Meta-analysis of Rare Adverse Events in Randomized Controlled Clinical Trials: a Case Study

Eunhee Hwang, Suna Barlas, Lingfeng Yang
Pfizer Inc., 500 Arcola Road, Collegeville, PA 19454

Abstract

Meta-analysis of randomized controlled clinical trials are often used to evaluate the effect of a new drug entity on rare adverse events as safety characterization plays an important role in the overall benefit risk assessment. In this case study, we apply commonly used meta-analysis techniques to our clinical trial data: combining summary statistics across trials via variance inflation or Newcombe method, Bayesian fixed/random effect model, generalized linear model, and generalized linear mixed model. We examine the impact of explicit and implicit assumptions in each technique and provide intuitive interpretation for some surprising results. Our choice of analysis method and metrics for the data are explained considering the issue of continuity correction, large sample approximation, bias, coverage probability and the maximum use of all available data. In this presentation, we limit our focus on the binomial outcomes without adjusting for the exposure.

Key Words: Meta-analysis, Rare events, Case study, Continuity correction, Coverage probability, Binomial outcome

1. Background

Safety characterization plays an essential role in assessing the overall benefit risk of a new drug entity. Increasingly, the scope of safety characterization includes evaluating risks for a series of rare serious adverse events. Meta-analyses techniques are useful in this setting to examine relevant data across the entire clinical program while adjusting for trial to trial variability. A wealth of publications has pointed out pitfalls in applying conventional meta-analysis techniques to a rare event setting -- the low coverage probability in normal approximation and the impact of continuity correction on the analysis result [1, 5, 7, 10]. Also controversial is an often-encountered practice of discarding trials with no observed events from meta-analyses [2, 10].

In this case study based on our clinical data, we quantify the risk for rare safety events associated with a new treatment compared with placebo by applying various meta-analysis methods that were suggested as remedies for conventional methods. Our choice of summary metrics and the overall interpretation of various analysis results are guided in consideration of bias, coverage probability and the maximum use of all available data.

2. Introduction

The meta-analysis data set for this case study is presented in Table 1. Number of subjects who experienced adverse events of interest is summarized by treatment group from the five randomized, double-blind, placebo-controlled, parallel-group studies in this clinical development program.

Table 1. Meta-analysis data from 5 eligible clinical trials

Study	Treatment Duration (Week)	# Subject with Events/ Safety Population (Rate per 1,000 Subjects)	
		Active Treatment	Placebo
1	104	6/2547 (2.4)	0/427
2	52	1/710 (1.4)	0/172
3	12	0/255	0/63
4	12	0/437	0/105
5	52	1/919 (1.1)	0/474

The following observations provide useful insights on the data. First, the observed event rates are indeed very low (<<1%) for both treatment arms across studies indicating that the summary based on the large sample theory would yield an inadequately low coverage probability [1, 5, 7, 10].

Second, the event rate in Study 1, while still very low in absolute sense, is higher compared with those in other studies. The analysis should therefore either be stratified by study or include this effect in the model [4].

Third, the number of subjects between the two arms is not balanced within each study, mainly due to more than 1 dose groups, with the most extreme ratio of about 6:1 for active treatment vs. placebo in Study 1. Applying a constant continuity correction in this unbalanced and rare event setting results in bias penalizing the placebo group [10].

Fourth, note the absence of any events in the placebo arm across all studies with Studies 3 and 4 having no event in the active treatment arm, either. Some authors exclude trials with zero events from analyses based on the fact that these do not provide any information on the risk ratio or the odds ratio. On the other hand, Cai et al. [3] reported that this practice might lead to a bias. We choose the metric of rate differences instead of ratios for this case study. The advantages include the following: utilization of all available data, not having to use continuity correction, and its applicability in benefit/risk assessment [4].

Lastly but not least, also observed is the huge variation in the treatment duration across studies ranging from 12 weeks to 2 years. While out of scope for this publication, exposure-adjusted methods complement shortcomings of analyzing binary outcomes.

3. Methods Applied

This section provides brief introduction of each meta-analyses method we applied to our case study data. We applied two variations of methods that combine summary statistics from each study. We applied the generalized linear models assuming both fixed and random treatment differences that were fit using SAS® PROC GENMOD or NLMIXED [8]. In addition to this frequentist approach, we also tried Bayesian fixed/random effect models that were fit via WinBUGS.

3.1 Combining Results Across Studies: Variance Inflation and Extended Newcombe Method

In this section, we presents two methods that directly combine results from individual studies. Both methods use the Mantel-Haenszel weight when combining studies to prevent studies with smaller point estimates from having undue larger weights. Both methods also estimate the rate difference within each study using the observed proportions without continuity correction. The difference lies in the way the variance is calculated within each study.

We first introduce the variance inflation method which is a simple modification of the Wald method. Let π_{i1} and π_{i2} be the true event rate for active treatment and placebo, respectively, in study *i*. Let p_{i1} and p_{i2} be the observed rates x_{i1}/n_{i1} and x_{i2}/n_{i2} where x_{ij} and n_{ij} represent the number of events and total for treatment *j*, respectively. The Wald method utilizes the large sample theory to derive a confidence interval (CI) based on the following estimates for the rate difference and its variance in each study:

$$d_i = p_{i1} - p_{i2} \text{ and}$$
$$\text{var}\,(d_i) = p_{i1}(1-p_{i1})/n_{i1} + p_{i2}(1-p_{i2})/n_{i2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

The most common meta-analysis based on Wald method combines the results across studies by weighing each study according to the inverse of its estimated variance in (1).

Koch suggested the following modification. The variance of d_i is obtained after slightly over-estimating the variance for each proportion:

$$\text{var}\,(p_{ij}) = \frac{q_{ij}(1-q_{ij})}{n_{ij}-1}, \text{ where } q_{ij} = \frac{x_{ij}+0.5}{n_{ij}+1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Then, combine the results across studies using the Mantel-Haenszel weight:

$$w_i = (1/n_{i1} + 1/n_{i2})^{-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

The modification (2) circumvents the zero variance estimate for treatments with no event while alleviating the problem of low coverage probability of the Wald method. Note that the rates are estimated without any continuity correction thus not penalizing the treatment with smaller *n*. Further, the modification (3) improves the overall estimates by preventing studies with low event rates from being heavily weighted.

The second method is the extension of the Newcombe method. Newcombe [7] extended Wilson score method for the interval estimation of a single proportion into two sample cases. Wilson score method derives the CI for π_{ij} by solving the quadratic equation (4) instead of plugging in the observed proportion p_{ij} on the right hand side as in (1).

$$|\pi_{ij} - p_{ij}| \le z\sqrt{\frac{\pi_{ij*}(1-\pi_{ij})}{n_{ij}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

Newcombe proposed the following CI of (L_i, U_i) for the rate difference d_i after obtaining each CIs of $(l_{i1},\ u_{i1})$ and $(l_{i2},\ u_{i2})$ for π_{i1} and π_{i2} based on (4).

$$L_i = d_i - \text{int}_{i,l},\ U_i = d_i + \text{int}_{i,u},$$
$$\text{where}\quad \text{int}_{i,l} = \sqrt{(p_{i1}-l_{i1})^2 + (u_{i2}-p_{i2})^2}, \text{ and}$$

$$\text{int}_{i,u} = \sqrt{(u_{i1} - p_{i1})^2 + (p_{i2} - l_{i2})^2} \quad \text{.....(5)}$$

The lower interval, int_{il}, combines the smallest possible departure from p_{i1} and the largest one from p_{i2}. The upper interval is similarly constructed. This interval, along with the intuitive interpretation and other advantages, provides an improved coverage probability over the Wald method [5, 7].

We further extend this method to a meta-analysis setting utilizing that both $\text{int}_{i,l}$ and $\text{int}_{i,u}$ in (5) can be expressed as a separate multiple of the *z* score: for each study *i*,

$$\text{int}_{i,l} = z * \text{pseudo standard error}_{i,l}, \text{ and}$$
$$\text{int}_{i,u} = z * \text{pseudo standard error}_{i,u} \quad \text{.....(6)}$$

We combine results across studies using the Mantel-Hanszel weight in (3). The overall estimated difference *D* is obtained as $D = \Sigma w_i d_i / \Sigma w_i$ with the CI of $(D - L_D,\ D + U_D)$ where $L_D = z * sqrt\ \{\Sigma\left[{w_i}^2 * \ (pseudo\ se_{i,l)})^2\right]/(\Sigma \text{w}_\text{i})^2\ \}$ and U_D is similarly defined.

3.2 Bayesian Fixed/Random Effect Model

The two methods in Section 3.1 assume a fixed treatment difference δ between the active treatment and the placebo across all studies by setting

$$\delta = \pi_{i1} - \pi_{i2} \quad \text{.....(7)}$$

Bayesian fixed effect model assumes that this δ has a statistical distribution. We adopt a model suggested by Warn et al. [11] where a non-informative prior distribution is assigned for δ as well as for the placebo rate of π_{i2}:

$$\delta \sim \text{Uniform (-1, 1) and } \pi_{i2} \sim \text{Uniform (0, 1)} \quad \text{.....(8)}$$

These result in the following rate for the active treatment:

$$\pi_{i1} = \pi_{i2} + \min\ (\max(\delta, -\pi_{i2}), 1 - \pi_{i2}) \quad \text{.....(9)}$$

We also follow Warn et al. [11]'s suggestion for the Bayesian random effect model by setting the observed treatment difference d_i in (1) as a realization of random variable δ_i:

$$\delta_i \sim \text{N}(\delta, \tau^2) \quad \text{.....(10)}$$

The same priors in (8) for δ and π_{i2} are used; the resulting π_{i1} is similarly defined as in (9) with δ replaced by δ_i. Further, the prior distribution for τ is set as Uniform (0, 2).

3.3 Generalized Linear Fixed/Random Effect Model

We also apply the generalized linear fixed/random effect model to our data. For the generalized linear fixed effect model [6], we use the identity link to connect the binomial model parameter π_{ij} to each study and treatment as shown below. Note that this model results in the identical treatment difference across studies as in (7).

$$x_{ij} \sim \text{binomial}\ (n_{ij}, \pi_{ij}) \text{ with } \pi_{ij} = \text{study}_{\ i} + \text{treatment}_{\ j} \quad \text{.....(11)}$$

For the generalized linear random effect model, we modified Stijnen's approach. Stijnen [9] introduced random effects on the two rate parameters by assuming a bivariate normal distribution for a pair of (logit π_{i1} , logit π_{i2}). Given the rate parameter π_{ij}, the number of events $x_{ij}|\ \pi_{ij}$ follows an independent binomial distribution (n_{ij}, π_{ij}). As our choice of metric is the rate difference, we assumed a bivariate normal distribution directly

for the rate parameters (π_{i1}, π_{i2}). This leads to the same distributional assumption for the treatment difference as seen in (10).

Note an inherent difficulty of fitting the rate difference within the generalized linear model framework due to a boundary condition in the rate difference.

4. Analyses Results

4.1 Combining Results Across Studies: Variance Inflation and Extended Newcombe Method -- Results

Table 2 presents the meta-analyses results of the two methods that combine each individual study result. Both methods yield identical point estimates as both utilize Mantel-Haenszel weight without any continuity correction. The CIs in both methods are larger compared with those in Wald method, therefore improving the coverage probability. The CIs in the Newcombe method are in general wider suggesting that the quick and easy adjustment in the variance inflation method may not be satisfactory; Dann and Koch [5] showed that Newcombe method yielded CIs with adequate coverage probability. Note that the asymmetric CIs around the point estimates in Newcombe method reflect higher uncertainty in the placebo rates due to smaller numbers of samples compared with those in active treatment.

In addition to the meta-analyses of all five studies, we also conducted meta-analyses of studies with duration >=1 year that included studies 1, 2 and 5. These three studies also happened to be the studies with larger sample sizes. Within each method, the long term analysis slightly shifted the point estimate upward with a reduced length of CI interval, especially for the extended Newcombe method. We consider this observation desirable given relatively small sample sizes especially in the placebo arm in studies with <1 year duration.

Table 2. Meta-analysis using variance inflation method and extended Newcombe method

Study	# Subject with Events/ Safety Population (Rate per 1,000 Subjects)		Mean Excess Risk of Active Treatment per 1,000 Subjects over Placebo (90% CI)	
	Active Treatment	Placebo	Variance Inflation	Extended Newcombe
1	6/2547 (2.4)	0/427	2.4 (-0.8, 5.5)	2.4 (-4.0, 4.5)
2	1/710 (1.4)	0/172	1.4 (-5.9, 8.7)	1.4 (-14.1, 6.3)
3	0/255	0/63	0.0 (-18.9, 18.9)	0.0 (-41.2, 10.5)
4	0/437	0/105	0.0 (-11.4, 11.4)	0.0 (-25.1, 6.2)
5	1/919 (1.1)	0/474	1.1 (-2.2, 4.4)	1.1 (-4.6, 4.9)
All Studies	--	--	1.5 (-0.9, 3.9)	1.5 (-3.5, 3.3)
>= 1 year Studies (1, 2, 5)	--	--	1.7 (-0.6, 4.0)	1.7 (-2.8, 3.6)

4.2 Bayesian Fixed/Random Effect Model -- Results

Bayesian models introduced in Section 3.2 were fit in WinBUGS. Continuity correction had to be applied for the Markov Chain Monte Carlo simulation to update the samples. Given the severe imbalance in sample sizes between the treatment, we followed Sweeting et al. [10]'s suggestion that the correction be proportional to the sample size totaling to 1 within each study. For illustration purpose, we start with the random effect model analysis result shown in Table 3.

Table 3. Meta-analysis using Bayesian Random Effect Model

Study	# Subject with Events/ Safety Population (Rate per 1,000 Subjects)		Median Excess Risk of Active Treatment per 1,000 Subjects over Placebo (90% CI)	
	Active Treatment	Placebo	All 5 Studies	>= 1 Year Studies
1	6/2547 (2.4)	0/427	0.9 (-2.5, 3.1)	1.0 (-2.7, 3.4)
2	1/710 (1.4)	0/172	0.5 (-3.6, 3.6)	0.6 (-5.2, 4.4)
3	0/255	0/63	0.5 (-5.0, 5.0)	
4	0/437	0/105	0.3 (-5.1,3.3)	
5	1/919 (1.1)	0/474	0.7 (-2.5, 3.2)	0.6 (-3.3, 3.7)
All 5 Studies	--	--	0.5 (-3.2, 3.2)	
>= 1 Year Studies (1, 2, 5)				0.7 (-8.5, 8.9)

We first noticed that the estimated median excess risks were similar across all studies ranging from 0.3 to 0.9 per 1,000 subjects in the analysis of all studies, contrary to the wider range of observed excess risks that ranged from 0 to 2.4. This similarity was not due to the continuity correction as Sweeting et al. [10]'s suggestion was adopted. Rather, it was due to the benign non-informative prior distribution given to the rates. The uniform (0, 1) distribution assigned to the placebo rate π_{i2} in (8), together with the binomial distribution for the number of events in the placebo treatment [$x_{i1}|\pi_{i2} \sim \text{bin}(n_{i2}, \pi_{i2})$], results in the posterior distribution of $\pi_{i2}| x_{i2}$ with the mean as $(x_{i2} + 1) / (n_{i2} + 2)$. The impact of the prior distribution on the posterior mean for the rate for the active treatment rate, π_{i1}, was similar. The severe imbalance in sample sizes between the two arms resulted in the placebo group being unduly penalized, thus pulling the excess risk toward null.

As the uniform distribution belongs to the beta distribution family denoted as Beta (α, β), we varied prior distribution for π_{i2} with varying α while fixing β as 1. Table 4 shows that the point estimate for the median excess risk increases as α decreases with the model failing at α=0.05.

Table 4. Meta-analysis of All Studies with Varying Prior Distribution* for the Placebo Rate using Bayesian Random Effect Model

Excess Risk of Active Treatment per 1,000 Subjects over Placebo Median (90% CI)				
α=1	α=0.5	α=0.3	α=0.1	α=0.05
0.5 (-3.1, 3.2)	1.6 (-1.5, 4.3)	2.0 (-1.0, 4.4)	2.3 (0.5, 4.3)	Simulation fails

*: Beta (α, 1) distribution

Also noted were similar lengths of credible intervals across studies despite huge differences in sample sizes. This in trun results in a considerable difference in precision between the two meta-analyses of all studies vs. long studies. This seems to be due to the random effect assumption (10) where the true treatment differences $\delta_i's$ share the same inherence variance τ^2. In other words, the variability expressed in the credible interval for each study has two sources: one from the random effect assumption, another, a sample variation in estimating δ_i. The former forces the credible/confidence intervals to have similar lengths compared with those in the fixed effect model where only the latter is counted. With the eligible case studies varying so much in study sizes, we would like to voice concern applying the random effect model in addition to the inherent difficulty of discerning variability across studies in the rare event setting.

Table 5 presents the Bayesian random effect model and fixed effect model analysis side by side. The impact of the prior distribution on event rates is still shown in the fixed model expressed as the median excess risk pulled toward null. Note that, in the fixed effect model, the meta- analysis of all studies produced a similar length of credible interval compared with that for studies of >=1 year. This supports our previous conjecture that the huge difference shown in the Bayesian random effect model analyses is likely from the random effect assumption, not from the Bayesian assumption per se.

Table 5. Comparison of Bayesian Meta-analysis: Random versus Fixed Effect Model

Meta-analysis	Median Excess Risk of Active Treatment per 1,000 Subjects over Placebo (90% CI)	
	Random Effect	Fixed Effect
All 5 Studies	0.5 (-3.2, 3.2)	0.7 (-1.8, 2.4)
>= 1 Year Studies	0.7 (-8.5, 8.9)	0.9 (-1.7, 2.9)

4.3 Generalized Linear Fixed/Random Effect Model -- Results

The generalized linear fixed effect model in Section 3.3 was fit using PROC GENMOD in SAS® [8]. Sweeting et al. [10]'s continuity correction was also applied for model convergence. The model with the identity link function converged with our data despite the boundary condition problem. Table 6 presents the meta analyses results along with those in Section 4.1 to compare performances among the three fixed effect models based on frequentist approach we considered.

Table 6. Comparison of Fixed Effect Models based on Frequentist approach

Meta-analysis	Mean Excess Risk of Active Treatment per 1,000 Subjects over Placebo (90% CI)		
	Generalized Linear Model	Variance Inflation	Extended Newcombe
All 5 Studies	1.6 (-0.8, 4.1)	1.5 (-0.9, 3.9)	1.5 (-3.5, 3.3)
>= 1 Year Studies	1.9 (-0.6, 4.4)	1.7 (-0.6, 4.0)	1.7 (-2.8, 3.6)

Note that the results are similar between the two meta-analyses within the generalized linear model, thus indicating that an appropriate weight is given to each study. This comes from the fact that each event or no-event is treated with equal weight in model (11). In this model, the variance is estimated based on the large sample theory of maximum likelihood estimates and has slightly low coverage probability with smaller CI lengths compared with those in the extended Newcombe method.

The point estimate from the generalized linear fixed effect model is similar to the ones from the other two methods suggesting that continuity correction in the former did not have a negative impact on the analysis.

We tried to fit the generalized linear mixed effect model in Section 3.3 using PROC NLMIXED in SAS® [8]. We were unable to start the NLMIXED fitting using the best fixed starting point with grid and boundary conditions. It seems that using WinBUGS under Bayesian approach is more advantageous as the random effect specification is far more flexible with Bayesian approach especially for non-canonical links.

5. Conclusion

We considered several meta-analysis methodology with our case study. These methods can be grouped into four using the criteria of "fixed versus random effect" as well as "frequentist versus Bayesian approach." With Bayesian approaches, we demonstrated the surprisingly negative impact of the prior distribution on the outcome when the sample sizes were not balanced between the treatment groups within a study. We also noted that the random effect assumption tended to equalize the impact of studies that had vast differences in sample sizes. In addition, we were unable to fit the random effect model in the frequentist setting that uses non-canonical parameters.

We therefore opted for the extended Newcombe method, a fixed effect model based on frequentist approach. Additional advantage is found in that it is relatively straight forward to compare the extended Newcombe method results with each individual study result due to a restricted set of model assumptions. We emphasize that this choice is by no means a universal recommendation. Our recommendation is to practice a similar process we presented; that is, to carefully evaluate several methods and choose the one that is most appropriate for the data.

References

[1] Agresti A, Coull BA. Approximate is better than exact for interval estimation of binomial proportions. The American Statistician, 1998; **52**: 119-126.

[2] Bradburn MJ, Deeks JJ, Berlin JA, Localio AR. Much ado about nothing: a comparson of the performance of meta-analystical methods with rare events. *Statistics in Medicine*, 2007; **26**:53-77.

[3] Cai T, Parast L, Ryan L. Meta-analysis for rare events. *Statistics in Medicine*, 2010; **29**: 2078-2089.

[4] Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. *Pharmaceutical Statistics*, 2011,
10:3-7.

[5] Dann SR, Koch GG. Methods for one-sided testing of the difference between proportions and sample size considerations related to non-inferiority clinical trials. *Pharmaceutical Statistics*, 2008; **7**: 130–141.

[6] McCullagh P, Nelder JA. Generalized Linear Models, 2nd edition. Chapman and Hall, 1989.

[7] Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 1998; **17**:873-890.

[8] SAS Institute Inc. SAS/STAT_ 9.2 Users Guide. SAS Institute Inc.: Cary, NC, 2008.

[9] 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.

[10] 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.

[11] Warn DE, Thompson SG, Spiegelhalter DJ. 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.