Integrated Data Analysis for Assessing Treatment Effect through Combining Information

Hui Quan[1], Bingzhi Zhang[1], Christy Chuang-Stein, Byron Jones[2]
[1]Sanofi, and [2]Novartis

Abstract

It is critical to use a precise estimate of treatment effect when drawing conclusions, evaluating benefit/risk or designing a new study. Utilization of data from all sources in an integrated data analysis/meta-analysis will help us move closer to meeting this need. Depending on the data sources and objectives, there are many approaches for integrated analyses. These include network meta-analysis, multivariate meta-analysis, model-based meta-analysis as well as methods of borrowing historical data. In this paper, we discuss these methods with additional details for implementation and interpretation. We consider information adaptive repeated cumulative meta-analyses. We also discuss how to apply three integrated analysis approaches that take into account the variability of the overall treatment effect estimate to determine sample size for a new trial. Some computation and simulation results are provided.

Key words: meta-analysis, historical data, surrogate endpoint, sample size calculation.

1. Introduction

There are many methods for combining data in an integrated data analysis. They include the methods of network meta-analysis [1-6], multivariate meta-analysis [7-8], borrowing of historical data [9-14], model-based meta-analysis [15-17] and other types of analysis [18-20]. For network meta-analysis, Jones et al. (2011) and Achana et al. (2013) focus on the case of a binary endpoint. Mavridis et al. (2014) discuss a model for accounting for publication bias in a full network meta-analysis. Rucker and Schwarzer (2014) use an example to illustrate the inclusion of studies with direct and indirect comparisons for more efficient comparisons between any two treatments. To implement the methods of network meta-analysis, one needs information on the correlations between any pair of between-treatment differences within study. Unfortunately, publications typically do not provide this information. Rucker and Schwarzer (2014) assume certain values for the correlations in their work. In reality, these correlations depend on the sample size or the number of events of the individual treatments. In addition, adjustment for study level covariates may be necessary to meet the assumption of consistency of treatment effects across studies for a network meta-analysis to be valid. We will discuss this and other considerations related to network meta-analysis in Section 2.1.

When there are two endpoints, a univariate analysis is usually applied to each endpoint separately, especially when one endpoint is a continuous endpoint and the other is a binary or time-to-event endpoint. This practice applies to meta-analysis also. Riley et al. (2007) consider a bivariate meta-analysis and provide formulas for the overall estimates of treatment effects on individual endpoints. Even when the two endpoints are of a different type, the formula derived from bivariate meta-analysis may still be used for potentially more efficient estimation of treatment effects of the individual endpoints. In Section 2.2, we compare the performances of the bivariate approach with the univariate approach. The comparison focuses on how much information from the treatment effect estimate on one endpoint can contribute to the estimation of the treatment effect on the other endpoints under different scenarios.

In recent years, many trialists have realized the value of using historical data when designing and analyzing a trial. Many of the applications focus on borrowing historical control data for the purpose of reducing the sample size of the control arm in a new study [10-12]. Still, there is an

increasing trend to conduct a meta-analysis of all drugs in the same class and apply the overall effect estimate to all drugs in the same class, at least for safety assessment (Kramer, 2009 [21]). This idea, appropriately modified, can be used to help estimate the efficacy effect of a new drug also. The first step is to determine how much historical information to borrow. One approach is to discount data from other drugs using a shrinkage estimation approach [22]. Under this approach, less information will be borrowed from other drugs if the between-drug variability is high. Another approach is to use power likelihood [12-14] for the other drugs. The idea is to use a power (exponent) parameter (between 0 and 1) to control the extent of data borrowing from each existing drug. The power parameter is determined based on how similar an existing drug is to the new drug. We discuss these approaches in Section 2.3.

There are also other types of approaches for integrated data analysis such as the model-based meta-analysis. Model-based meta-analysis is especially appealing as it can adjust for design differences between studies and allow for the development of a prediction model. We include examples of model-based meta-analysis in Section 2.4.

During a new drug development process, studies are being completed sequentially. Sequential repeated meta-analyses have been proposed to take advantage of the cumulative data to make strategic decisions in a timely manner. For example, if the amount of data anticipated is likely to be inadequate for a specific objective based on the conditional power calculation, adaptation on program-wise sample size may be made. When this occurs, the analysis needs to take into account the potential sample size increase to control the overall type I error rate. We discuss adaptive repeated cumulative meta-analysis including the option for sample size re-estimation in Section 3.1 and sample size calculation for a new study with different variability for the estimate of treatment effect in Section 3.2. We conclude the paper with additional remarks in Section 4.

2. Methods of integrated data analysis

In this section, we summarize and provide new insights on some existing methods for integrated/meta-analysis.

2.1. Network Meta-analysis

Network meta-analysis is a useful tool to combine information from all trials to compare between any pair of treatments. A network meta-analysis includes trials that contribute to the comparison between two treatments directly (e.g., comparing Treatment 1 versus Treatment 2) and trials that contribute indirectly to a comparison (e.g., comparing Treatment 1 versus Treatment 2 through a comparison between Treatment 1 versus Treatment 3 and a comparison between Treatment 2 versus Treatment 3). Extensive literature exists on network meta-analysis [1-6].

Suppose there are a total of K treatments and a total of S studies that have at least one pair of the K treatments. Let $\theta = (\theta_1,..., \theta_K)'$ denote the vector of parameters of the individual within-treatment effects. For certain endpoints, the estimates for $\{\theta_i\}$ may not be available from publications or study reports. Let $\mu_{ij} = \theta_i - \theta_j$ represent the difference in effect between treatments i and j. These parameters are more likely to appear in a publication than the $\{\theta_i\}$. For example, for the analysis of a time to event endpoint, hazard rates for individual treatment groups potentially as functions of time may not be always available. However, the hazard ratio for between-treatment comparison is typically available. We borrow the data example of Rucker and Schwarzer (2013) as shown in Table 1 for illustration.

Notice that if a study has more than two treatments, some of the between-treatment differences can be expressed as a linear combination of the other between-treatment differences. For

example, for a study with 3 treatment arms 1, 2 and 3 like Study 1 in Table 1, the design matrix $\begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ for the three comparisons for Treatment 1 versus 2, 1 versus 3 and 2 versus 3 or

$$\begin{pmatrix} \mu_{12} \\ \mu_{13} \\ \mu_{23} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$$

is singular because difference $\theta_2 - \theta_3$ can be expressed as the difference between $\theta_1 - \theta_3$ and $\theta_1 - \theta_2$. The corresponding full rank design matrix concerning $\theta = (\theta_1, ..., \theta_K)'$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & ... & 0 \\ 1 & 0 & -1 & 0 & ... & 0 \end{pmatrix} \qquad (1)$$

can be obtained by deleting the third row and adding other columns with zero elements for θ_k, k>3. For the other studies with perhaps other treatments, the full rank design matrices like (1) can also be specified in a similar fashion. Putting all these full rank design matrices for all studies together and deleting the last column, we have a full rank design matrix (otherwise the sum of all the columns will be zero) treating θ_K as a reference. For the studies and between-treatment comparisons in Table 1, we have

$$\begin{pmatrix} \theta_1 - \theta_2 \\ \theta_1 - \theta_3 \\ \theta_1 - \theta_2 \\ \theta_1 \\ \theta_1 - \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \qquad (2)$$

or $\mu = A\theta$.

Table 1. Fictional Example from Rucker and Schwarzer (2013): 4 treatments and 4 studies

Study ID	First Treatment in a Pair	Second Treatment in a Pair	Difference	Variance of difference
1	1	2	0.50	0.03
1	1	3	0.75	0.04
1	2	3	0.25	0.05
2	1	2	0.40	0.02
2	1	4	0.60	0.05
2	2	4	0.20	0.05
3	1	2	0.45	0.05
4	3	4	0.25	0.05

Because the estimators of treatment effects from different studies are independent, the corresponding covariance matrix is full rank and block diagonal with the diagonal elements from the variance column in Table 1

$$\Sigma = \begin{pmatrix} 0.03 & r_1 & 0 & 0 & 0 & 0 \\ r_1 & 0.04 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.02 & r_2 & 0 & 0 \\ 0 & 0 & r_2 & 0.05 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.05 \end{pmatrix}. \qquad (3)$$

Both r_1 and r_2 in (3) are not directly available from Table 1 and may also not be directly available from publications. Since $\text{cov}(\hat{\theta}_1 - \hat{\theta}_2, \hat{\theta}_1 - \hat{\theta}_3) = \text{var}(\hat{\theta}_1)$, they are actually the estimates of the variance of $\hat{\theta}_1$ in study 1 and study 2, respectively. One assumption for the application of the network meta-analysis is consistency of treatment effects across studies. This strong assumption is particularly important for indirect comparisons and may not hold if different studies were conducted in patients with different characteristics. To make this assumption more attainable, it may be better to incorporate study level baseline characteristics into the analysis.

Suppose *B* is the design matrix for the baseline covariates of interest and ξ is the vector of the coefficients or effects of the covariates. Combining the two design matrices, we have

$$\hat{\mu} = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} \theta \\ \xi \end{pmatrix} + \varepsilon = X\beta + \varepsilon \qquad (4)$$

where $\hat{\mu}$ is the vector of the observed values of μ, $X = \begin{pmatrix} A & B \end{pmatrix}$, $\beta = \begin{pmatrix} \theta \\ \xi \end{pmatrix}$ and ε has a multivariate normal distribution with mean zero and known covariance matrix Σ. Based on (4), the least square estimate of β is

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\hat{\mu} \qquad (5)$$

and the corresponding estimate of the covariance is $\text{var}(\hat{\beta}) = (X'\Sigma^{-1}X)^{-1}$. With $\hat{\theta}$, the estimate and the variance of the relative treatment effect between any pair of treatments for the network meta-analysis can be derived. For example, $\hat{\theta}_i$ is the estimate of treatment effect of Treatment *i* compared to the reference treatment, while $\hat{\theta}_1 - \hat{\theta}_2$ is the estimate of treatment effect between Treatments 1 and 2. The covariance matrix $\text{var}(\hat{\theta})$ for $\hat{\theta}$ through (5) is the upper left corner submatrix of $\text{var}(\hat{\beta})$. If baseline covariates are not considered in the analysis (see (2) for example), an alternative estimate for θ can be obtained through

$$\hat{\theta}^* = (A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}\hat{\mu} \qquad (6)$$

and the corresponding covariance matrix $\text{var}(\hat{\theta}^*) = (A'\Sigma^{-1}A)^{-1}$. Since $(A'\Sigma^{-1}A)$ is the upper left corner submatrix of $X'\Sigma^{-1}X$, theoretically, $\text{var}(\hat{\theta}^*) < \text{var}(\hat{\theta})$. We will demonstrate this through simulations later.

For a binary endpoint, sample sizes and numbers of events for individual treatments are generally available from publications. Jones et al. (2011) discuss method for network meta-analysis for a binary endpoint. Their methods can be easily extended to incorporate study level baseline covariates in the analysis.

If only two treatments are considered and all the included studies had the two treatments, one of the treatments can be treated as a reference and (5) is essentially a meta-regression analysis. Note that in this case all off-diagonal elements of the covariance matrix are zero.

We conducted simulations to compare the performances of $\hat{\theta}$ in (5) (denoted as Method 1) with $\hat{\theta}^*$ in (6) (denoted as Method 2). Results are reported in Table 2. For the simulations, we assumed that $\hat{\mu} \sim N(X\beta, \Sigma)$ and used 10,000 runs. For the case of 6 data points

$$X_6 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 0.03 & 0.01 & 0 & 0 & 0 & 0 \\ 0.01 & 0.04 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.02 & 0.01 & 0 & 0 \\ 0 & 0 & 0.01 & 0.05 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.05 \end{pmatrix}.$$

As anticipated, the variance of Method 1 is greater than that of Method 2 for all cases. When there is no covariate effects (ξ =0), both methods provide consistent estimates and the 95% confidence intervals also have the nominal coverage probabilities. When there are large covariate effects, Method 2 provides biased estimates and the coverage probabilities of the 95% confidence intervals were much smaller than the nominal level probably due to the bias and partially due to the smaller estimated variance particularly when the number of observations is large. Basically, incorporating covariates in the analysis provides more robust results. The choice of ξ for the second scenario was to assess the ultimate impact of the covariate effect on the estimation of treatment effect.

Table 2. Simulation results for network meta-analysis

		θ_1	θ_2	θ_3	ξ_1	ξ_2	ξ_3
	True value	1.00	0.50	0.25	-0.05	-0.05	0.05
$\bar{\hat{\beta}}$	Method 1	0.9990	0.4981	0.2409	-0.0385	-0.0647	0.0441
	Method 2	0.9336	0.4739	0.1948	NA	NA	NA
$\overline{\widehat{var}(\hat{\beta})}$	Method 1	0.0800	0.0500	0.1000	0.1500	0.1400	0.0800
	Method 2	0.0315	0.0326	0.0311	NA	NA	NA
Coverage rate of 95% CI	Method 1	0.9560	0.9540	0.9470	0.9510	0.9350	0.9430
	Method 2	0.9280	0.9510	0.9380	NA	NA	NA
		θ_1	θ_2	θ_3	ξ_1	ξ_2	ξ_3
	True value	1.00	0.50	0.25	-0.30	-0.20	0.20
$\bar{\hat{\beta}}$	Method 1	1.0090	0.5022	0.2513	-0.2978	-0.2041	0.1858
	Method 2	0.6503	0.3506	-0.0418	NA	NA	NA
$\overline{\widehat{var}(\hat{\beta})}$	Method 1	0.0800	0.0500	0.1000	0.1500	0.1400	0.0800
	Method 2	0.0315	0.0326	0.0311	NA	NA	NA
Coverage rate of 95% CI	Method 1	0.9510	0.9500	0.9420	0.9370	0.9390	0.9360
	Method 2	0.5040	0.8650	0.6240	NA	NA	NA

2.2. Multivariate meta-analysis and the use of surrogate endpoint

Multivariate meta-analysis has been proposed for joint synthesis of treatment effects on multiple endpoints [7-8]. This type of analysis allows borrowing information from correlated endpoints

when assessing treatment effect on one specific endpoint. The borrowing may result in potentially more efficient estimates. For example, if a Phase IIb study uses only a surrogate endpoint as the study endpoint, the estimate of treatment effect on the surrogate endpoint can be utilized to estimate treatment effect on a clinical endpoint.

Let's consider the case of two endpoints. Riley et al (2007) propose a random effects meta-analysis approach for the evaluation of bivariate outcomes. They provide close form formulas for the estimates of treatment effects of individual endpoints and the corresponding estimates of variances. The use of a random effects model demands a robust estimate of the between-study covariance that needs a reasonably large number of studies. Unfortunately, for most new drug development programs, the number of studies, particularly the number of studies with the clinical endpoint as the primary study endpoint, is usually small. Therefore, the use of a fixed effects model is more reasonable even though it does not take into account between-study variability.

Suppose the vector of estimates of treatment effects on the two endpoints from *S* sources has the following distribution

$$\begin{pmatrix}\hat{\eta}_{1i}\\ \hat{\eta}_{2i}\end{pmatrix} \Big| \begin{pmatrix}\eta_{1i}\\ \eta_{2i}\end{pmatrix} \sim BVN\left(\begin{pmatrix}\eta_{1i}\\ \eta_{2i}\end{pmatrix}, \Sigma_i\right), \; i=1, \ldots, S.$$

For a fixed effects model, $\begin{pmatrix}\eta_{1i}\\ \eta_{2i}\end{pmatrix} = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}$. Despite different patient populations for different studies, it may be reasonable to assume a constant treatment effect across studies upon the selection of an appropriate metric to measure the magnitude of the treatment effect. For example, if hazard ratio is used for measuring treatment effect (log hazard ratio will be used for statistical inference), the treatment effect observed from a Phase II study with a low risk patient population may be similar to that of a Phase III study with an enriched patient population. Besides choosing an appropriate metric, adopting the same standard for endpoint measurement and the same adjudication process within the same drug development program as well as clever adjustment of baseline characteristics could also contribute to a reasonably constant treatment effect across studies.

Denote the covariance matrix of the treatment effect estimates on the two endpoints in study *i* by

$$\hat{\Sigma}_i = \begin{pmatrix}\hat{\sigma}_{1i}^2 & \rho\hat{\sigma}_{1i}\hat{\sigma}_{2i}\\ \rho\hat{\sigma}_{1i}\hat{\sigma}_{2i} & \hat{\sigma}_{2i}^2\end{pmatrix}, \; i=1, \ldots, S.$$

To simplify the notations, we use $\lambda_i = \rho\hat{\sigma}_{1i}\hat{\sigma}_{2i}$ to denote the estimated covariance between the treatment effect estimates in study *i*. The estimated variance $\hat{\sigma}_{ji}^2$ depends on the amount of information in study *i* and can differ from study to study. We assume a common correlation ρ across studies here.

When individual patient data are available, we can estimate the correlation by applying a multivariate analysis to the two endpoints. A univariate approach may be more convenient when the two endpoints are of different types, e.g., one endpoint is continuous and the other is a time-to-event endpoint. When a univariate approach is used to analyze the two endpoints separately, one can apply the Jackknife or bootstrapping method to obtain an estimate of the correlation as suggested by Daniels and Hughes (1997). The weighted average of the estimated correlations with sample sizes as the weights across studies is an estimate for the common correlation. This estimate can be used for studies without individual patient data that can be used to estimate the correlation.

Setting between-study covariance matrix to zero in the formulas of Riley et al. (2007), we obtain the overall estimate of treatment effect for endpoint *j* as

$$\hat{\eta}_j = \frac{\sum_{i=1}^{S} \frac{\hat{\eta}_{ji}}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2} [\sum_{k=1}^{S} \frac{\hat{\sigma}_{j'i}^2 \hat{\sigma}_{jk}^2 - \lambda_i \lambda_k}{\hat{\sigma}_{jk}^2 \hat{\sigma}_{j'k}^2 - \lambda_k^2}] + \sum_{i=1}^{S} \frac{\hat{\eta}_{j'i}}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2} [\sum_{k=1}^{S} \frac{\lambda_k \hat{\sigma}_{ji}^2 - \lambda_i \hat{\sigma}_{jk}^2}{\hat{\sigma}_{jk}^2 \hat{\sigma}_{j'k}^2 - \lambda_k^2}]}{\sum_{i=1}^{S} \frac{\hat{\sigma}_{ji}^2}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2} \sum_{i=1}^{S} \frac{\hat{\sigma}_{j'i}^2}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2} - (\sum_{i=1}^{S} \frac{\lambda_i}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2})^2} \quad (7)$$

and an estimate for its variance

$$Var(\hat{\eta}_j) = \frac{\sum_{i=1}^{S} \frac{\hat{\sigma}_{ji}^2}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2}}{\sum_{i=1}^{S} \frac{\hat{\sigma}_{ji}^2}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2} \sum_{i=1}^{S} \frac{\hat{\sigma}_{j'i}^2}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2} - (\sum_{i=1}^{S} \frac{\lambda_i}{\hat{\sigma}_{ji}^2 \hat{\sigma}_{j'i}^2 - \lambda_i^2})^2}, \; j=1, 2 \quad (8)$$

where $j' = 2$ if j=1 and $j' = 1$ if j=2. If the correlations for all the studies are zero, (7) and (8) reduce to

$$\hat{\eta}_j = \frac{\sum_{i=1}^{S} \frac{\hat{\eta}_{ji}}{\hat{\sigma}_{ji}^2}}{\sum_{i=1}^{S} \frac{1}{\hat{\sigma}_{ji}^2}} \quad \text{and} \quad \text{var}(\hat{\eta}_j) = \frac{1}{\sum_{i=1}^{S} \frac{1}{\hat{\sigma}_{ji}^2}}. \quad (9)$$

These are the formulas for the univariate meta-analysis based on a fixed effects model when the inverse variance weighting is used. For a study with no data for an endpoint, Riley et al. (2007) proposed setting the corresponding variance of the estimate of the treatment effect for the endpoint to infinity and still use formulas (7) and (8).

Table 3. Simulation results for comparing (7) and (8) versus (9)
with data for the second endpoint available for only some of the studies

ρ	Method	$\hat{\eta}_1$	$\hat{\eta}_2$	Var($\hat{\eta}_1$)	Var($\hat{\eta}_2$)	v̂ar($\hat{\eta}_1$)	v̂ar($\hat{\eta}_2$)
0.2	1	0.15017	0.49931	0.00018	0.00664	0.00019	0.00661
	2	0.15017	0.49938	0.00019	0.00677	0.00019	0.00675
0.4	1	0.15015	0.49926	0.00017	0.00623	0.00019	0.00626
	2	0.15015	0.49937	0.00019	0.00677	0.00019	0.00675
0.6	1	0.15013	0.49930	0.00015	0.00539	0.00020	0.00575
	2	0.15013	0.49936	0.00019	0.00677	0.00019	0.00675
0.8	1	0.15012	0.49953	0.00010	0.00375	0.00024	0.00547
	2	0.15010	0.49936	0.00019	0.00677	0.00019	0.00675

We conducted simulations to compare the performance of (7) and (8) versus (9), assuming the endpoint follows a normal distribution. We consider the case of 5 studies. The true parameters for treatment effects on the two endpoints are $\eta_1 = 0.15$ and $\eta_2 = 0.5$. The corresponding standard deviations are 0.32 and 1.41. We select sample sizes of 124, 158, 198, 252, 337 per group, so that power to detect non-zero treatment effects for the two endpoints in the 5 studies at level of 0.05 are 0.74, 0.84, 0.91, 0.96, 0.99, and 0.50, 0.60, 0.70, 0.80, 0.90 correspondingly. We assume that data on endpoint 2 are not observed in study 1, 2 and 3. Results are summarized in Table 4. Method 1 is based on (7) and (8), and Method 2 is based on (9). Both methods continue to provide consistent estimates of the true treatment effects. As the correlation increases, the variance of the estimate of endpoint 2 from Method 1 becomes smaller than that of Method 2 through borrowing information from the other endpoint. The sample variance of the point estimate is similar to the average of the estimates of the corresponding variance when ρ is small, but as ρ increases the average of the variance estimates becomes smaller than the sample

variance of the estimates. For endpoint 1 we find that Var($\hat{\eta}_1$), the average of the variance estimate in (8) under Method 1 decreases as ρ increases. By comparison, the sample variance of the estimates $\hat{\text{v}}\text{ar}(\hat{\eta}_1)$ under Method 1 goes in the opposite direction and becomes greater than that of Method 2 when ρ is large.

2.3. Use of historical data or data of other drugs in the same class

There is a rich literature on borrowing control data from historical trials with the goal of reducing sample size of the control arm of a new trial [9-14]. Even though there is still the consistency issue when borrowing data from multiple historical studies that were conducted for different drugs, there is less concern if only control data are utilized. Also, it should be less controversial if data from studies conducted on the same drug are included in a meta-analysis. However, if one intends to borrow data of both active treatment and control arms from studies conducted on different drugs (albeit in the same class) to assess treatment effect of a specific drug, one should consider discounting part of the historical data. We will discuss discounting approaches in this section.

The usual practice is to first conduct a meta-analysis for individual drugs in the same class to obtain the estimate $\hat{\delta}_i$ and variance $\hat{\sigma}_i^2$ for drug i, i=1, …, D. Models for these meta-analyses could be different. For example, some of them could be a fixed-effects model if the treatment effects are homogeneous across the studies of the same drug while others could be a random-effects model. A second layer of meta-analysis based on $\hat{\delta}_i$ and $\hat{\sigma}_i^2$, i=1, …, D can then be performed to obtain an overall estimate of treatment effect for the entire class. If the number of drugs in the class is reasonably large, one could assume a random effects model for the true treatment effect for drug i

$$\delta_i \sim N(\delta, \tau^2),\ i=1, \ldots, D \qquad (10)$$

with between-drug variability/variance τ^2. The overall estimate $\tilde{\delta}$ of the true treatment effect δ in (10) for the whole drug class is

$$\tilde{\delta} = \sum_{i=1}^{D} w_i \hat{\delta}_i / \sum_{i=1}^{D} w_i = \sum_{i=1}^{D} w_i \hat{\delta}_i \ / \ w \sim N(\delta, 1/w)$$

where $\hat{\tau}^2$ is the estimate of τ^2 and $w_i = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$ is the estimate of the inverse of the variance of $\hat{\delta}_i$, and $w = \sum_{i=1}^{D} w_i$. An empirical shrinkage estimator for δ_i is

$$\tilde{\delta}_i = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}_i^2}\hat{\delta}_i + \frac{\hat{\sigma}_i^2}{\hat{\tau}^2 + \hat{\sigma}_i^2}\tilde{\delta}.$$

The estimator $\tilde{\delta}_i$ is derived by borrowing information from the other drugs through $\tilde{\delta}$. The larger the between-drug variance $\hat{\tau}^2$ is or the smaller the $\hat{\sigma}_i^2$ is, the less information will be borrowed from the other drugs. One concern for the use of the empirical shrinkage estimator is that it may borrow too much information from the other drugs when estimating δ_i [22]. Also, the empirical shrinkage method relies on the use of a random effects model that requires a reasonably large number of drugs in the class and the estimator has a relatively large variability. Variance of $\tilde{\delta}_i$ under a random effects model can be estimated. When the number of drugs in the class is small, it is more realistic to use a fixed effects model for the second layer meta-analysis where the true treatment effect δ_i for drug i is treated as a parameter rather than a random variable as in a random-effects model. When the δ_i's are the same across all drugs, this is equivalent to τ^2=0 in a random-effects model. In this case, the overall estimate of treatment effect for the whole class based on the fixed effects model is

$$\hat{\delta} = \sum_{i=1}^{D} \hat{\delta}_i / \hat{\sigma}_i^2 / \sum_{i=1}^{D} 1/\hat{\sigma}_i^2 \ .$$

Another estimate for δ_i is the James-Stein (J-S) estimator that takes the following form

$$\breve{\delta}_i = c\hat{\delta}_i + (1-c)\delta_0 = \delta_0 + c(\hat{\delta}_i - \delta_0)$$

The J-S estimator minimizes the following mean square error under a fixed effects model

$$WMSE = \sum_{i=1}^{D} E(\breve{\delta}_i - \delta_i)^2 / \sigma_i^2 = \sum_{i=1}^{D} (c-1)^2 (\delta_i - \delta_0)^2 / \sigma_i^2 + c^2 D \, .$$

The optimal parameters that minimize the WMSE are

$$\delta_0 = \bar{\delta} = \frac{\sum_{i=1}^{D} \delta_i / \sigma_i^2}{\sum_{i=1}^{D} 1/\sigma_i^2} \qquad (11)$$

and

$$c = \frac{\sum_{i=1}^{D} (\delta_i - \bar{\delta})^2 / \sigma_i^2}{\sum_{i=1}^{D} (\delta_i - \bar{\delta})^2 / \sigma_i^2 + D} \qquad (12)$$

where $\sum_{i=1}^{s} (\delta_i - \bar{\delta})^2 / \sigma_i^2$ is a measure of the variability among $\delta_i, i = 1,2,...,D$. After replacing the unknown δ_i in (11) and (12) by $\hat{\delta}_i$, we obtain $\hat{c}$ and $\breve{\delta}_i = \hat{c}\hat{\delta}_i + (1-\hat{c})\hat{\delta}$. The J-S estimator differs from the empirical shrinkage estimate in that the shrinkage factor for the J-S estimator is constant across different drugs. For a fixed $\hat{c}$ which is a consistent estimate of *c*, the variance of the J-S estimator can be estimated by

$$\text{var}(\breve{\delta}_i) = \hat{c}^2 \sigma_i^2 + (1-\hat{c})(1+\hat{c}) / \sum_{j=1}^{D} 1/\sigma_j^2 \ .$$

The J-S estimator borrows information from the other drugs through $\hat{\delta}$. The larger the between drug variability is, the less information from the other drugs will be borrowed. Compared to the empirical shrinkage estimate, the J-S estimator generally borrows less information from the other drugs (Quan et al. 2014) [22].

The empirical Bayes and the J-S estimation approaches explicitly express the amount of information to be borrowed from the other drugs. A more flexible alternative approach may be the use of power likelihood. This approach is often used for borrowing partial historical control data for a new study [12-14]. For our scenario, a power (or exponent) b_j ($0 \le b_j \le 1$) is applied to the likelihood of the estimate of the treatment effect of drug *j*. The magnitude of b_j depends on the similarity of drug *j* to the target drug *i*. If drug *j* is totally different from drug *i*, b_j can be set to zero. If drug *j* is believed to be very similar to drug *i*, b_j can be set close to 1. The overall likelihood for the parameter δ_i of the treatment effect of drug *i* is

$$F(\delta_i) = f(\hat{\delta}_i, \delta_i) \prod_{j \ne i}^{D} f^{b_j}(\hat{\delta}_j, \delta_i) \ , \ 0 \le b_j \le 1$$

where $f(\hat{\delta}_j, \delta_i)$ is the likelihood based on data of drug *j*. For example, it can be the density of an asymptotic normal distribution $\hat{\delta}_j \sim N(\delta_i, \hat{\sigma}_j^2)$. The determination of b_j is similar to the specification of the prior distribution of the effect of drug *j* toward δ_i in a Bayesian analysis setting. Ibrahim and Chen (2000) propose a Bayesian analysis with power prior. Basically, a prior distribution $\pi_0(\delta_i)$ for δ_i is added to the above likelihood.

2.4. Model-based integrated data analysis

Model-based integrated analysis can be applied to address questions that cannot be answered through a regular meta-analysis. They are generally specific depending on the nature of the

endpoint (continuous, binary, time to event or even recurrent event), the availability of data (e.g., PK data) and objective of the analysis. Therefore, it is difficult to discuss a broadly applicable model-based meta-analysis approach. We will illustrate the thinking via specific examples below.

Individual patient data from studies are typically available to the trial sponsor. When the sponsor performs an integrated analysis, individual patient data can be included in a more sophisticated model for potentially more efficient analysis. Quan et al (2014) discuss a case of modeling exposure-response relationship to justify the dose selection for a drug treating multiple sclerosis. Two phase III studies with very similar design and the same primary endpoint of annualized number of relapses (count data) demonstrated similar treatment effects for two doses of an experimental drug. A question arose as to whether a dose higher than the two existing doses would provide an improved treatment efficacy and a dose lower than the existing ones would also provide a similar treatment efficacy effect. Rather than designing a new Phase III study to address the question, a PK/PD modeling approach was applied. Data from the two Phase III studies were integrated in the analysis. A negative binomial regression model with study as a stratification factor, steady-state mean trough PK concentration and many other baseline values/charateristics as covariates was fitted to the data. The model reasonably depicted the relationship between PK parameters and the annualized relapse rate. After obtaining the predicted PK geometric means for all doses including doses not studied in the Phase III program (using a separate dose-PK model), treatment effects of different doses versus placebo control can be predicted through the negative binomial regression model.

When individual patient data are not available, model-based meta-analysis needs a reasonably large number of studies. Gross et al. (2013) apply a novel model-based meta-analysis to indirectly compare the efficacy of two DPP-4 inhibitors for treating type 2 diabetes mellitus using data from 25 studies. Moreover, Mercier et al. (2014) propose a model-based meta-analysis to make indirect comparison between two opioid drugs in their effect on pain intensity along with the tolerability characterized by the adverse event and dropout rates.

3. Additional applications of integrated and meta-analysis

Besides more precise quantification of treatment effect, integrated data analysis or meta-analysis can be used to generate hypothesis or help make a regulatory decision. For example, since safety evaluation needs a large data set, it may be acceptable to use a meta-analysis to conclude the non-inferiority of an experimental drug compared to a control on a safety endpoint. For example, a guidance issued by the Food and Drug Administration in the United States [23] states that a meta-analysis of cardiovascular (CV) event data from a new diabetes drug development program can be used to rule out certain level of CV risk associated with the drug.

During the development process, it may be necessary to assess, based on available data from completed studies, the amount of data required in future studies to achieve the desired power for the ultimate decision. As a result of the assessment, sample size for the whole program may be adapted. Depending on how the variability of the estimate of treatment effect is taken into account, the required sample size for a new study (or studies) can be very different. In this section, we will discuss these additional applications of integrated data analysis.

3.1. Adaptive and repeated meta-analyses and conditional power calculation

In a new drug development program, the Clinical Development Plan (CDP) outlines the number of studies along with the primary objective and the required sample size for each study. Using the CDP, one can estimate the amount of cumulative data over time from the program. Repeated meta-analyses can be planned to address a question that cannot be answered using data of individual studies alone. That is, repeated meta-analyses can be performed upon the completions

of additional major studies to closely monitor treatment effect. This also allows adjustment to be made to the development strategy, if necessary (Quan et al 2014 [18]). Suppose the null and alternative hypotheses are

$$H_0 : \delta \geq \Delta \qquad \text{versus} \qquad H_a : \delta < \Delta\,, \qquad (13)$$

where δ is the true overall treatment effect. Assuming small δ is more desirable, Δ in (13) will be set to 0 for testing superiority and $\Delta > 0$ for testing non-inferiority.

Among many things, the analysis plan should pre-specify the total number of meta-analyses *K* and how the overall type-I error rate will be controlled. For example, if there is no intention to stop the new drug development program early unless there is a serious safety concern, then no alpha will be spent at the intermediate meta-analyses and the entire alpha will be reserved for the final analysis. This will ensure the highest power for the final analysis. Pre-specifications contribute to the credibility of results from the repeated meta-analysis.

Suppose a conditional power calculation is performed after the k^{th} meta-analysis. Given the observed results from the previous analyses, conditional power as a function of the true treatment effect δ at the final analysis for the pre-specified weight w_i ($\sum_{i=1}^{K} w_i^2 = 1$) is

$$\begin{aligned} \text{CP}(\delta) &= \Pr\Big(\Big(\sum_{i=1}^{K} w_i \hat{\delta}_i / \hat{\sigma}_i + z_{1-\alpha}\Big) \Big/ \Big(\sum_{i=1}^{K} w_i / \hat{\sigma}_i\Big) \leq \Delta \mid \delta, \sum_{j=1}^{k} w_j \hat{\delta}_j / \hat{\sigma}_j\Big) \\ &= \Pr\Big(Z \leq \frac{\Delta \sum_{i=1}^{K} w_i / \hat{\sigma}_i - \sum_{j=1}^{k} w_j \hat{\delta}_j / \hat{\sigma}_j - z_{1-\alpha} - \delta \sum_{i=k+1}^{K} w_i / \hat{\sigma}_i}{\sqrt{\sum_{i=k+1}^{K} w_i^2}} \mid \delta, \sum_{j=1}^{k} w_j \hat{\delta}_i / \hat{\sigma}_i\Big)\,, \end{aligned}$$

where $\hat{\delta}_i$ and $\hat{\sigma}_i^2$ are the estimate of treatment effect and the corresponding variance based on data between the *i*-1 and *i*th meta-analyses, respectively; also, where $Z \sim N(0,1)$ asymptotically. To have $1-\beta'$ conditional power, we need

$$(\Delta - \delta) \sum_{i=k+1}^{K} w_i / \hat{\sigma}_i = \sum_{j=1}^{k} w_j (\hat{\delta}_j - \Delta) / \hat{\sigma}_j + z_{1-\alpha} + z_{1-\beta'} \sqrt{\sum_{i=k+1}^{K} w_i^2}\,. \qquad (14)$$

The above conditions suggest that we may need to add new studies into the development program to have enough data from the future studies such that $\sum_{i=k+1}^{K} w_i / \hat{\sigma}_i$ is large enough for (14) to hold. When new studies are added between the originally planned i-1 and *i*th meta-analyses, additional data will make $\hat{\sigma}_i$'s smaller.

Sutton et al. (2007) consider the problem of designing a new study and adding it to an updated meta-analysis. They propose to first conduct a meta-analysis of the existing studies. The estimate of treatment effect from this meta-analysis is used to design the new study and to calculate the required sample size to test the hypothesis in the updated meta-analysis. Sample size for the new study obtained this way can be considered a sample size adaptation based on the observed treatment effect of the previous studies. Without appropriate adjustment in the updated meta-analysis, the type I error rate may not be controlled. For example, if the observed treatment effect from the meta-analysis of the existing studies is already significant, the new study will be small. If the actual sample sizes are used as weights in the updated meta-analysis, a small new study will contribute very little to the updated meta-analysis and thus make it easier to demonstrate a significant treatment effect. However, if the weights are pre-specified and not dependent on the early result, a smaller new study will have a larger variability in the estimate and may therefore not guarantee a significant treatment effect in the updated meta-analysis.

3.2. Application of the integrated analysis to design a new study

One of the purposes for an integrated data analysis is to obtain an estimate of treatment effect for sample size calculation for a new study. In this section, as in the case of designing a new Phase III study using existing data, we assume that only data from the new study are used in the statistical inference and will not be combined with historical data for an updated meta-analysis. Suppose $(\hat{\delta}_i \mid \delta_i) \sim N(\delta_i, \sigma_i^2)$ and $\delta_i \sim N(\delta, \tau^2)$, i=1, …, H are from the historical studies potentially for different compounds but for the same indication. The true treatment effect δ^* for the new study is also assumed to follow the same normal distribution

$$\delta^* \sim N(\delta, \tau^2). \qquad (15)$$

The analysis approach for the new study is the frequentist hypothesis-testing approach. For example, to declare a positive result, the upper bound of the confidence interval should satisfy $\hat{\delta}^* + z_{1-\alpha}\sigma\sqrt{2/N^*} < \Delta$ where N^* is the sample size per treatment group and σ^2 is the variance of the endpoint for the new study. There are different approaches for calculating the sample size for the new study for the desired power. A common approach is to simply use formula

$$N^{1*} = 2\sigma^2(z_{1-\alpha} + z_{1-\beta})^2/(\Delta - \hat{\delta})^2. \qquad (16)$$

Some researchers use the overall estimate $\hat{\delta}$ (defined in Section 2.3) of the treatment effect from existing studies based on a fixed-effects model as the true treatment effect for the new study and consider no variability for δ^*. In other words, they treat $\delta^* = \delta = \hat{\delta}$ and $\tau^2 = 0$.

An estimate of the overall treatment effect $\tilde{\delta}$ under a random-effects model can also be used to replace $\hat{\delta}$ in (16). To take the uncertainty or variability of δ^* into account, another sample size for the new study N^{2*} is the solution of

$$f(N^{2*}) = E_{\delta^*}\Pr(\hat{\delta}^* + z_{1-\alpha}\sigma\sqrt{2/N^{2*}} < \Delta \mid \delta^*) = 1-\beta,$$

where the expectation is with regard to

$$\delta^* \sim N(\tilde{\delta}, \tau^2). \qquad (17)$$

The estimate of the treatment effect for the new study, $\hat{\delta}^*$, follows $\hat{\delta}^* \mid \delta^* \sim N(\delta^*, 2\sigma^2/N^{2*})$. This N^{2*} takes into account the variability of δ^* but replaces the unknown parameter δ (in (15)) with $\tilde{\delta}$ without considering the variability associated with $\tilde{\delta}$. When $\tau^2 = 0$, N^{1*} and N^{2*} are the same. For $\tau^2 > 0$, N^{2*} is greater than N^{1*}. Since $\tilde{\delta}$ itself is an estimate based on existing/historical data, it is prudent to consider its variability. This leads to a posterior distribution for δ as

$$(\delta \mid \hat{\delta}_1, ..., \hat{\delta}_H) \sim N(\tilde{\delta}, 1/\sum_{i=1}^{H} 1/(\sigma_i^2 + \tau^2))$$

where H represents the number of prior studies included in determining $\tilde{\delta}$. Thus,

$$(\delta^* \mid \hat{\delta}_1, ..., \hat{\delta}_H) \sim N(\tilde{\delta}, 1/\sum_{i=1}^{H} 1/(\sigma_i^2 + \tau^2) + \tau^2). \qquad (18)$$

This leads to a third sample size formula for the new study N^{3*} which is the solution of

$$g(N^{3*}) = E_{\delta^*}\Pr(\hat{\delta}^* + z_{1-\alpha}\sigma\sqrt{2/N^{3*}} < \Delta \mid \delta^*) = 1-\beta$$

where the expectation is with regard to the distribution in (18). Since the variance in (18) is larger than that in (17), N^{3*} is greater than N^{2*} even when $\tau^2 = 0$. The unknown parameters σ^2 and

τ^2 can be estimated using historical data. Rather than using the normal distribution in (18) with an unrestricted support $(-\infty,\infty)$, one can consider using a truncated distribution with a support $(-\infty,\Delta^*)$ (where $\Delta^*(<\infty)$ is a finite value) for power and sample size calculation. Restricting the support to $(-\infty,\Delta^*)$ will result in a sample size smaller than N^{3*}.

We compare the required sample size per group for different desired power when estimate of treatment effect based on historical data δ =0.3, σ =1, one-sided significance level α =0.025, τ =0.01, 0.025, 0.05, 0.075 and the sample sizes for 5 historical studies are 113, 143, 163, 191, 235. For example, when τ =0.05, the required sample sizes for the three approaches for 90% power are 234, 263 and 302, respectively; and the difference between N^{1*} and N^{3*} is 68. The sample size difference increases with desired power. Taking into account all the variability, N^{3*} gives us the largest sample size. If the estimate of treatment effect based on very limited historical data is much better than what we anticipated and populations across studies are different, we may be willing to consider a relatively larger sample size N^{3*} for the new trial.

One issue with a large sample size for a study is the study's ability to detect a very small treatment effect which may not be clinically meaningful. For superiority assessment (Δ =0), the minimum observed treatment effect that will give a significant P-value is

$$\hat{\delta}^* = -z_{1-\alpha}\sigma\sqrt{2/N^*}\ .$$

They are -0.181, -0.171 and -0.160 for the three sample sizes (234, 263 and 302) under the design parameters specified above. That is, if sample size is 302, the minimum observed treatment effect to be statistically significant at level of 0.025 is -0.160. If -0.16 is a clinically meaningful treatment effect, it will be reasonable to consider a trial with a sample size of 302 per group. On the other hand, if a clinically meaningful treatment effect is at least -0.181, it will be more appropriate to consider 234 per group with the understanding that we will not have a positive trial if the observed estimate is larger than -0.181.

4. Discussion

In general, the main purpose of an integrated data analysis is to obtain a more precise estimate of the treatment effect. Results from the analysis can be used to generate hypothesis for future studies or modify a current development program. They are rarely used for making a regulatory claim in efficacy. Integrated data analysis has been regularly applied to safety data to assess patients' adverse reactions to a drug. This is by necessity because the typical studies in a clinical development program are rarely designed from the safety perspective. In recent years, integrated data analysis has been used to rule out a safety concern based on a formal non-inferiority assessment. For example, the guidance for a new diabetes drug issued by the FDA in 2008 requires the drug sponsor to rule out a 1.8 relative risk for a major cardiovascular adverse event (MACE) before the drug could be approved for marketing. The sponsor needs to rule out a relative risk of 1.3 post-marketing. The assessment can be done either through a meta-analysis or a single CV outcome trial (Marchenko et al, 2015 [24]).

Recently, two major CV outcome trials (EXAMINE and SAVOR [25-26]) of two DPP-4 inhibitors for type 2 diabetes (alogliptin and saxagliptin) successfully ruled out a 1.3 relative risk for MACE for these two drugs. Detailed trial results were discussed at an FDA Endocrinologic and Metabolic Drug Advisory Committee meeting on April 14 2015. In addition to EXAMINE and SAVOR, another CV outcome trial (TECOS) of a different DPP-4 inhibitor (sitagliptin) was recently completed and found to meet the primary objective of ruling out a 1.3 relative CV risk for MACE also [27]. Since a long term and large CV outcome trial demands a great amount of

resources, a series of positive CV outcome trials have prompted the question of how we could learn from completed CV outcome trials to help assess CV risk of new type 2 diabetes drugs in a more efficient manner. For example, can a future drug borrow information on CV risk from previous drugs? In addition, if several drugs are developed concurrently, can we include all these drugs and one control in one trial since recent CV outcome trials share much similarity in inclusion/exclusion criteria, primary endpoints, endpoint adjudication and analysis strategy. Including several drugs in one trial saves on control subjects and allows for head-to-head comparisons between drugs.

The demand of personalized medicines also needs more data for subgroup and biomarker analysis. Since individual studies are usually not designed to have sufficient power for subgroup analyses, it will be beneficial to combine data from multiple studies, and at the same time to borrow information from different subgroup by using methods described in this paper. Health economics assessment may be another area that relies on integrated analysis to quantify country or region specific benefit and cost ratio to compare the utility of a new drug to existing ones.

As data become more readily available through various data transparency initiatives, we can expect greater access to patient level data for researchers. Data could come from very different sources including randomized clinical trials and observational studies. In this paper, we focus on analytical approaches for integrated data analysis. An important question is the appropriateness of combining data from very different sources. We did not address this question in this paper. We refer readers to a paper by Chuang-Stein et al (2016) [28] on how to use data from different sources to enable benefit-risk assessment. We hope to address questions related to data sources for integrated data analysis in a future paper.

References

1. Jones B, Roger J, Lane PW, Lawton A, Fletcher C, Cappelleri JC, Tate H, Moneuse P. Statistical approaches for conducting network meta-analysis in drug development. Pharmaceutical Statistics 2011; 10: 523-531.
2. Rucker G and Schwarzer G. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. Stat in Med 2014; 33: 4353-4369.
3. Konig J, Krahn U and Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Statistics in Medicine 2013; 32: 5414-5429.
4. Mavridis D, Welton NJ, Sutton A and Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. Statistics in Medicine 2014; ?
5. Achana FA, Cooper NJ, Dias S, Lu G, Rice SJC, Kendrick D and Sutton A. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Statistics in Medicine 2013; 32: 752-771.
6. Giovane CD, Vacchi L, Mavridis D, Filippini G and Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. Statistics in Medicine 2013; 32: 25-39.
7. Riley RD, Abrams KR, Lambert PC, Sutton AJ and Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Statistics in Medicine 2007; 26: 78-97.
8. Daniels MJ and Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. Statistics in Medicine 1997 16: 1965-1982.
9. Neuenschwander B, Capkun-Niggli G, Branson Michael and Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. Clinical Trials. 2010; 7: 5-18.
10. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim JG, Kinnersley N, Lindborg S, Micallef S, Roychoudhury S and Thompson L. Use of historical

control data for assessing treatment effects in clinical trials. Pharmaceutical Statistics. 2014; 13: 41-54.
11. Gsteiger S, Neuenschwander B, Mercier F and Schmidli H. Using historical control information for the design and analysis of clinical trials with overdispersed count data. Statistics in Medicine. 2013; 32: 3609-3622.
12. Chen MH, Ibrahim JG, Lam P, Yu A and Zhang Y. Bayesian design of noninferiority trials for medical devices using historical data. Biometrics. 2011; 67: 1163-1170.
13. Ibrahim JG, Chen MH, Xia HA and Liu T. Bayesian meta-experimental design: evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. Biometrics. 2012; 68: 578-586.
14. Ibrahim JG and Chen MH. Power prior distributions for regression models. Statistical Science. 2000; 15: 46-60.
15. Quan H, Mao X, Wei L and Wang L. A case study of modeling and exposure-response prediction for count data. *Journal of Biopharmaceutical Statistics*. 2014; 24: 1073-1090.
16. Mercier F, Claret L, Prins K and Bruno R. A Model-based meta-analysis to compare efficacy and tolerability of Tramadol and Tapentadol for the treatment of chronic non-malignant. Pain Therapy. 2014; 3: 31-44.
17. Gross JL, Rogers J, Polhamus D, Gillespie W, Friedrich C, Gong Y, Monz BU, Patel S, Staab A and Retlich S. A novel model-based meta-analysis to indirectly estimate the comparative efficacy of two medications: an example using DPP-4 inhibitors, sitagliplitin and linagliptin, in treatment of type 2 diabetes mellitus. BMJ Open 2013; 3: e001844.
18. Quan H, Ma Y, Yan Zheng, Meehyung Cho, Christelle Lorenzato, Carole Hecquet. Adaptive and Repeated Cumulative Meta-Analyses for Safety Signal Detection during a New Drug Development Process. *Pharmaceutical Statistics*. 2015; 14: 161-171.
19. Chen MH, Ibrahim JG, Xia HA, Liu T and Hennessey V. Bayesian sequential meta-analysis design in evaluating cardiovascular risk in a new antidiabetic drug development program. Statistics in Medicine. 2014; 33: 1600-1618.
20. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR and Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. Stat in Med 2007; 26: 2479-2500.
21. Kramer JM. Balancing the benefits and risks of inhaled long-acting beta-agonists--the influence of values. *N Engl J Med*, 360(16):1592-5, 2009.
22. Quan H, Mao X, Chen J, Shih WJ, Ouyang SP, Zhang J, Zhao PL and Binkowitz B. Multi-regional clinical trial design and consistency assessment of treatment effects. Statistics in Medicine. 2014; 33: 2191-2205.
23. FDA, Guidance for Industry: diabetes mellitus-evaluating cardiovascular risk in new antidiabetic therapies to treat Type 2 diabetes, December 2008.
24. Marchenko O, Jiang Q, Chakravarty A, Ke C, Ma H, Maca J, Russek-Cohen E, Sanchez-Kam M, Zink RC, Chuang-Stein C. Evaluation and review of strategies to assess cardiovascular risk in clinical trials in patients with type 2 diabetes mellitus. To appear in *Statistics in Biopharmaceutical Research Journal*, 2015.
25. Willaim et al. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. The New England Journal of Medicine 2013.
26. Benjamin et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. The New England Journal of Medicine 2013.
27. Green et al. Effect of sitagliptin on cardiovascular outcomes in Type 2 diabetes. NEJM 2015; June 8:1-11.
28. Chuang-Stein C, Quartey G, He W, Jiang Q, Ma H, Norton J, Scott J, Berlin J. Sources of data to enable benefit-risk assessment. In *Benefit-Risk Assessment Methods in Medicinal Product Development: Bridging Qualitative and Quantitative Assessments* (eds, He W and Jiang Q). CRC Press Taylor & Francis Group, 2016.