An adjusted random-effects model for binary-data meta-analysis

Rose Baker
School of Business
University of Salford, UK

Keywords

meta-analysis, random effect, binary data, beta distribution.

An adjusted random-effects model for binary-data meta-analysis

Rose Baker
School of Business
University of Salford, UK

Abstract

A new model with a variable size of random effect is introduced for the meta-analysis of 2×2 tables. The random-effects parameter has a simple interpretation in terms of sample size and offers a new measure of heterogeneity.

1. Introduction

The focus of this article is binary-data meta-analysis applied in medicine and epidemiology, and the language of those subjects is used. However, the results here apply also to meta-analysis generally. Beyond medicine and epidemiology, there is a very wide range of application areas, e.g. education.

By far the most common trial design for binary data is a parallel study, in which one patient group receives treatment A, and the other receives treatment B. Often but not always, 'treatment A' will be a placebo. An event (some measure of recovery or the reverse, such as death) occurs to some members of each group. There are usually no individual patient covariates such as age or disease duration, so the results can be summarized in a 2×2 table. In epidemiological studies, typically group A is the control group,

and group B has been exposed to some hazard.

In a meta-analysis we seek to estimate the 'treatment effect' θ and its standard error from a number of such 2×2 tables, each one giving the results of a study that a systematic review has found to be of acceptable quality. One major problem is that studies often disagree by more than their quoted statistical errors would indicate. This disagreement may arise in medicine because of differing patient mixes among the various studies, varying operational procedures, or use of a wrong model of treatment effect by the analyst. This problem appears in nearly half of binary-data meta-analyses and is even more prevalent for continuous outcomes (Alba *et al*, 2016); the random effect is larger for 'softer' outcomes and lowest for 'hard' outcomes such as mortality (Turner *et al*, 2012). The Higgins *et al* (2003) I^2 statistic is often used to give a measure of this extra variability (for caveats, see Borenstein *et al* 2017).

A standard approach is to model the excess variability by assuming that the observable treatment effect varies from study to study, so that the ith study 'sees' $\theta_i = \theta + \epsilon_i$, where $\epsilon_i \sim N[0, \tau^2]$. In this paper a modified form for the random effect is introduced, in which rather than simply assuming that the effective value of θ varies from study to study, it is considered that this variation is induced by the probabilities of the event occurring under treatments A and B varying randomly between studies, so that the ith study has effective event probabilities that differ from the correct values, for the reasons given. Modelling this variation in probabilities using the beta distribution rather than modelling the variation in θ_i directly gives a slightly different form for the random effect, in which its scale is geared to

the variance that $\hat{\theta}_i$ would have for a constant sample size, say unity, in each group.

Note that if the treatment A and B event probabilities do not vary much from study to study, it does not matter which model of treatment effect is used, and also the random effects model proposed here and conventional models would give similar fits to data and give rise to similar conclusions. However, these probabilities usually do vary appreciably across studies.

The new random-effects submodel is derived, some examples are given, and the paper ends with some brief conclusions.

2. The new model

2.1. Some notation

Let there be n studies; table 1 gives notation for the observed data from a study. Study suffices will often be suppressed for clarity.

Let p be the probability of an event for treatment A (often control/placebo) and q the corresponding probability for treatment B. Let the treatment effect be $\theta = g(q, p)$. The currently-used treatment effects can all be written as $\theta = T(q) - T(p)$, for some monotonic function T, but this simplification is not needed here. Thus the widely-used log-odds ratio is

$$\theta = \ln(q/(1-q)) - \ln(p/(1-p)) = \ln(q(1-p)/p(1-q)). \qquad (1)$$

The methodology is exemplified using the log-odds ratio throughout, but is

quite general. Two-stage models give

$$\hat{\theta}_i = g(\hat{q}_i, \hat{p}_i) \tag{2}$$

from the ith study, where $\hat{\theta}_i$ is assumed to be approximately normally distributed with mean θ. Sample sizes are taken as N_q, N_p respectively.

2.2. Derivation of the model

The task is to find the variance of $\hat{\theta}$, which is assumed to be approximately normally distributed. Besides sampling error, the variance must include the random errors in P and Q; we use P to show the method, and results are similar for Q. For small changes $\delta p, \delta q$ we have that $\delta\theta \simeq (\partial g/\partial p)\delta p + (\partial g/\partial q)\delta q$, and assuming that $\delta p, \delta q$ are independent, the delta method (e.g. Oehlert, 1992) gives

$$\mathrm{var}(\hat{\theta}) \simeq \{(\partial g/\partial p)\}^2 \mathrm{var}(P) + \{(\partial g/\partial q)\}^2 \mathrm{var}(Q).$$

Taking N_p, N_q as fixed, assume that the effective probability of an event is a random variable from some distribution. The beta distribution, as the conjugate prior for the binomial distribution, is a natural choice here. With parameters α, β the mean is $p = \alpha/(\alpha + \beta)$. Then using the notation in table 1, n_{11} is the realization of a beta-binomial random variate (see e.g. Prentice, 1986), with mean $N_p\alpha/(\alpha + \beta)$, variance

$$\mathrm{var}(n_{11}) = \frac{N_p\alpha\beta(\alpha + \beta + N_p)}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \tag{3}$$

Reparameterizing, one parameter, the random effect size, is taken as $\rho = 1/(\alpha+\beta+1)$, and this is also the intra-study correlation. The method of moments gives $\hat{p} = n_{11}/N_p = \rho\hat{\alpha}/(1-\rho)$. Hence $\hat{\beta} = (1-\rho)(1-\hat{p})/\rho$, and so from (3) the estimated variance of $\hat{p}$ is

$$\text{var}(n_{11}/N_p) \simeq \hat{p}(1-\hat{p})\{1/N_p + (1-1/N_p)\rho\}. \tag{4}$$

Hence finally

$$\text{var}(\hat{\theta}) \simeq \{(\partial g/\partial p)\}^2\hat{p}(1-\hat{p})\{1/N_p+(1-1/N_p)\rho\}+\{(\partial g/\partial q)\}^2\hat{q}(1-\hat{q})\{1/N_q+(1-1/N_q)\rho\}, \tag{5}$$

which can be written for the ith study

$$\text{var}(\hat{\theta}_i) = \sigma_i^2 + v_i\rho.$$

The parameter ρ also has the meaning that $\alpha+\beta = (1-\rho)/\rho$ is the notional or effective sample size per group for the inter-study variation in θ_i. The practical application of this is that there is little point in new studies having much larger sample size than ρ^{-1}. The statistic $\hat{\rho}$, expressed as a percentage, is a new measure of heterogeneity, not related to statistical significance as is I^2. For example, as sample size per study increases, I^2 will tend to 100%, but $\hat{\rho}$ will remain constant. Borenstein *et al* (2017) stress that I^2 is not 'an absolute measure of heterogeneity' although it is often taken as such; $\hat{\rho}$ is an absolute measure.

This model has the implication that the size of random error for a study is geared to the sizes of $\hat{p}$ and $\hat{q}$; for values of $\hat{p}, \hat{q}$, giving a tiny variance,

the random effect variance is similarly tiny. Thus for the log-odds ratio $\sigma_i^2 = \frac{1}{\hat{q}(1-\hat{q})N_q} + \frac{1}{\hat{p}(1-\hat{p})N_p}$, so that the extra variance $\rho\{\frac{1}{\hat{q}(1-\hat{q})} + \frac{1}{\hat{p}(1-\hat{p})}\}$ is smallest when $\hat{p}, \hat{q} = 1/2$. This is the difference from the standard model, where the size of random error is constant.

Note that for the arc-sine transformation, designed to stabilize variance, $\theta = \sin^{-1}(\sqrt{q}) - \sin^{-1}(\sqrt{p})$, we have that $v_i = 1/2$, so the new method reduces to the old. For risk difference $\theta = q - p$ we have that $v_i = \hat{p}(1 - \hat{p}) + \hat{q}(1-\hat{q})$. The size of the random effect is now greatest where $\hat{p}, \hat{q} \simeq 1/2$. Also, if $N_q = N_p$ is constant for all trials, the new method amounts to a multiplicative rescaling of variance, like the methods used in particle physics (Baker and Jackson, 2013)

2.3. Changes to the DerSimonian and Laird and Mandel-Paule procedures

The DerSimonian and Laird (1986) method for random-effects meta-analysis has been very popular in the life sciences. Writing $w_i = 1/\sigma_i^2$, the sum of squares $Q = \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2/\sigma_i^2$ (where $\bar{\theta} = \sum_{i=1}^n w_i\hat{\theta}_i / \sum_{i=1}^n w_i$) is equated to its expectation, under the assumption that $\hat{\theta}_i$ has variance $\sigma_i^2 + \tau^2 = 1/w_i^*$. This yields the DSL estimator

$$\hat{\tau}^2 = \max\{0, \frac{Q - (n-1)}{\sum_{i=1}^n w_i - \sum_{i=1}^n w_i^2 / \sum_{i=1}^n w_i}\},$$

where now θ is estimated as $\sum_{i=1}^n w_i^*\hat{\theta}_i / \sum_{i=1}^n w_i^*$ with standard error $(\sum_{i=1}^n w_i^*)^{-1/2}$.

We follow the same derivation, under the assumption that $\hat{\theta}_i$ has variance $\sigma_i^2 + v_i\rho$, which yields the estimator

$$\hat{\rho} = \max\{0, \frac{Q - (n-1)}{\sum_{i=1}^n v_i w_i - \sum_{i=1}^n v_i w_i^2 / \sum_{i=1}^n w_i}\},$$

and $\sigma_i^2 + v_i\rho = 1/w_i^*$, so that θ and its standard error are now estimated as before.

The Mandel-Paule estimator (e.g. Hartung, Knapp and Sinha, 2008) minimises $Q(\tau^2) = \sum_{i=1}^{n} \frac{(\hat{\theta}_i - \tilde{\theta})^2}{\sigma_i^2 + \tau^2}$ for $\tilde{\theta}$, so that $\tilde{\theta} = \frac{\sum_{i=1}^{n} \hat{\theta}_i/(\sigma_i^2+\tau^2)}{\sum_{i=1}^{n} 1/(\sigma_i^2+\tau^2)}$ and estimates τ^2 iteratively by setting $Q = n - 1$. It can be trivially changed to $Q(\rho) = \sum_{i=1}^{n} \frac{(\hat{\theta}_i - \tilde{\theta})^2}{\sigma_i^2 + v_i\rho}$, minimized for $\tilde{\theta}$, and ρ then estimated by setting $Q(\rho) = n - 1$. Hence the standard methods of meta-analysis can be easily tweaked to make the adjustment recommended here.

2.4. Examples

Table 2 shows details of the 15 datasets used as examples. These were datasets that were comparatively easily available, and some such as the tuberculosis dataset are very well known and much studied.

The results are shown in table 3. A maximum likelihood fit for the parameters θ and τ^2 or ρ was done to evaluate the relative goodness of fit of the conventional and proposed new model, but the estimates of treatment effect θ shown with standard errors are derived using the DerSimonian and Laird procedure, and the modified version of this. This was done rather than quoting results from the maximum likelihood analysis, because the DSL method is very widely used.

Of the 15 meta-analyses studied, 10 had higher likelihood with the new method, 2 had lower likelihood (Eclamp and Resp), while 3 (Lamo, Steroid and Strep) fitted with random effect going to zero, so both methods gave identical results. In two of these latter cases the results are shown, because $\hat{\theta}$ differed using the DSL method. In the 2 cases where the new method gave

a worse fit, the decrease in log-likelihood was small, not more than 0.25. However when the fit was better, the improvement could be much larger, of the order of 2.0. This shows convincingly that the new method fits data better. The estimates of treatment effect have smaller modulus in 11 cases, and are larger in 4, big enough changes to make the new model of practical interest. The maximum change was something like 15% of treatment effect. The standard errors were always slightly smaller for the new method, which is another advantage.

The random effect $\hat{\rho}$ estimated from the modified DerSimonian and Laird method is also shown as a percentage. It correlates only weakly with the I^2 statistic, because I^2, although a measure of heterogeneity, derives from a measure of significance, whereas ρ is an absolute measure of heterogeneity.

3. Conclusions

A modification of the standard normal random effects model for 2×2 table data has been derived, where the size of the random effect is scaled to the size of the variances for the event probabilities for treatments A and B (at a fixed sample size). There is thus an effect on inference: a sample treatment effect measure such as the log-odds ratio (1) has smallest variance when $\hat{p} \simeq 1/2, \hat{q} \simeq 1/2$. The random effect will be smaller there, so that studies where $\hat{p}$ are small or large will have a larger random effect variance ρ and so be downweighted. Thus the weighting accorded to the various studies is altered. This model can be seen as the conversion of the beta-binomial regression model mentioned by Kuss (2015) to a 2-stage model. Hence the attractive properties of the beta-binomial model found by Kuss

(2015) should carry over to the 2-stage model.

This modification of a familiar model gave slightly better fits to data than the conventional model for most of the sample of datasets studied. It can be used wherever random effects occur in binary-data meta-analysis, e.g. in multivariate and network meta-analysis, in meta-regression, or in diagnostic meta-analysis, where the diagnostic odds ratio (DOR) can be given this type of random effect. Further experience with this model by other workers is of course necessary, and its computation would require only a small tweak of existing software.

A spin-off from this model is that the random-effects statistic $\hat{\rho}$ derived in section 2.2 is more meaningful than the usual variance τ^2, being a measure of heterogeneity on a scale from 0 to 1, and having the second meaning that the effective sample size for the random variation is $\simeq 1/\rho$. As pointed out by Borenstein (2017), the familiar I^2 statistic has the drawback that it is not an absolute measure of heterogeneity.

For 1-stage meta-analyses, this model offers a computationally simple method that often gives better fits to data, confirming the results of Kuss (2015). This analysis is available in appendix A in the supplementary material available online for this paper. Future work could be the comparison of the adjusted model with the standard model for a much larger sample of meta-analyses.

References

[1] Alba, A. C., Alexander, P. E., Chang, J., MacIsaac, J., DeFry, S. and Guyatt, G. H. (2016). High statistical heterogeneity is more frequent in

meta-analysis of continuous than binary outcomes, *Journal of Clinical Epidemiology*, **70**, 129-135.

[2] Baker R. D., Jackson D. (2013). Meta-analysis inside and outside particle physics: two traditions that should converge? *Research Synthesis Methods* **4**, 109124.

[3] Borenstein, M., Higgins, J. P. T., Hedges, L. V. and Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, DOI: 10.1002/jrsm.1230

[4] Borenstein, M., Hedges, L. V., Higgins, J. P. T. and Rothstein, H. R. (2009). Introduction to meta-analysis, Wiley, New York

[5] Chen D-G D, Peace KE (2013). Applied Meta-Analysis with R, Chapman and Hall/CRC.

[6] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7:** 177–188.

[7] Egger, M., Altman, D. G., Smith, G. D. (2001). Systematic Reviews in Health Care 2e: Meta-analysis in Context, 2nd. ed., BMJ Publishing Group, London

[8] Fergusson, D., Glass, K. C., Hutton, B., and Shapiro, S. (2005). Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding?. *Clinical Trials*, 2(3):218–229.

[9] Hartung, J., Knapp, G. and Sinha, B. K. (2008). Statistical meta-analysis with applications, Wiley, New York.

[10] Higgins, J., Thompson, S. G., Deeks, J. J. and Altman, D. G. (2003). measuring inconsistency in meta-analyses, *British Medical Journal*, **327**, 557-560.

[11] Kuss, O. (2015). Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. *Statistics in Medicine*, **34** (7), 1097-1116.

[12] Oehlert, G. W.. (1992) A note on the delta method, *The American Statistician*, **46** (91), 27-29.

[13] Pani P. P., Trogu, E., Pacini M. and Maremmani, I. (2014). Anticonvulsants for alcohol dependence, Cochran Systematic Review CD008544, analysis 1.1

[14] Prentice, R. L. (1986). Regression using an extended beta-binomial distribution,*Journal of the American Statistical Association*, **81**, 321-327.

[15] Ramaratnam, S., Marson, A. G. and Baker, G.A. (2001). Lamotrigine add-on for drug resistant partial epilepsy. Cochrane Database of Systematic Reviews 2001, 3.

[16] Sharp, S. J., Thompson, S. G. and Altman, D. G. (1996). *British Medical Journal* **313**, 735–738.

[17] Sidik, K. and Jonkman, J. N. (2008). Estimation using non-central hypergeometric distributions in combining tables. *Journal of Statistical Planning and Inference*, **138** (12): 3993–4005.

[18] Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. and Thompson, S. G. (2000). A multilevel framework for meta-analysis of clinical trials with binary outcomes, *Statistics in Medicine* **19**, 3417-3432.

[19] Turner,R. M., Davey, J., Clarke,M. J., Thompson, S. G. and Higgins, J. P. T. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews, *International Journal of Epidemiology*, **41**, 818-827.

[20] Warn, D. E., Thompson, S. G. and Spiegelhalter, D. J., Bayesian random-effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales, *Statistics in Medicine* (2002). **21**, 1601–1623.

Tables

event↓,group→	Treatment A	Treatment B
Yes	n_{11}	n_{12}
No	n_{21}	n_{22}
Total	N_1 or N_p	N_2 or N_q

Table 1: Notation for 2×2 tables; columns are the group, rows the event, e.g. recovery or death.

Name	Description	Reference	Studies
Amant	Amantidine for influenza	Higgins *et al.* [10]	8
Anti	Anticonvulsants for alcoholism, dropouts	Pani *et al.* [13]	17
Apro	Aprotinin to reduce perioperative bleeding	Fergusson [8]	64
Circum	Circumcision and HIV prevalence	Borenstein *et al.* [4]	33
Cis	Cisapride for nonulcer dyspepsia	Hartung *et al.* [9]	13
Eclamp	Diuretics for pre-eclampsia	Turner *et al.* [18]	9
Endo	New surgical therapy for bleeding peptic ulcers	Sidik and Jonkman [17]	41
Ibup	Ibuprofen for post-operative pain	Warn *et al.* [20]	46
Lamot	Lamotrigine for drug-resistant partial epilepsy	Ramaratnam *et al.* [15]	11
Resp	Selective decontamination for RPI	Turner *et al.* [18]	22
Sclero	Endoscopic sclerotherapy for liver disease	Sharp *et al.* [16]	19
Smoking	Smoking cessation	Baker & Jackson [2]	111
Steroid	Steroids for reducing neonatal deaths	Chen and Peace [5]	7
Strept	Intravenous streptokinase in MI	Egger *et al.* [7]	22
TB	BCG vaccine for TB prevention	Hartung *et al.* [9]	13

Table 2: Summary of the 15 datasets used in our empirical investigation. RPI denotes respiratory tract infection and MI denotes myocardial infarction.

Name	$I^2\%$	$-\ell$	$-\ell_{\text{new}}$	$\hat{\theta}$	$\hat{\theta}_{\text{new}}$	$100\hat{\rho}$
Amant	47.0	1.207	0.888	-1.098 (.231)	-1.011 (.208)	0.73
Anti	57.7	4.584	4.403	-0.144 (.197)	-0.156 (.195)	3.13
Apro	44.0	19.814	19.501	-1.032 (.088)	-1.006 (.086)	1.41
Circum	92.3	6.999	5.983	-0.513 (.150)	-0.586 (.129)	3.11
Cis	71.2	7.547	7.503	1.491 (.308)	1.361 (.306)	7.98
Eclamp	70.8	1.222	1.421	-0.518 (.204)	-0.425 (.181)	0.74
Endo	54.6	32.468	30.480	-0.997 (.189)	-0.924 (.184)	0.56
Ibup	50.7	30.475	29.022	2.022 (.151)	1.797 (.140)	2.55
Lamo	0	0.530	0.530	0.950 (.167)	0.950 (.167)	0.00
Resp	63.8	8.406	8.658	-1.221 (.167)	-1.153 (.163)	2.09
Sclero	56.2	3.266	2.675	-0.339 (.175)	-0.333 (.173)	2.96
Smoking	26.3	-31.947	-32.354	0.567 (.0392)	0.562 (.0387)	0.20
Steroid	12.5	-0.3851	-0.3851	-0.6292 (.192)	-0.616 (.190)	0.13
Strept	33.3	-7.431	-7.431	-0.245 (.0622)	-0.247 (.0621)	0.09
TB	95.3	1.3964	-0.5435	-0.856 (.225)	-1.009 (.145)	0.18

Table 3: 2-stage meta-analysis results using standard and new methods for the log-odds ratio. Minus the log-likelihood is shown, plus treatment effect estimates obtained using the DerSimonian and Laird procedure, with standard errors in parentheses, and the random-effect correlation $\hat{\rho}$.