Contemporary Clinical Trials Communications 10 (2018) 141–147

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

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

An integrative shrinkage estimator for random-effects meta-analysis of rare binary events

Lie Li, Ou Bai, Xinlei Wang*

Department of Statistical Science, Southern Methodist University, 3225 Daniel Avenue, P O Box 750332, Dallas, TX 75275-0332, USA

ARTICLE INFO

Keywords:
Bias
Estimation efficiency
Log odds ratio
Mean squared error
Sparse binary data

ABSTRACT

Meta-analysis has been a powerful tool for inferring the treatment effect between two experimental conditions from multiple studies of rare binary events. Recently, under a random-effects (RE) model, Bhaumik et al. developed a simple average (SA) estimator and showed that with the continuity correction factor 0.5, the SA estimator was the least biased among a set of commonly used estimators. In this paper, under various RE models that allow for treatment groups with equal and unequal variability (in either direction), we develop an integrative shrinkage (iSHRI) estimator based on the SA estimator, which aims to improve estimation efficiency in terms of mean squared error (MSE) that accounts for the bias-variance tradeoff. Through simulation, we find that iSHRI has better performance in general when compared with existing methods, in terms of bias, MSE, type I error and confidence interval coverage. Data examples of rosiglitazone meta-analysis are provided as well, where iSHRI yields competitive results.

1. Introduction

In medical research, when events of interest are rare, a single randomized study rarely has sufficient power to provide reliable information regarding the treatment effect between two experimental conditions (say, treatment vs. control). Therefore, meta-analysis is often used to combine information from multiple studies of rare binary events. In the past, various meta-analysis approaches have been developed to estimate the overall treatment effect, based on either fixed-effect (FE) models [e.g., 18] or random-effects (RE) models [e.g., 10]. Note that FE models assume a common treatment effect across different studies while RE models allow the treatment effects to vary from study to study.

When dealing with rare binary events, under the FE assumption, commonly used methods for estimating the treatment effect include the Mantel-Haenszel [MH, 18] with a constant zero-cell correction factor 0.5, Peto [30], and inverse variance methods [10,14]. A previous study [3] has compared the performance of twelve FE methods. The general recommendation is to use the MH method with some appropriate continuity corrections, which is consistent with the findings of Sweeting et al. [25]. In practice, RE models seem to be less restrictive, especially for clinical trials, because doses and follow-up time can be different in individual studies. Through a meta-analysis of multiple rosiglitazone studies, Shuster et al. [22] pointed out that the performance based on RE models is superior to that based on FE models. In this paper, we focus on meta-analysis of rare binary events based on RE models.

For RE meta-analysis, Shuster [21] showed that the widely used DerSimonian and Laird [DSL, 10] method can be highly biased for rare events, and suggested to use the simple (unweighted) average of estimates from individual studies. Recently, Bhaumik et al. [2] formally proposed a simple average (SA) estimator based on a RE model, which is the unweighted average of estimated log odd ratios (with a positive continuity correction factor a) in individual studies. Bhaumik et al. [2] showed that, when $a = \frac{1}{2}$, the SA estimator (SA_0.5) is asymptotically unbiased and has superior bias performance when compared with existing estimators, including MH, empirical logit [EL, 10], and DSL methods. However, Li and Wang [17] pointed out that the RE model they assumed is restrictive in the sense that it forces the variability in the treatment group to be no less than that in the control group, and more importantly, SA_0.5 fails to minimize the mean squared error (MSE), which is an established measure of estimation performance that takes into account the bias-variance tradeoff.

Based on the SA estimator, we aim to develop a shrinkage estimator with smaller MSE to improve estimation efficiency. Shrinkage methods, which shrink some "standard" estimator toward zero or any other fixed value, have been widely used in various fields [24,26–29]. Many shrinkage methods [7,8,29] were developed under a rigorous Bayesian framework via empirical Bayes (EB) approaches, which shrink a point estimate from the sample to the prior. By contrast, the others were derived based on statistical decision theory directly (by minimizing

* Corresponding author.
E-mail address: swang@smu.edu (X. Wang).

https://doi.org/10.1016/j.conctc.2018.04.004
Received 27 December 2017; Received in revised form 8 March 2018; Accepted 9 April 2018
Available online 16 April 2018

some loss function), of which an early example is the famous James–Stein estimator [6,16]. A recent example is that, to estimate the failure rate of rare events from multiple heterogeneous systems, Xiao and Xie [28] developed a plug-in shrinkage estimator based on Poisson distributions, which shrinks the maximum likelihood estimator (MLE) of the failure rate toward a predetermined point in the parameter space, and so lifts degenerated zero estimates to some reasonable positive values. This estimator can be expressed as a weighted average of the MLE and a data-dependent point.

In the context of meta-analysis of rare binary events, we develop an integrative shrinkage estimator (iSHRI) by using the SA method in Bhaumik et al. [2] to combine shrinkage estimators from individual studies, which intends to shrink the estimated log odd ratios with $a = \frac{1}{2}$ toward a predetermined point. The shrinkage factor is obtained from minimizing the MSE and then plug in estimates from the data. We thoroughly compare the bias and MSE of the proposed method with existing methods via simulation. We further examine their performance on hypothesis testing and interval estimation using the type I error and coverage probability of confidence intervals (CIs). Besides, two data examples of rosiglitazone meta-analysis are provided for further comparison.

2. An integrative shrinkage estimator

Consider a meta-analysis consisting of K randomized studies. In the kth study, let n_{kt} and n_{kc} be the number of subjects in the treatment and control groups, respectively. Assume $x_{kt} \sim \text{Binomial}(n_{kt}, p_{kt})$ and $x_{kc} \sim \text{Binomial}(n_{kc}, p_{kc})$, where $x_{kt}(x_{kc})$ is the number of observed events of interest in the kth study, and $p_{kt}(p_{kc})$ is the probability of observing an event in the treatment (control) group. Let $q_{kt} \equiv 1 - p_{kt}$ and $q_{kc} \equiv 1 - p_{kc}$. To measure the treatment effect in study k, the log odds ratio is used throughout this paper, i.e., $\theta_k \equiv \ln\left(\frac{p_{kt}/q_{kt}}{p_{kc}/q_{kc}}\right)$. We denote the mean treatment effect across component studies by θ, and the between-study heterogeneity among individual treatment effects by τ_θ^2, satisfying $\tau_\theta^2 > 0$. Below we consider three RE models, to accommodate realistic situations where treatment and control groups can have equal variability or unequal variability (in either direction). All the three models involve random terms θ_k's and μ_k's, where $\theta_k \overset{iid}{\sim} N(\theta, \tau_\theta^2)$, $\mu_k \overset{iid}{\sim} N(\mu, \tau_\mu^2)$, and any two components of $(\mu_1,\cdots,\mu_K; \theta_1,\cdots,\theta_K)$ are assumed to be independent. Here, μ_k can represent the log odds of the control or treatment group or the average of the two groups, depending on model specification below.

1. Model I is the one used in Bhaumik et al. [2], which implicitly assumes that the variance of $\text{logit}(p_{kt})$ is greater than that of $\text{logit}(p_{kc})$:
$$\text{logit}(p_{kc}) = \mu_k, \quad \text{logit}(p_{kt}) = \mu_k + \theta_k.$$

2. Model II is the one used in Smith et al. [23], which assumes equal variances between the two groups:
$$\text{logit}(p_{kc}) = \mu_k - \frac{\theta_k}{2}, \quad \text{logit}(p_{kt}) = \mu_k + \frac{\theta_k}{2}.$$

3. Model III assumes the variance in the treatment is less than that in the control:
$$\text{logit}(p_{kc}) = \mu_k - \theta_k, \quad \text{logit}(p_{kt}) = \mu_k.$$

Although originally developed under Model I, the SA estimator $\hat{\theta}_a$ can be used with any of the three models to estimate the mean treatment effect θ, given by:

$$\hat{\theta}_a = \frac{1}{K}\sum_{k=1}^{K} \hat{\theta}_{ka}, \tag{1}$$

where $\hat{\theta}_{ka}$ is an estimator of the individual treatment effect θ_k in the kth study:

$$\hat{\theta}_{ka} = \ln\frac{x_{kt}+a}{n_{kt}-x_{kt}+a} - \ln\frac{x_{kc}+a}{n_{kc}-x_{kc}+a}, \tag{2}$$

and a is a positive continuity correction factor. Bhaumik et al. [2] proved that $\hat{\theta}_a$ is asymptotically unbiased when $a = \frac{1}{2}$ under Model I; that is, $E\left(\hat{\theta}_{\frac{1}{2}}\right) = \theta + O(n^{-2})$, where $n \equiv min[(n_{kt}, n_{kc})_{k=1}^K]$ is the overall minimum number of subjects. It can be shown that the asymptotic unbiasedness of $\hat{\theta}_{\frac{1}{2}}$ also holds under Models II and III. This theoretical property ensures that SA_0.5 (i.e., $\hat{\theta}_{\frac{1}{2}}$) performs well in terms of bias for large sample sizes. However, Li and Wang [17] proved that SA_0.5 is suboptimal in terms of MSE, and showed via simulation that it can have poor MSE performance especially for small sample sizes.

Motivated by Xiao and Xie [28], in the kth study, we consider a shrinkage estimator of θ_k based on $\hat{\theta}_{k\frac{1}{2}}$, denoted by $\hat{\theta}_{SHRI.k}$,

$$\hat{\theta}_{SHRI.k} = c\hat{\theta}_{k\frac{1}{2}} + (1-c)\theta_0,$$

where θ_0 is a fixed point in the parameter space of θ, and c is a shrinkage factor. Then our integrative shrinkage estimator for the overall treatment effect θ can be given by

$$\hat{\theta}_{iSHRI} = \sum_{k=1}^{K}\frac{\hat{\theta}_{SHRI.k}}{K} = c\hat{\theta}_{\frac{1}{2}} + (1-c)\theta_0.$$

Clearly, the iSHRI estimator shrinks SA_0.5 toward the predetermined point θ_0.

Based on the asymptotic unbiasedness of $\hat{\theta}_{\frac{1}{2}}$, it is easy to show

$$\text{Bias}(\hat{\theta}_{iSHRI}) = (1-c)(\theta-\theta_0) + O(n^{-2}).$$

Further, we can show that the variance of $\hat{\theta}_{iSHRI}$ is given by

$$V(\hat{\theta}_{iSHRI}) = c^2V\left(\hat{\theta}_{\frac{1}{2}}\right) = \frac{c^2}{K^2}\sum_{k=1}^{K}\left[E_{p_{kt}}\left(\frac{1}{n_{kt}p_{kt}q_{kt}}\right) + E_{p_{kc}}\left(\frac{1}{n_{kc}p_{kc}q_{kc}}\right)\right] + \frac{c^2}{K}\tau_\theta^2 + O(n^{-2}).$$

Thus, regardless of the model specification, $\hat{\theta}_{iSHRI}$ has the following MSE:

$$\text{MSE}(\hat{\theta}_{iSHRI}) = (1-c)^2(\theta-\theta_0)^2 + \frac{c^2}{K}\tau_\theta^2 + \frac{c^2}{K^2}\sum_{k=1}^{K}\left[E_{p_{kt}}\left(\frac{1}{n_{kt}p_{kt}q_{kt}}\right) + E_{p_{kc}}\left(\frac{1}{n_{kc}p_{kc}q_{kc}}\right)\right] + O(n^{-2}).$$

To minimize the asymptotic MSE of $\hat{\theta}_{iSHRI}$, we ignore the term $O(n^{-2})$ and set the shrinkage factor c to

$$c^* = \frac{(\theta-\theta_0)^2}{(\theta-\theta_0)^2 + \frac{\tau_\theta^2}{K} + \frac{1}{K^2}\sum_{k=1}^{K}\left[E_{p_{kt}}\left(\frac{1}{n_{kt}p_{kt}q_{kt}}\right) + E_{p_{kc}}\left(\frac{1}{n_{kc}p_{kc}q_{kc}}\right)\right]}, \tag{3}$$

satisfying $c^* \le 1$. The minimized asymptotic MSE of iSHRI, denoted by mAMSE, is always less than or equal to the asymptotic MSE of SA_0.5:

$$\text{mAMSE}(\hat{\theta}_{iSHRI}) = c^*\left\{\frac{\tau_\theta^2}{K} + \frac{1}{K^2}\sum_{k=1}^{K}\left[E_{p_{kt}}\left(\frac{1}{n_{kt}p_{kt}q_{kt}}\right) + E_{p_{kc}}\left(\frac{1}{n_{kc}p_{kc}q_{kc}}\right)\right]\right\} = c^*\text{AMSE}\left(\hat{\theta}_{\frac{1}{2}}\right) \le \text{AMSE}\left(\hat{\theta}_{\frac{1}{2}}\right).$$

To be able to calculate the shrinkage factor c^* in (3), we need to estimate θ, τ_θ^2, $E_{p_{kt}}\left(\frac{1}{n_{kt}p_{kt}q_{kt}}\right)$ and $E_{p_{kc}}\left(\frac{1}{n_{kc}p_{kc}q_{kc}}\right)$ from the data as well as choosing an appropriate value of θ_0. We estimate θ using SA_0.5. Based on some preliminary simulation, we find that how to estimate τ_θ^2 (e.g., the popular DSL estimator or other estimators introduced in Bhaumik et al. 2), has not much effect on the performance of iSHRI. Thus, we set $\hat{\tau}_\theta^2 = 0$ for simplicity. We also adopt the estimators in Gart et al. [12] for $E_{p_{kt}}\left(\frac{1}{n_{kt}p_{kt}q_{kt}}\right)$ and $E_{p_{kc}}\left(\frac{1}{n_{kc}p_{kc}q_{kc}}\right)$: