Published in final edited form as:

Commun Stat Simul Comput. 2016 ; 45(8): 3036–3052. doi:10.1080/03610918.2014.911901.

Meta-analysis of Proportions of Rare Events—A Comparison of Exact Likelihood Methods with Robust Variance Estimation

Yan Ma^{1,2}, Haitao Chu³, and Madhu Mazumdar¹

¹Department of Public Health, Weill Medical College of Cornell University, New York, NY 10021

²Hospital for Special Surgery, New York, NY 10021

³School of Public Health, University of Minnesota, Minneapolis, MN 55455

Abstract

The conventional random effects model for meta-analysis of proportions approximates within-study variation using a normal distribution. Due to potential approximation bias, particularly for the estimation of rare events such as some adverse drug reactions, the conventional method is considered inferior to the exact methods based on binomial distributions. In this paper, we compare two existing exact approaches—beta binomial (B-B) and normal-binomial (N-B)—through an extensive simulation study with focus on the case of rare events that are commonly encountered in medical research. In addition, we implement the empirical (“sandwich”) estimator of variance into the two models to improve the robustness of the statistical inferences. To our knowledge, it is the first such application of sandwich estimator of variance to meta-analysis of proportions. The simulation study shows that the B-B approach tends to have substantially smaller bias and mean squared error than N-B for rare events with occurrences under five percent, while N-B outperforms B-B for relatively common events. Use of the sandwich estimator of variance improves the precision of estimation for both models. We illustrate the two approaches by applying them to two published meta-analysis from the fields of orthopedic surgery and prevention of adverse drug reactions.

1. Introduction

Proportions such as the incidence of clinical events among a cohort of patients or the response rate in patients taking a certain treatment regimen are commonly reported outcomes in epidemiologic and medical research. Often these are rare events and the resulting proportions are very low. For example, a recent study of outcomes after total knee arthroplasty (TKA) for 8325 patients in the global orthopaedic registry found a 0.2% incidence of death as well as incidences of 1.4% and 0.8% for the most common in-hospital complications, DVT and cardiac events, respectively (Cushner et al. 2010). It is extremely difficult to estimate such rare events with adequate statistical power at a single institution or in a single study due to the fact that when the numerator of a proportion is small, the size of the denominator needed for adequate precision is even higher than otherwise. Meta-analysis, the method for pooling the estimated outcomes of interest from independent studies while taking into account the between-study heterogeneity, helps enhance the quality of evidence and produces improved power and precision. Meta-analysis of rare clinical events are therefore very common in published literature (e.g., Lazaros et al. 1998, Muller et al. 2010,

Warycha et al. 2009). In this paper, we consider the methodologies underlying the meta-analysis of proportions with focus on non-comparative (i.e., single-arm) studies, where a sample size and a number of events are reported in each study.

Two recently published meta-analyses motivated us to perform this research. The first meta-analysis (Dy et al. 2012) estimated the incidence of a set of complications following patello-femoral arthroplasty (PFA). The study consisted of 23 relatively small studies (average sample size= 50, standard deviation (SD)= 29, median=46, 1st quartile ($Q1$)= 26, 3rd quartile ($Q3$)= 63). Most of the complications were associated with low incidences. For example, one of the outcomes of interest named “other complications” had a mean incidence of 1% across a total of 1154 patients. Further, 18 (78%) of the 23 studies reported 0 incidence and the highest incidence found was 12%. More than 95% of the studies reported an incidence of “other complications” under 5% (Figure 1 (a)). The second meta-analysis (Hakkarainen et al. 2012) evaluated the proportion of patients with preventable adverse drug reactions (PADRs). It described 16 large studies (average sample size= 3050, SD= 5046, median=987, $Q1$ = 640, $Q3$ = 1911). The average incidence of PADRs was estimated to be 3.5% amongst outpatients over 48797 emergency visits or hospital admissions. The reported incidence ranged from 0.08% to 9% with 85% of them being less than 6% (Figure 1(b)).

The conventional statistical method for meta-analysis of proportions is based on the following random effects model (DerSimonian and Laird, 1986):

$$Y_i = \beta + b_i + \varepsilon_i; i=1, \dots, K, \quad (1)$$

where for K independent studies, Y_i represents the chosen measure of effect, β the population effect, b_i the random between-study effect, and ε_i the sampling error. The outcome measure Y_i is a function of summary statistics such as the logit proportion in a non-comparative study. By convention, both b_i and ε_i are assumed to follow normal distributions, $b_i \sim N(0, \tau^2)$, $\varepsilon_i \sim N(0, \sigma_i^2)$, and τ^2 and $\sigma_i^2 (i=1, \dots, K)$ represent between- and within-study variances, respectively. The model (1) is equivalent to a hierarchical model in the form of

$$Y_i \sim N(\beta_i, \sigma_i^2) \quad (2)$$

and

$$\beta_i \sim N(\beta, \tau^2), i=1, \dots, K, \quad (3)$$

describing the within- and between-study distributions, respectively. Typical estimation procedures for the random effects model (1) include likelihood-based methods (e.g., maximum likelihood (Hardy and Thompson, 1996) or restricted maximum likelihood (Raudenbush and Bryk, 1985)) or the method of moments (DerSimonian and Laird, 1986). In all cases, the parameter of primary interest β is estimated as a weighted average of

estimated effects in individual studies $\hat{\beta} = \frac{\sum_{i=1}^K w_i y_i}{\sum_{i=1}^K w_i}$, where y_i denotes the sample version of Y_i and $w_i = (\hat{\tau}^2 + \sigma_i^2)^{-1}$. The within-study variance σ_i^2 is considered known and usually estimated from a normal distribution. Specifically, for the outcome of proportion, the effect size is taken as the $\text{logit} Y_i = \log\left(\frac{P_i}{1 - P_i}\right)$ where P_i represents the proportion of event within the i th study. The within-study variance of Y_i is then estimated by

$$\hat{\sigma}_i^2 = \frac{1}{x_i} + \frac{1}{n_i - x_i}, \quad (4)$$

where x_i and n_i denote the number of events and sample size in the i th study, $i = 1, \dots, K$.

Despite the popularity of the conventional random effects model, limitations have been pointed out when the model is applied to binary outcomes. These issues have been discussed thoroughly in the published literature, especially in the context of meta-analysis of comparative studies (Bhaumik et al. 2012, Chu et al. 2010, Kuss and Gromann 2007, Shuster et al. 2007, Sweeting et al. 2004). Here we highlight two major limitations of the conventional model for handling rare events. First, the within-study distribution is approximated by normal distribution (2). When the within-study sample size n_i is small

(e.g., PFA study) or the proportion $p_i = \frac{x_i}{n_i}$ is close to 0 (e.g., PFA and PADR studies), the normality assumption of the model might not hold resulting in biased estimation and invalid inference (Hamza et al. 2008, Platt et al. 1999, Stijnen et al. 2010). Second, when handling “zero”-event (i.e., $x_i = 0$), the logit proportion and its within-study variance (4) become undefined. To get around the issue of “zero”-event, a widely used strategy is to add an arbitrary positive number (e.g., 0.5) to the x_i and n_i (named “continuity correction”). However, such a correction has been shown to induce further bias (Sweeting et al. 2004, Hamza et al. 2008). One suggestion made for confronting these issues has been to replace the normal approximation of the within-study distribution with the exact distribution—binomial distribution—through nonlinear mixed effects models. Hamza et al. (2008) proposed a Normal (between-study distribution)–Binomial (within-study distribution) model and compared it with the Normal–Normal (N-N) random effects model (2, 3) for pooling logit proportions from non-comparative studies. In the related simulation study, they found the N-N model to have large bias with poor coverage rates, while the Normal-Binomial (N-B) model was consistently superior. In addition to the N-B model, Young-Xu and Chan (2008) introduced meta-analysis of proportions based on the Beta-Binomial (B-B) distribution. Chu et al. (2010) and Kuss et al. (2013) extended the B-B distribution to bivariate meta-analysis of diagnostic accuracy studies. In the presence of two exact methods, a clear guidance on which one to use in what scenario is needed. However, to the best of our knowledge, no study has compared these methods for analysis of rare events yet.

Another issue of concern is the fact that all of the methods mentioned above assumes between-study distributions either normal or beta distribution, while in practice the true distribution of proportion is unknown. To minimize the impact of model misspecification

and provide robust statistical inferences, we propose to integrate the sandwich variance estimator within both exact methods. The sandwich estimator (White, 1982), also known as the empirical variance-covariance matrix estimator, had been a very useful tool for variance estimation. Its main property is that it generates consistent estimates of the variance-covariance matrix for the parameter estimates without the need for fulfillment of the underlying distributional assumptions adequately. Since its introduction, the sandwich estimator has been applied to a variety of models including the generalized estimating equations (Liang and Zeger, 1986), the Cox proportional hazards model (Lin and Wei, 1989), and the conditional logistic regression (Fay et al. 1998). Although both the nonlinear mixed effects models and the sandwich variance estimators are well-established in the literature, there has been no comprehensive study of their combined use and detailed description of their performance in the context of meta-analysis of proportions. Our goal is to compare the two models and study their performances with and without sandwich variance estimators through an extensive simulation study and provide guidance for the use of these models in meta-analysis of proportions.

The outline of the paper is as follows: we briefly describe the B-B and N-B models and incorporate the sandwich variance estimator in both models in Section 2. A statistical simulation comparing the two models and describing their operating characteristics in meta-analysis of rare events follows in Section 3. Analyses of data from the PFA and PADR studies are described in Section 4. Section 5 includes our conclusions and ideas for future research.

2. Models

The B-B and N-B models for meta-analysis of proportions consider K independent studies asking the same research question. Let X_i denote the total number of events of interest observed from a total of n_i subjects in the i th study with a probability of events p_i , $i = 1, 2, \dots, K$. Both models obtain an estimate of the pooled proportion in two stages. In the first stage, conditional on (n_i, p_i) , the X_i is assumed to follow a binomial within-study distribution $B(n_i, p_i)$, that is,

$$P(X_i = x_i | n_i, p_i) = \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}, 0 < p_i < 1, i = 1, 2, \dots, K. \quad (5)$$

In the second stage, the marginal distribution of p_i is specified. Specifically, the N-B model assumes a normal distribution of p_i on a logit scale while the B-B model assumes a beta-distribution of p_i on its original scale.

2.1 Beta-Binomial Model

Although the number of events X_i is binomial in nature, the between-study variation in p_i can produce X_i that may be more variable than expected under a binomial distribution. This phenomenon, often referred to as overdispersion, is common when modeling proportions. In meta-analysis, between-study heterogeneity due to differences in factors such as sample size, study design, physician expertise, and patient health condition can cause

overdispersion. To account for the heterogeneity, the B-B model assumes that the probability p_i follows a beta distribution $B(\alpha, \beta)$ with probability density function

$$f(p_i|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1-p_i)^{\beta-1} \quad (6)$$

where $\alpha > 0$, $\beta > 0$, and the mean of p_i is given by

$$\mu_{B-B} = \frac{\alpha}{\alpha+\beta}. \quad (7)$$

The beta distribution is very flexible for modeling proportions, as its density can display a variety of shapes depending on the value of the parameters α and β . Further, by combining the beta-distribution (6) and the binomial distribution (5), the density function of the beta-binomial distribution is given by

$$P(X_i=x_i) = \binom{n_i}{x_i} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x_i)\Gamma(\beta+n_i-x_i)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n_i)}. \quad (8)$$

Let $\gamma = \frac{1}{\alpha+\beta+1}$, which is a measure of overdispersion accounting for heterogeneity in meta-analysis (Griffiths 1973). The mean and variance of X_i under the beta-binomial distribution are functions of α and β , and can be parameterized in terms of μ_{B-B} (7) and γ so that $E(X_i) = n_i \mu_{B-B}$ and $Var(X_i) = n_i \mu_{B-B} (1 - \mu_{B-B}) \gamma$. The B-B model is a true random effects model in the sense that the number of events follows a binomial distribution, conditional on a random probability p_i , and the p_i , $i = 1, 2, \dots, K$, follow a common beta-distribution. In addition, the B-B model has a closed-form likelihood, so all the complicated estimation methods (e.g., penalized quasi-likelihood, Markov chain Monte Carlo) for normal random effects are not necessary.

2.2 Normal-Binomial Model

The N-B model considers a generalized linear mixed effects model on a logit scale of p_i with a normal distribution

$$\text{logit}(p_i) = \mu_0 + b_i, \quad b_i \sim N(0, \tau^2), \quad (9)$$

where μ_0 and τ^2 represent mean and between-study variance of the transformed p_i , respectively. The mean of p_i in the N-B can be obtained through the integration

$$\mu_{N-B} = \int_{-\infty}^{+\infty} \text{logit}^{-1}(\mu_0 + z) f_{b_i}(z) dz, \quad (10)$$

where $f_{b_j}()$ denotes the normal density function of the random effect b_j in (9). In this paper, the B-B and N-B models are compared via the mean proportions (7) and (10).

2.3 Sandwich Estimator of Variance

To introduce the sandwich estimator, let $l_K(\eta) = -\log L_K(\eta)$, where $L_K(\eta)$ denotes the likelihood function of the N-B or the B-B model and η the vector of parameters of interest. Specifically, $\eta = (\mu_0, \tau^2)^\top$ in the N-B model and $\eta = (\mu, \gamma)^\top$ in the B-B model. Define

$$\hat{V}_K(\hat{\eta}) = \sum_{i=1}^K v_i(\hat{\eta}) v_i(\hat{\eta})^\top, \quad \hat{J}_K(\hat{\eta}) = l_K''(\hat{\eta}), \quad (11)$$

where v_i is the first derivative of the contribution to l_K by the i th study and l_K'' is the second derivative matrix of l_K with respect to η . Then the sandwich estimator of the variance-covariance matrix of $\hat{\eta}$ is given by (White, 1982)

$$\hat{J}_K(\hat{\eta})^{-1} \hat{V}_K(\hat{\eta}) \hat{J}_K(\hat{\eta})^{-1}. \quad (12)$$

The variance of the pooled proportion can then be obtained using the Delta method.

2.4 Parameter Estimation

Since both N-B and B-B belong to the class of nonlinear mixed effects models, we fit the two models using the SAS procedure NLMIXED (SAS Institute Inc. Cary, NC). The maximum likelihood estimation method is implemented through the procedure. The sandwich estimator of variance is readily obtained with the “empirical” option. In the appendix, we have given the syntax to fit the models in SAS. Since no built-in likelihood function is available in PROC NLMIXED for the B-B model, we calculated the sandwich estimator in R (R development core team. Vienna, Austria, 2012) by following formula (11) and (12). The additional programming in R is available from the first author upon request. A specially prepared program in SAS that can calculate the sandwich estimator for B-B is also provided in the appendix. Because parameter estimation and statistical inference for nonlinear mixed effects models have been discussed extensively in the literature (Hamza et al. 2008, Young-Xu et al. 2008), we are skipping description of the related technical issues in this paper.

3. Simulation study

3.1 Simulation method

A simulation study was carried out to compare the performance of the B-B and N-B models and study the properties of the sandwich estimator of variance in the context of meta-analysis. We assessed the effects of the following parameters in a variety of scenarios: the number of studies in the meta-analysis, the mean within-study sample size, and the true mean and variance of the proportions. Data were generated in three steps:

Step 1. We conducted the simulation study for two sample sizes $K = 25$ and $K = 50$ studies per meta-analysis. We considered two sets of within-study sample sizes (n_i), which had similar mean and SD to those observed in the PFA example dataset (1st set: mean = 50, SD = 30) and PADRs example dataset (2nd set: mean = 3000, SD = 5000). The first set was generated from a normal distribution $N(50, 30^2)$ and any n_i smaller than 10 was set to 10. If the second set of n_i s were also generated from a normal distribution and truncated at 10, the mean and SD of the simulated n_i s would deviate substantially from the assumed values (i.e., 3000 and 5000). Instead, setting $n_i = z_i^{1.5}$, where z_i was generated from an exponential distribution with mean of 175, we were able to obtain n_i s that had a mean and SD of approximately 3000 and 5000 and followed a distribution similar to that of the PADRs study.

Step 2. To mimic the empirical distributions of the proportions in the two motivating examples, we generated p_i from a mixture of two uniform distributions (Craigmire and Tirrerington, 1997). The density function of the distribution is given by

$$f(p_i|\pi, 0, b) = \pi U(0, \theta) + (1 - \pi)U(\theta, b), \quad 0 < \pi < 1, 0 < \theta < b < 1, \quad (13)$$

where $U(a_1, a_2)$ denotes a uniform distribution on the interval $[a_1, a_2]$ and π represents the probability associated with the first uniform distribution $U(0, \theta)$. When the values of π and θ approach 1 and 0, respectively, the majority of the probabilities p_i will have small values, simulating outcomes with rare events. We plotted the empirical and model (mixture of two uniform distributions)-based probability density functions (PDF) in Figures 1(a) and 1(b) for the PFA and PADRs studies, respectively. Specifically, the parameters (π, θ, b) in (13) were set to be (0.95, 0.025, 0.14) for Figure 1(a) and (0.85, 0.06, 0.12) for Figure 1(b). The PDFs estimated from the model appear to capture the key patterns of the data. In this simulation study, we set $\pi = 0.9$ and chose a wide range of values for the mean $\mu_p \equiv E(p_i) = \{0.5\%, 1\%, 3\%, 5\%, 10\%, 25\%\}$. With these selected values, we were able to study thoroughly the performance of the two models for rare events, particularly when $\mu_p < 10\%$. In addition, because meta-analysis of proportions involves binary outcomes with associated probabilities symmetric around 0.5, it is sensible to study values below 0.5 for the mean rather than its entire range. Let $b = \mu_p + \delta$, where δ denotes the distance between μ_p and the upper bound b of p_i . For fixed μ_p and π it follows to derive $\theta = \delta(\pi - 1) + \mu_p(1 + \pi)$. Applying the constraint $0 < \theta < b < 1$, δ follows

$$\frac{\mu_p \pi}{2 - \pi} < \delta < \min \left(1 - \mu_p, \frac{\mu_p(1 + \pi)}{(1 - \pi)} \right). \quad (14)$$

For each μ_p , two values of δ were selected at the 10th and 90th percentiles of the interval (14). Because of the monotonic relationship between the variance $\sigma_p^2 \equiv \text{Var}(p_i)$ and those δ s complying with (14), two σ_p^2 s representing small and large variances were obtained accordingly by plugging the two δ s into variance formulation. Hence, a total of twelve scenarios were defined in the simulation study by the six μ_p s and two σ_p^2 s.

Step 3. The number of events X_j was generated from the binomial distribution (5) with n_j and p_j from Steps 1 and 2.

3.2 Reported summary statistics

We report the relative bias of the estimated mean proportion $\left(\frac{\hat{\mu}_p - \mu_p}{\mu_p} \times 100\%\right)$ over 2000

Monte Carlo replications, where $\hat{\mu}_p = \frac{1}{2000} \sum_{j=1}^{2000} \hat{\mu}_{jp}$ and $\hat{\mu}_{jp}$ is the estimated mean proportion from the j th replication. The mean squared error $MSE(\hat{\mu}_p)$ was estimated using

$\frac{1}{2000} \sum_{j=1}^{2000} (\hat{\mu}_{jp} - \mu_p)^2$ and compared through the ratio of MSE for B-B to MSE for N-B. To

assess the precision of $\hat{\mu}_p$, the empirical standard deviation $\left(\sqrt{\frac{1}{2000-1} \sum_{j=1}^{2000} (\hat{\mu}_{jp} - \hat{\mu}_p)^2}\right)$ was compared with the means of the model-based standard error (SE-M) and the sandwich estimator of standard error (SE-S) over the 2000 simulations. The comparison was represented in terms of the relative change in SE-M and SE-S with respect to the empirical

standard deviation (ESD), i.e., $R - SE - M = \left(\frac{SE - M - ESD}{ESD} \times 100\%\right)$ and

$R - SE - S = \left(\frac{SE - S - ESD}{ESD} \times 100\%\right)$. We calculated the 95% confidence interval (CI) of $\hat{\mu}_{jp}$ using $\hat{\mu}_{jp} \pm t_{0.025, (K-1)} \sqrt{\text{Var}(\hat{\mu}_{jp})}$, where $t_{0.025, (K-1)}$ denotes the 2.5th percentile of a t -distribution with $(K-1)$ degrees of freedom. The proportions of CIs that cover the true μ_p are also reported.

3.3 Simulation result

3.3.1 Small number of studies, small mean within-study sample size—Table 1 shows the simulation results when there were 25 studies per meta-analysis and the mean(SD) within-study sample size was 50(30).

When the mean proportion $\mu_p = 5\%$ (scenarios 1–8), N-B tended to have larger relative bias and MSE (except for scenario 8) than B-B (e.g., for scenario 2, relative bias: 103% for N-B vs. 19% for B-B; the ratio of MSEs: $MSE(B-B)/MSE(N-B) = 17\%$). For each $\mu_p < 5\%$, when σ_p^2 increased, an increasing trend was observed for N-B in the relative bias, the ratio of MSEs (N-B/B-B), and the relative change in SE-M. Large SE-Ms relative to ESDs were also observed for B-B but with much smaller magnitudes than for N-B. The low estimation precision induced by the inflated SE-Ms caused high coverage rate for both models (e.g., scenario 2: $R-SE-M(B-B) = 37\%$ and $CR-M(B-B) = 96\%$; $R-SE-M(N-B) = 65\%$ and $CR-M(N-B) = 98\%$). When μ_p increased to 5%, as σ_p^2 increased, both models tended to have small SE-Ms relative to ESDs, resulting in low coverage rates (scenario 8: $R-SE-M(B-B) = -10\%$ and $CR-M(B-B) = 83\%$; $R-SE-M(N-B) = -2\%$ and $CR-M(N-B) = 79\%$).

When the mean proportion $\mu_p > 5\%$ (scenarios 9–12), the bias (except for scenario 9) and MSE for B-B were larger than for N-B. The SE-Ms for both models tended to be larger than ESDs when σ_p^2 was small, leading to high coverage rates. The opposite was found as σ_p^2 increased (e.g., scenarios 10, 12).

In all scenarios, the use of sandwich SEs reduced the differences between the model-based SEs and ESDs, implying improved estimation precision of $\hat{\mu}_p$. For example, the relative difference in SE for scenario 2 decreased from 37% (R-SE-M) to 9% (R-SE-S) for B-B and from 65% to –29% for N-B. These different SEs led to different coverage rates. Specifically, the coverage rate would decrease (increase) if R-SE-S was smaller (larger) than R-SE-M. In addition, it was noted that for scenarios 2 and 4 where N-B had large bias and high R-SE-M, the coverage rate for N-B dropped significantly with the use of sandwich standard error (e.g., scenario 4: CR-M= 95% vs. CR-S= 68%).

3.3.2 Small number of studies, large mean within-study sample size—Table 2 shows the simulation results when there were 25 studies per meta-analysis and the mean(SD) within-study sample size was 3000(5000).

When the mean proportion $\mu_p = 5\%$, B-B tended to have smaller relative bias than N-B. For small mean proportions (i.e., $\mu_p = 0.5\%$ or 1%), the MSEs for B-B were smaller than for N-B. As μ_p increased, B-B had smaller MSEs when σ_p^2 was small (scenarios 5 and 7). For those scenarios (6 and 8) where σ_p^2 s were large and N-B had smaller MSEs than B-B, large negative biases and very low coverage rates were also observed for N-B. The low coverage rates for N-B were due not only to the small SE-Ms but also to the underestimated μ_p s. When the mean proportion $\mu_p > 5\%$, B-B tended to have larger relative bias and MSE than N-B (except for scenario 9).

The use of sandwich SE improved estimation precision of $\hat{\mu}_p$ for both models in all scenarios. Especially for those scenarios (e.g., 2, 4, 6, 8, 10, 12) where B-B and N-B had large negative R-SE-Ms, the corresponding R-SE-Ss increased significantly, leading to higher coverage rates. For example, the sandwich standard errors increased the coverage rates for scenario 6 from 61% to 68% for N-B and from 75% to 88% for B-B.

3.3.3 Large number of studies—As expected, increasing the number of studies per meta-analysis from $K = 25$ to $K = 50$ decreased the MSE and SE for both models. For both small and large mean within-study sample sizes, the conclusions for comparisons between B-B and N-B remained and the differences between the two models became more marked than those shown in Tables 1 and 2. Additional information for the simulation results will be made available by the first author upon request.

4. Applications

We apply the B-B and N-B models to the two motivating examples described in Introduction, one with small and the other with large mean within-study sample size.

4.1 Example 1. A meta-analysis of complications after patello-femoral arthroplasty in the treatment of isolated patello-femoral osteoarthritis

Patello-femoral arthroplasty (PFA) is a successful treatment for isolated patello-femoral osteoarthritis, but there are concerns about post-treatment complications. A meta-analysis was performed to assess the incidence of a set of complications following PFA (Dy et al. 2012). For illustrative purposes, we considered two outcomes: revision and “other complications” (excluding mechanical failure, pain, and progression of osteoarthritis). The data presented in Table 3 contain the frequency of the two outcomes from 23 independent studies. The incidence of an outcome is defined by the ratio of the number of events (e.g., revision) to the total number of operated knees. The mean within-study sample size was small (50).

Table 4 shows the pooled incidence estimate $\hat{\mu}_p$ along with model-based standard error, sandwich standard error, and 95% CI. Descriptive statistics including sample mean $\bar{p}$ and variance s_p^2 of the incidence across all studies are also reported. Low incidences were found for both revision and “other complications” ($\hat{\mu}_p < 5\%$). For these two outcomes, the pooled incidence estimate $\hat{\mu}_p$ tended to be slightly higher for N-B than for B-B. N-B was associated with higher SEs and wider confidence intervals than B-B. For example, the large SE of the incidence of revision for N-B (N-B: 0.016 vs B-B: 0.013) led to a 24% wider CI than for B-B. The sandwich SEs were smaller than the model-based SEs, leading to narrower CIs. For example, the width of CI-S for “other complications” was 34% and 17% less than that of CI-M for N-B and B-B, respectively. The change in the width of the CIs was more prominent for N-B since difference between SE-M and SE-S for N-B was larger than for B-B. These results agreed with the simulation study where we found that N-B was associated with larger estimate of μ_p and higher SE when the proportion was small (e.g., Table 1, scenario 4).

4.2 Example 2. A meta-analysis of adverse drug reaction

Drug-related adverse events, including adverse drug reactions (ADRs), have been among the leading causes of morbidity and mortality (Lazaros et al. 1998, de Vries et al. 2008). According to the World Health Organization, ADRs are responsible for a substantial portion of health care costs in many countries. A meta-analysis (Hakkarainen et al. 2012) was recently conducted to estimate the percentage of patients with preventable ADRs (PADRs). In this example, we consider the proportion of adult outpatients with PADRs. The data set in Table 5 includes the number of healthcare visits (hospitalizations or emergency care visits) and the number of healthcare visits with PADRs from 16 independent studies. These studies had large mean within-study sample size (3000).

The results are summarized in Table 6. The sample mean incidence of death (3.4%) and the variance (7×10^{-4}) were very low in the studies comprising this meta-analysis. The pooled proportion estimate of PADRs was 16% higher for N-B than for B-B. N-B was associated with higher SEs (SE-M: 63% higher, SE-S: 33% higher) and wider confidence intervals than B-B. Compared to the SE-M, the SE-S built on the sandwich estimator was smaller, resulting in 37% and 19% narrower CIs for N-B and B-B, respectively. These results agreed with our findings that N-B had larger estimate of μ_p and SE than B-B in a similar scenario of the simulation study (Table 2, scenario 5).

5. Discussion

Two exact likelihood methods, B-B and N-B, for meta-analysis of proportions were discussed and compared in this study. Both methods utilize the binomial distribution to model within-study variation. For between-study variation, N-B assumes the proportions on a logit scale follow a normal distribution, while B-B assumes a beta distribution on the original scale. The sandwich estimator of variance was integrated into the two models and was expected to increase their robustness.

Through an extensive simulation study, we were able to compare B-B and N-B for different proportions, variations, and sample sizes. We demonstrated that the B-B approach tended to have smaller bias, MSE, and SE than N-B for small proportions ($\mu_p \leq 5\%$) and N-B outperformed B-B for $\mu_p > 5\%$. When assessing the numerical robustness of the two methods, B-B had a less number of non-converged simulation runs (1% of 1000 runs) compared to N-B (15% of 1000 runs). The use of sandwich SEs reduced the differences between the ESDs and model-based SEs for both models, leading to improved estimation precision. Therefore, the sandwich estimator is recommended for meta-analysis of proportions when using B-B or N-B.

Different studies often have different designs and patient characteristics such as patient age, proportion of female patients, and proportion of patients that have a certain comorbidity. Like other random effects models, B-B and N-B can readily incorporate these study-specific attributes by expanding the mean μ of the beta distribution (7) and μ_0 of the normal distribution (9), respectively, with study-level covariates. Incorporating study-level covariates allows us to explore and explain between-study heterogeneity by specific factors. Although meta-regression can be readily implemented, a systematic investigation and comparison of the two models in a meta-regression setting is needed. In addition, adding study-level covariates can be challenging in meta-analysis of proportions as the models may not converge for rare events. Statistical analysis may also be hampered by missing data when some study-level covariates are not available in all studies. Furthermore, meta-regression always requires careful consideration, as additional biases are introduced by including covariates from different studies, possibly leading to spurious results (Sutton et al. 2000). These issues are beyond the scope of the current study and will be considered in future research.

Acknowledgements

The following grants supported this study: Agency for Healthcare Research and Quality's health services research grant (AHRQ R01HS021734) and Clinical Translational Science Center (UL1-RR024996). We sincerely thank the anonymous reviewers for their constructive critiques that help improve this manuscript immensely.

Appendix

In this appendix the SAS code is given to perform meta-analysis using B-B and N-B models through PROC NLMIXED.

```
***SAS code for Beta-Binomial model***;
```

```

proc nlmixed data=sasdata df=1000 gtol=1e-10;
parms mu=0.005 gama=0.02; **initial value**;
A=mu*(1-gama)/gama;
B=(1-mu)*(1-gama)/gama;
loglike=(lgamma(n+1)-lgamma(event+1) -lgamma(n-event+1))+lgamma(A+event)
+lgamma(n+Bevent)+lgamma(A+B)-lgamma(A+B+n)-lgamma(A)-lgamma(B); ***log
likelihood function of betabinomial***;
model event~general(loglike); ***event=the number of events, n=sample
size***;
estimate "mu(B-B) " A/(A+B); **estimate of mean proportion mu(B-B)**;
run;

***A specially prepared SAS code that can calculate the sandwich
estimator for Beta-Binomial model (invoke a RANDOM statement but skip
estimation of the random effect variance tau*tau and keep tau fixed at 0)***
proc nlmixed data=sasdata empirical;
parms gama=0.1 Intercept=1.14 tau=0;
bounds 0:00001 <=gama<= 0:99999;
bounds 0 <=tau<= 0;
eta = Intercept + u;
mu= exp(eta)/ (1 + exp(eta));
varianz=(1-gama)/gama;
A=mu*(1-gama)/gama;
B=(1-mu)*(1-gama)/gama;
ll= lgamma(n+1)+lgamma(event+A)+lgamma(n-event+B)+lgamma(A+B) -
lgamma(event+1)-lgamma(n-event+1)-lgamma(n+A+B) -lgamma(A)-lgamma(B);
model y ~ general(ll);
random u ~ normal(0,tau*tau) subject=study;
run;

***SAS code for Normal-Binomial model***;
***calculate the mean proportion of N-B***;
%macro plogit;
Pexact = 0
%do j=-50 %to 50;
+ sqrt(tau2)/10 * 1/(1+exp(-mu0 - sqrt(tau2)*j/10))*pdf('normal',
sqrt(tau2)*j/10, 0, sqrt(tau2))
%end;
;
%mend plogit;

proc nlmixed data=sasdata df=1000 gtol=1e-10 EMPIRICAL; ***EMPIRICAL: to
obtain sandwich estimator of of variance ***;
parms mu0=-6.5 tau2=2;
bounds tau2>0;
beta=mu0+u;
pred=exp(beta)/(1+exp(beta));

```

```

pai=constant("pi");
model event~binomial(n, pred);
random u ~normal(0, tau2) subject=study;
%plogit;
estimate "mu(N-B)" Pexact;**estimate of mean proportion mu(N-B)**;
run;

```

References

- Bhaumik DK, Amatya A, Normand ST, Greenhouse J, Kaizar E, Neelon B, Gibbons RD. Meta-analysis of rare binary adverse event data. *Journal of the American Statistical Association*. 2012; 107:555–567. [PubMed: 23734068]
- Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Med Decis Making*. 2010; 30:499–508. [PubMed: 19959794]
- Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for absolute risk difference and relative risk. *Statistical Methods in Medical Research*. 2010; 21:621–633. [PubMed: 21177306]
- Craigmile PF, Titterton DM. Parameter estimation for finite mixtures of uniform distributions. *Communications in Statistics-Theory and Methods*. 1997; 26:1981–1995.
- Cushner F, Agnelli G, Fitzgerald G, Warwick D. Complications and functional outcomes after total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). *Am J Orthop*. 2010; 39:22–28. [PubMed: 21290028]
- de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: A systematic review. *Qual Saf Health Care*. 2008; 17:216–223. [PubMed: 18519629]
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986; 7:177–188. [PubMed: 3802833]
- Dy CJ, Franco N, Ma Y, Mazumdar M, McCarthy MM, Gonzalez Della Valle A. Complications after patello-femoral versus total knee replacement in the treatment of isolated patello-femoral osteoarthritis: a meta-analysis. *Knee Surg Sports Traumatol Arthrosc*. 2012; 20:2174–2190. [PubMed: 21987361]
- Fay MP, Graubard BI, Freedman LS, Midthune DN. Conditional logistic regression with sandwich estimators: application to a meta-analysis. *Biometrics*. 1998; 54:195–208. [PubMed: 9544516]
- Griffiths DA. Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*. 1973; 29:637–648. [PubMed: 4785230]
- Hakkarainen KM, Hedna K, Petzold M, Hagg S. Percentage of patients with preventable adverse drug reactions and preventability of adverse drug reactions: a meta-analysis. *PLoS ONE*. 2012; 7:e33236. [PubMed: 22438900]
- Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology*. 2008; 61:41–51. [PubMed: 18083461]
- Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*. 1996; 15:619–629. [PubMed: 8731004]
- Kuss O, Gromann C. An exact test for meta-analysis with binary endpoints. *Methods Inf Med*. 2007; 46:662–668. [PubMed: 18066416]
- Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Statistics in Medicine*. 2013 [Epub ahead of print].
- Lazaros J, Pomeranz BH, Corey PH. Incidence of adverse drug reactions in hospitalised patients: a meta-analysis of prospective studies. *JAMA*. 1998; 15:1200–1205.

- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*. 1989; 84:1074–1078.
- Lindstrom MJ, Bates DM. Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics*. 1990; 46:673–687. [PubMed: 2242409]
- Muller M, Wandel S, Colebunders R, Attia S, Furrer H, Egger M. IeDEA Southern and Central Africa. Immune reconstitution inflammatory syndrome in patients starting antiretroviral therapy for HIV infection: a systematic review and meta-analysis. *Lancet Infect Dis*. 2010; 10:251–261. [PubMed: 20334848]
- Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Statistics in Medicine*. 1999; 18:643–654. [PubMed: 10204195]
- R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- Raudenbush SW, Bryk AS. Empirical bayes meta-analysis. *Journal of Educational Statistics*. 1985; 10:75–98.
- SAS Institute Inc.. SAS/STAT 9.2 User's guide. Cary, NC: SAS Institute Inc.; 2008.
- Shuster JJ, Jones LS, Salmon DA. Fixed vs random effects meta-analysis in rare event studies: the Rosiglitazone link with myocardial infarction and cardiac death. *Statistics in Medicine*. 2007; 26:4375–4385. [PubMed: 17768699]
- Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*. 2010; 29:3046–3067. [PubMed: 20827667]
- Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*. 2004; 23:1351–1375. [PubMed: 15116347]
- Sutton, AJ.; Abrams, KR.; Jones, DR.; Sheldon, TA.; Song, F. *Methods for Meta-Analysis in Medical Research*. New York: Wiley; 2000.
- Warycha, Zakrzewski, Ni, Shapiro, Berman, Pavlick, Polsky, Mazumdar, Osman. Meta-analysis of sentinel lymph node positivity in thin melanoma. *Cancer*. 2009; 115:869–879. [PubMed: 19117354]
- White H. Maximum likelihood estimation of misspecified models. *Econometrica*. 1982; 50:1–26.
- Young-Xu Y, Chan KA. Pooling overdispersed binomial data to estimate event rate. *BMC Medical Research Methodology*. 2008; 8:58. [PubMed: 18713448]

Figure 1.
Empirical PDF (histogram) and model-based estimated PDF (dashed line) of proportion.

Table 1

Simulation results over 2000 Monte Carlo replications when $K = 25$ studies per meta-analysis and mean(SD) within-study sample size= 50(30). The results include the following summary statistics for the pooled proportion estimate $\hat{\mu}_p$: relative bias(RB), ratio of mean squared errors(R-MSE), relative change for model-based standard error (R-SE-M) with respect to empirical standard deviation (ESD), relative change for sandwich standard error (R-SE-S) with respect to ESD, coverage rate of 95% CI using model-based standard error (CR-M), coverage rate of 95% CI using sandwich standard error (CR-S).

Scenario	Model	RB(%)	R-MSE	R-SE-M(%)	R-SE-S(%)	CR-M(%)	CR-S(%)
(1) $\mu_p = 0.005$	B-B	11.049	0.798	20.771	16.092	97.936	97.662
	N-B	17.009	–	39.429	14.981	98.452	98.245
$\sigma_p^2 = 0.013 \times 10^{-3}$	B-B	18.580	0.173	36.681	9.096	96.242	94.394
	N-B	103.010	–	64.789	–29.055	97.707	80.063
(3) $\mu_p = 0.01$	B-B	4.524	0.983	9.654	6.706	95.313	95.240
	N-B	5.650	–	14.498	8.211	95.767	95.616
$\sigma_p^2 = 0.053 \times 10^{-3}$	B-B	7.043	0.198	27.609	1.953	91.134	88.888
	N-B	81.895	–	51.768	–22.677	95.394	67.933
(5) $\mu_p = 0.03$	B-B	0.608	0.984	1.486	–0.776	94.940	94.456
	N-B	1.286	–	4.792	–0.632	95.371	94.886
$\sigma_p^2 = 0.479 \times 10^{-3}$	B-B	–0.454	0.750	8.048	–3.801	84.780	84.244
	N-B	–1.807	–	9.088	–7.486	87.781	82.797
(7) $\mu_p = 0.05$	B-B	0.260	0.988	0.069	–2.316	94.132	93.618
	N-B	0.962	–	4.330	–2.128	94.801	93.875
$\sigma_p^2 = 0.001$	B-B	3.831	1.186	–9.672	–8.905	83.403	84.193
	N-B	–	–	–	–	–	–

Scenario	Model	RB(%)	R-MSE	R-SE-M(%)	R-SE-S(%)	CR-M(%)	CR-S(%)
$\sigma_p^2=0.025$	N-B	-12.994	-	-1.636	0.822	78.872	78.872
(9) $\mu_p=0.1$	B-B	-0.340	1.003	4.744	-1.372	94.997	94.087
$\sigma_p^2=0.004$	N-B	0.677	-	10.321	-1.221	95.704	94.441
(10) $\mu_p=0.1$	B-B	9.515	1.567	-33.879	-12.569	79.899	83.919
$\sigma_p^2=0.025$	N-B	-0.614	-	-21.220	-6.969	82.261	83.266
(11) $\mu_p=0.25$	B-B	-1.595	1.014	4.724	-1.755	95.582	94.578
$\sigma_p^2=0.021$	N-B	-0.899	-	9.267	-1.732	96.084	94.929
(12) $\mu_p=0.25$	B-B	1.488	1.083	-9.499	-4.433	93.129	93.179
$\sigma_p^2=0.035$	N-B	0.751	-	-3.395	-3.260	93.530	94.533

Table 2

Simulation results over 2000 Monte Carlo replications when $K = 25$ studies per meta-analysis and mean(SD) within-study sample size= 3000(5000). The results include the following summary statistics for the pooled proportion estimate $\hat{\mu}_p$: relative bias(RB), ratio of mean squared errors(R-MSE), relative change for model-based standard error (R-SE-M) with respect to empirical standard deviation(ESD), relative change for sandwich standard error (R-SE-S) with respect to ESD, coverage rate of 95% CI using model-based standard error (CR-M), coverage rate of 95% CI using sandwich standard error (CR-S).

Scenario	Model	RB(%)	R-MSE	R-SE-M(%)	R-SE-S(%)	CR-M(%)	CR-S(%)
(1) $\mu_p = 0.005$	B-B	-0.048	0.838	3.141	-2.990	94.663	92.797
	N-B	4.365	-	22.885	-1.648	97.098	94.145
$\sigma_p^2 = 0.013 \times 10^{-3}$	B-B	1.066	0.603	-21.983	-9.496	76.541	82.497
	N-B	-7.940	-	-13.786	-5.615	67.032	74.294
(3) $\mu_p = 0.01$	B-B	0.065	0.747	4.293	-3.011	94.892	93.360
	N-B	5.894	-	26.555	-2.528	97.701	94.637
$\sigma_p^2 = 0.053 \times 10^{-3}$	B-B	1.295	0.923	-18.869	-10.912	84.936	91.790
	N-B	-26.854	-	-22.016	-5.241	59.948	66.649
(5) $\mu_p = 0.03$	B-B	-0.168	0.737	5.426	-2.496	95.570	93.380
	N-B	6.597	-	33.314	-0.662	98.930	95.773
$\sigma_p^2 = 0.479 \times 10^{-3}$	B-B	7.457	1.237	-17.744	-9.921	74.595	87.983
	N-B	-40.390	-	-31.240	-15.246	60.759	68.070
(7) $\mu_p = 0.05$	B-B	-0.412	0.779	6.172	-1.354	95.608	94.144
	N-B	5.966	-	34.565	0.268	98.435	95.961
$\sigma_p^2 = 0.001$	B-B	5.425	1.317	-17.822	-11.746	78.746	87.912
	N-B	-	-	-	-	-	-

Scenario	Model	RB(%)	R-MSE	R-SE-M(%)	R-SE-S(%)	CR-M(%)	CR-S(%)
$\sigma_p^2=0.025$	N-B	-46.359	-	-28.996	-18.760	52.838	64.614
(9) $\mu_p=0.1$	B-B	-0.921	0.856	12.957	-1.470	96.415	94.497
$\sigma_p^2=0.004$	N-B	4.372	-	41.001	-0.429	98.738	94.447
(10) $\mu_p=0.1$	B-B	9.318	1.800	-31.810	-15.227	74.237	82.298
$\sigma_p^2=0.025$	N-B	-1.805	-	-23.034	-8.549	81.372	82.026
(11) $\mu_p=0.25$	B-B	-2.296	1.063	8.842	-2.249	95.965	94.402
$\sigma_p^2=0.021$	N-B	-0.894	-	24.055	-1.271	97.529	95.158
(12) $\mu_p=0.25$	B-B	1.966	1.143	-11.909	-6.490	92.222	91.969
$\sigma_p^2=0.035$	N-B	1.403	-	1.256	-4.247	95.959	93.333

Table 3

Data of example 1: complications after patello-femoral arthroplasty in the treatment of isolated patello-femoral osteoarthritis

Study	n(knees)	n(revision)	n(other complications)
1	79	0	2
2	20	0	0
3	30	0	0
4	109	0	0
5	25	0	0
6	66	7	3
7	59	0	0
8	16	0	0
9	30	3	0
10	56	7	0
11	45	2	0
12	26	0	0
13	76	10	3
14	45	3	0
15	16	1	2
16	25	2	0
17	51	1	0
18	85	11	0
19	122	6	0
20	22	1	0
21	50	0	0
22	55	0	0
23	46	0	1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Meta-analysis results of example 1: complications after patello-femoral arthroplasty in the treatment of isolated patello-femoral osteoarthritis. The results include the pooled proportion estimate $\hat{\mu}_p$ along with model-based standard error(SE-M), sandwich standard error(SE-S), 95%CI using model-based standard error(CI-M), and 95%CI constructed from the sandwich standard error(CI-S). The relative difference in width of the 95% CIs R-CI=(width (95%CI-S)-width (95%CI-M))/width (95%CI-M). The sample mean $\bar{p}$ and variance s_p^2 of the proportions across all studies are also reported.

Outcome	Model	$\hat{\mu}_p$	SE-M	SE-S	95% CI-M	95% CI-S	R-CI (%)
Revision							
$\bar{p}=0.042$ $s_p^2=0.002$	B-B	0.043	0.013	0.01	(0.016, 0.069)	(0.021, 0.065)	-17.0
	N-B	0.046	0.016	0.01	(0.014, 0.079)	(0.025, 0.068)	-33.9
Other complications							
$\bar{p}=0.011$ $s_p^2=0.001$	B-B	0.01	0.005	0.004	(-0.001, 0.021)	(0.001, 0.019)	-18.2
	N-B	0.012	0.008	0.005	(-0.004, 0.028)	(0.001, 0.022)	-34.4

Table 5

Data of example 2: proportion of patients with preventable adverse drug reactions (PADRs)

Study	$n(\text{visits})$	$n(\text{PADRs})$
1	10587	8
2	150	14
3	956	10
4	253	17
5	240	21
6	671	24
7	915	42
8	844	28
9	18820	880
10	6899	158
11	548	30
12	1756	78
13	1101	25
14	1802	28
15	2238	35
16	1017	19

Table 6

Meta-analysis results of example 2: proportion of patients with preventable adverse drug reactions (PADRs). The results include the pooled proportion estimate $\hat{\mu}_p$ along with model-based standard error(SE-M), sandwich standard error(SE-S), 95% CI using model-based standard error(CI-M), and 95% CI constructed from the sandwich standard error(CI-S). The relative difference in width of the 95% CIs: R-CI=(width (95%CI-S)-width (95%CI-M))/width (95%CI-M). The sample mean $\bar{p}$ and variance s_p^2 of the proportions across all studies are also reported.

Outcome	Model	$\hat{\mu}_p$	SE-M	SE-S	95% CI-M	95% CI-S	R-CI (%)
PADR							
$\bar{p}=0.034$	B-B	0.037	0.008	0.006	(0.021, 0.053)	(0.024, 0.050)	-18.8
$s_p^2=0.0007$	N-B	0.043	0.013	0.008	(0.016, 0.070)	(0.026, 0.060)	-37.0