonian and Laird [9], with a random effects analysis as the authors reported using this if heterogeneity was detected [2].

The results are shown in Table I. As a positive effect size represented more symptoms in the treated group, the effect sizes for studies 6 to 12 have been reversed in sign for the single analysis so that they were comparable with the effect size derived from the odds ratio for improvement in the treated compared to the control group. Neither of the published separate meta-analyses gave much support to an effect of house dust mite control measures, but confidence intervals were wide. The re-analysis of all 12 studies together provides a narrower confidence interval than the seven study effect size analysis, and confirms the conclusion.

4. DISCUSSION

This paper does not advocate the use of effect size. If all outcomes are continuous on the same scale then that is how they should be analysed. Greenland has warned against the use of standardized regression coefficients in meta-analysis [10, 11], and although it appears that it is standardizing the explanatory variable that is at the root of most of the problems he reports, caution should also be exercised over effect size. However, in some cases reviewers have no alternative.

It has been shown here that a meta-analysis of odds ratios is equivalent to a meta-analysis of effect size when there is an underlying continuous distribution, albeit with some loss of power. This also lends some justification to the combination of odds ratios from studies with different outcome variables, or from studies using different cut-off points of a continuous distribution. If combining effect size is justified, then meta-analysis of odds ratios is also warranted. From the viewpoint of Greenland's criticism of effect size this can be reversed; if a meta-analysis of effect size is rejected then so should one of odds ratios even if the exposure variables are all on the same scale.

While neither effect size nor odds ratio is ideal when the outcome is truly continuous, use of two separate meta-analyses of dichotomous and continuous outcomes can lead to a number of problems. First, neither will have as much power as the combined analysis, and an erroneous conclusion may be reached, a problem also identified by Whitehead *et al.* [4]. Secondly, results of the two analyses could conflict. Thirdly, information from one study may be used in both analyses, so the seemingly confirmatory results may be little more than a repetition in disguise. Combining the two forces the reviewer to choose one of the estimates from each study, that of direct effect size to be preferred on grounds of power over effect size derived from an odds ratio. The estimates from dichotomous and binary outcomes, but not the standard errors, will be of comparable size. This was noted by Whitehead *et al.* for studies that report both when Normal distributions with common variance are assumed [4] and found when the method here presented was applied to the estimates in the study of Hazell *et al.* [3]. Fourthly, the studies reporting odds ratios and a continuous outcome may differ in size, and examination of funnel plots [12] may erroneously conclude that there is publication bias, or such bias may be undetected because of the reduced number of studies in each analysis. Fifthly, heterogeneity may be undetected, or possibly erroneously reported. Possible causes of heterogeneity should be investigated [13] and the more studies that are included the more feasible this becomes.

In the example it was possible to combine estimates only because effect sizes were reported directly. Unless effect size, or a residual standard deviation allowing effect size to be calculated, is published for each eligible study, reviewers will be unable to carry out a single meta-analysis, and the above problems will continue. The method of Whitehead *et al.* also requires an estimate of standard deviation from each study [4].

Both methods assume an underlying Normal distribution, although Whitehead *et al*. also provide for estimation of ln(odds ratio) if the distribution is assumed log-Normal. Effect size assumes a common variance, and Whitehead *et al*. recommend using this assumption when plausible. The method given here for dichotomous and continuous outcomes is complementary to and simpler than that of Whitehead *et al*., and should be more powerful in that all the information is retained. The factor of 1.81 for converting ln(odds ratio) to effect size is an approximation, but a good one over the likely range of use.

No researcher should assume that results from a single study will be accepted without attempts at replication. To meet Greenland's objections [10, 11], provided the continuous outcome variables are on the same scale, a subanalysis of these alone should also be presented on the scale of measurement. If the residual standard deviations are very different then this is an indication of heterogeneity that needs to be investigated as much as that in the estimates themselves. The recommendation to report residual standard deviation needs to be incorporated in guidelines for publication by medical journals with an explanation of why this is required.

ACKNOWLEDGEMENTS

The author thanks an anonymous referee for helpful comments.

REFERENCES

1. Hasselblad V, Mosteller F, Littenberg B, Chalmers TC, Hunink MG, Turner JA, Morton SC, Diehr P, Wong JB, Powe NR. A survey of current problems in meta-analysis. *Medical Care* 1995; **33**(2):202–220.
2. Gøtsche PC, Hammarquist C, Burr M. House dust mite control measures in the management of asthma: meta-analysis. *British Medical Journal* 1998; **317**(7166):1105–1110.
3. Hazell P, O'Connell DO, Heathcote D, Robertson J, Henry D. Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *British Medical Journal* 1995; **310**(6984):897–901.
4. Whitehead A, Bailey A, Elbourne D. Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. *Journal of Biopharmaceutical Statistics* 1999; **9**(1):1–16.
5. Singh JP, Larson MG, Tsuji H, Evans, JC, O'Donnell, CJ, Levy D. Reduced heart rate variability and new-onset hypertension: insights into pathogenesis of hypertension: the Framingham Heart Study. *Hypertension* 1998; **32**(2): 293–297.
6. Finney DJ. *Statistical Method in Biological Assay*, *3*rd edn. Charles Griffin: London and High Wycombe, 1978; 362–368.
7. Kendal MG, Stuart A. *The Advanced Theory of Statistics Volume I.* 4th edn. Charles Griffin: London and High Wycombe, 1977; 126.
8. Stata Corporation. *Stata Statistical Software Release 5*. Stata Press, Texas, 1997.
9. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(3):177–188.
10. Greenland S, Schlesselman JJ, Criqui, MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology* 1986; **123**(2):203–208.
11. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* 1987; **9**:1–30.
12. Sutton AJ, Jones DR, Abrams K, Sheldon TA, Song F. Systematic reviews of randomised trials. In *Health Services Research Methods*, Black N, Brazier J, Fitzpatrick R, Reeves B (eds). BMJ Books: London, 1998; 175–186.
13. Thompson, S. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; **2**(2):173–192.

