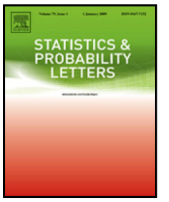
ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro



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

Rose Baker

School of Business, University of Salford, UK

ARTICLE INFO

Article history:
Received 24 May 2017
Received in revised form 27 October 2017
Accepted 28 October 2017
Available online xxxx

Keywords:
Meta-analysis
Random effect
Binary data
Beta distribution

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 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

E-mail address: r.d.baker@salford.ac.uk.

https://doi.org/10.1016/j.spl.2017.10.018
0167-7152/

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

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

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

$$\theta = \ln(q/(1-q)) - \ln(p/(1-p)) = \ln(q(1-p))/(p(1-q)). \tag{1}$$

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 δp, δq we have that $\delta\theta \simeq (\partial g/\partial p)\delta p + (\partial g/\partial q)\delta q$, and assuming that δp, δq 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

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

Hence finally

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

Please cite this article in press as: Baker R., An adjusted random-effects model for binary-data meta-analysis. Statistics and Probability Letters (2017), https://doi.org/10.1016/j.spl.2017.10.018.