Supplementary Materials for Bayesian Information Sharing and Interim Efficacy Monitoring for Equivalence Testing with an Application to Dose Proportionality Studies

Wenru Zhou[1], Samantha MaWhinney[1], Peter Anderson[2], and Alexander Kaizer[1]

[1]Department of Biostatistics and Informatics University of Colorado

[2]Department of Pharmaceutical Sciences University of Colorado

This file includes supplementary materials for the manuscript "Bayesian Information Sharing for Equivalence Testing with an Application to Dose Proportionality Studies." The sections below provide additional details on the MEM approach that are published elsewhere in greater detail (e.g., [1], [2]), additional figures and tables to summarize results from simulation studies presented in the main paper, and results for sensitivity analyses to evaluate the performance under different settings including a simpler parallel arm dose finding study design without a crossover that avoids correlation amongst repeated doses.

1 MEM Approach

In this section we briefly introduce the multisource exchangeability model approach to sharing information across multiple data sources. We do so in the context of PK studies, so we assume we are going to conduct a study to assess the dose proportionality of a drug.

For our PK study, we will use the power model and use the C_{ss} as the PK response.: $ln(C_{ss}) = \beta_{0,p} + \beta_{1,p} ln(dose_p) + \epsilon$, where $\epsilon \sim N(0, E)$ and p represents our current primary study. Meanwhile, there may be similar completed studies about the dose proportionality of our study drug. We want to borrow from these historic data sources for our primary study if we suspect the models are exchangeable. For each historic data source, the power model is similarly conducted: $ln(PK_h) = \beta_{0,h} + \beta_{1,h} ln(dose_h) + \epsilon$, where $\epsilon \sim N(0, E)$ and h represents each supplementary data source $h = 1, 2, \ldots, H$, where H is the total number of historical data sources. If $\beta_p = \beta_h$, we will conclude that the primary data and the h^{th} supplementary data are exchangeable based on the MEM approach [1]. Since there are H historical data sources, and we wish to determine whether each source is either "exchangeable" or "non-exchangeable" with the primary study, we are going to have $K = 2^H$ possible exchangeability configurations defined as Ω_k, where $k = 1, 2, \ldots, K$.

Figure 1 is an example of how MEMs are defined assuming there are two historical data sources ($Data_1$, $Data_2$) and one primary study $Data_p$. This leads to 4 assumed exchangeability patterns $(\Omega_1, \Omega_2, \Omega_3, \Omega_4)$ since we have $H = 2$ and $K = 2^2$. The arrows in Figure 1 indicate the data sources within each model that are pooled together to estimate the β_p parameter used to evaluate dose proportionality in our primary study, so arrows from supplemental data sources indicate ones with assumed exchangeability with the current study

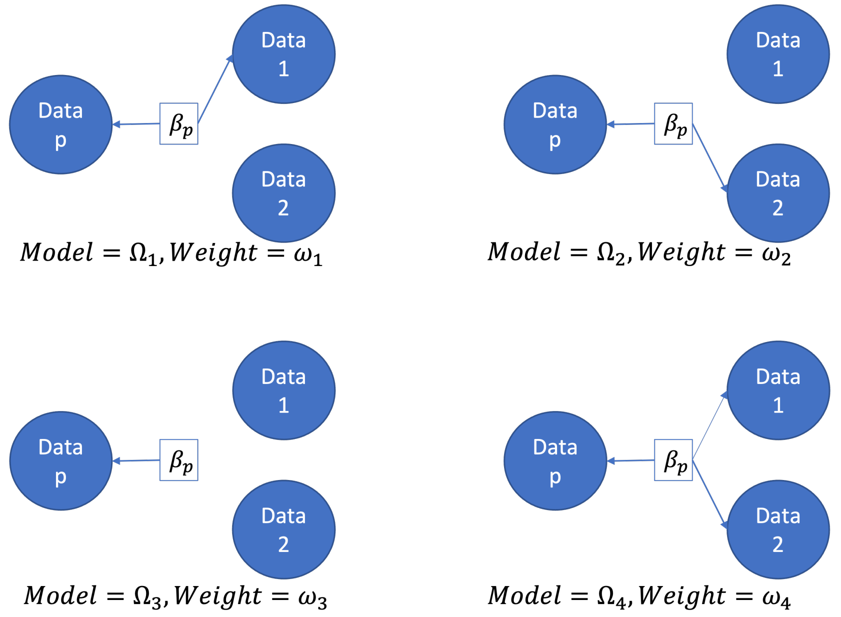

Data p
β_p
Data 1
Data 2
$Model = \Omega_1, Weight = \omega_1$
Data p
β_p
Data 1
Data 2
$Model = \Omega_2, Weight = \omega_2$
Data p
β_p
Data 1
Data 2
$Model = \Omega_3, Weight = \omega_3$
Data p
β_p
Data 1
Data 2
$Model = \Omega_4, Weight = \omega_4$

Figure 1: Example of Multisource Exchangeability Model with two supplemental trials ($H = 2$) resulting in four possible configurations of exchangeability ($K = 4$)

for a given configuration. If no arrow point to a data source, this data source is not assumed exchangeable with the primary data source.

From this approach, we can see that ω_k is influenced by $\pi(\Omega_k)$ and BIC from each configuration's different exchangeability patterns. Therefore, the choice of $\pi(\Omega_k)$ is important. Based on prior work by [1], within the MEM framework these priors are set on the *source*-level instead of the *configuration*-level, thus reducing the dimensionality from 2^H to H. Here we set our prior as π_e, which is defined as a common probability for each supplementary data source being exchangeable with the primary data source. If $\pi_e = 0.05$ and there is 1 supplementary source, the weight of the pattern of non-exchangeability is $\pi(\Omega_1) = (1 - \pi_e) = 0.95$, and for exchangeability is $\pi(\omega_2) = \pi_e = 0.05$. If there are 2 supplementary sources, like Figure 1, there will be 4 patterns: $\pi(\Omega_1) = 0.05 * (1 - 0.05)$, $\pi(\Omega_2) = (1 - 0.05) * 0.05$, $\pi(\Omega_3) = (1 - 0.05)^2$, and $\pi(\Omega_4) = 0.05^2$. As part of this Chapter, we will demonstrate that the choice of π_e is very important and may impact the ultimate trial results. Traditionally, it can be obtained by calibrating until desired frequentist trial operating characteristics (e.g., power and type I error rates) are achieved. One may also set a conservative value, for example, 0.05, in place of fine tuned calibration of π_e, to reduce the potential influential of supplemental data sources with a more conservative approach.

Since our aim is to assess dose proportionality, whether the coefficient of the dose equaling 1 is our primary interest. We obtain all coefficients from appropriate regression models, for example, the power model. Based on a real-world motivated PK setting, we assess the behavior of MEMs under settings with one or two supplementary data sources.

Consider a case with only one supplemental data source. For notation, define S=0 when data is from the primary study and S=1 when data is from the supplementary study. The regression model for the non-exchangeability configuration (i.e., assuming the supplemental study is not exchangeable with the current study) is:

$$ln(C_{ss}) = \beta_0 + \beta_1 ln(Dose) + \beta_2 I(S = 1) + \beta_3 Dose * I(S = 1) + \epsilon_{11} \tag{1}$$

where $\epsilon_{11} \sim N(0, E)$. For the exchangeability configuration (i.e., assuming the supplemental source is exchangeable with the current study) is:

$$ln(C_{ss}) = \beta_0 + \beta_1 ln(Dose) + \beta_2 I(S = 1) + \epsilon_{12} \tag{2}$$

where $\epsilon_{12} \sim N(0, E)$.

This configuration allows the coefficient of dose to be β_1 for both supplementary and primary data if they are assumed exchangeable. If they are not assumed exchangeable, the estimated coefficient for the dose will still be β_1 for primary data, but will be $\beta_1 + \beta_2$ for the supplemental data source.

This approach can be generalized to more historic sources. For instance, if there are 2 historical sources, $H = 2$, $S = 1, 2$, and $K = 2^2 = 4$, so there will be four configurations of exchangeability.

For two non-exchangeable sources:

$$ln(C_{ss}) = \beta_0 + \beta_1 ln(Dose) + \beta_2 I(S = 1) + \beta_3 I(S = 2) + \beta_4 Dose * I(S = 1) + \beta_5 Dose * I(S = 2) + \epsilon_{21} \tag{3}$$

where $\epsilon_{21} \sim N(0, E)$. If only one of the data sources is exchangeable with the primary study:

$$ln(C_{ss}) = \beta_0 + \beta_1 ln(Dose) + \beta_2 I(S = 1) + \beta_3 I(S = 2) + \beta_4 Dose * I(S = 1) + \epsilon_{22} \tag{4}$$

or

$$ln(C_{ss}) = \beta_0 + \beta_1 ln(Dose) + \beta_2 I(S = 1) + \beta_3 I(S = 2) + \beta_5 Dose * I(S = 2) + \epsilon_{23} \tag{5}$$

where $\epsilon_{22} \sim N(0, E)$ and $\epsilon_{23} \sim N(0, E)$. If both sources are exchangeable with primary study:

$$ln(C_{ss}) = \beta_0 + \beta_1 ln(Dose) + \beta_2 I(S = 1) + \beta_3 I(S = 2) + \epsilon_{24} \tag{6}$$

where ϵ_{24} $N(0, E)$.

2 Constrained MEM Approach

At the beginning of the study and its earlier interim analyses, the sample size in the primary study will be smaller, and therefore it may be misleading to borrow too much information from historical data. For example, the study might borrow a lot of information in the 1st interim analysis, but borrow much less information in the 2nd interim analysis as additional primary data clarifies the potential exchangeability. To address this challenge, we limit the maximum information borrowed from supplementary data at each interim analysis if the borrowed information exceeds some threshold, for example, the sample size at the next stage that will be enrolled in primary study. This was proposed by [2] as a more conservative approach to borrowing as compared to the "unconstrained" MEM approach introduced previously. While π_e can also be used to adjust the borrowing strength, it cannot adaptively adjust the borrowing strength based on the existing weights and sample size for the next stage like the constrained MEM does.

To identify if "too much" information is borrowed from supplementary data, we are going to first calculate the effective supplementary sample size (ESSS). Define $ESSS_{i,k}$ at interim analysis i and model k as $ESSS_{i,k} = n_{P,i}(\frac{P_k}{P_1} - 1)$, where $n_{P,i}$ is the sample size in primary study at interim analysis i, and P_1, P_k are the precisions from the posteriors for β_1 in the no-borrowing model and in the model k with some borrowing. Therefore, $ESSS_{i,k}$ can be interpreted as the additional sample size that should be enrolled in the primary study at analysis i to achieve the same precision of posterior from model k if no information is borrowed. We repeat this process for all exchangeability models, so that we have $ESSS_i = \sum_{j=1}^{K} ESSS_{i,j} = n_{P,i} \sum_{j=1}^{K} \omega_j (\frac{P_j}{P_1} - 1)$, where $K = 2^H$ is the total number of exchangeability models.

Then the sample size for the next stage for primary data $n_{P,i+1}$ is used as the threshold of adjustment. This means that if $ESSS_i > n_{P,i+1}$, a penalty on the weights for information borrowing models will be applied. The process of adjustment is following:

1. If $ESSS_i < n_{P,i+1}$, just use the original weights and then sample β_1 from the weighted mixture posterior distribution $q(\beta|D_p, D_1, D_2...D_H) = \sum_{j=1}^{K} \omega_j q(\beta|\Omega_j)$. Otherwise, go to the next step.

2. Define a shrinkage factor $s = n_{P,i+1}/ESSS_i$

3. Change the posterior weight of the no-borrowing model from w_1 to $w_1^{'} = w_1 + (1 - s)(1 - w_1)$

4. Change the posterior weight of some borrowing models j from w_j to $w_j^{'} = s * w_j$, $j = 1, 2, \ldots, K$

5. Sample β_1 from the new weighted mixture posterior distribution $q(\beta|D_p, D_1, D_2...D_H) = \sum_{j=1}^{K} \omega_j^{'} q(\beta|\Omega_j)$

3 Practical Considerations for Success Criteria with Incorrect Dosing

One question that people may raise is what will happen if the given dose is out of the range of minimum dose and maximum dose in the experiment? In the design stage of the experiment, we need first to identify the minimum and maximum doses and then establish the criteria of dose proportionality. However, people may want to use the result of dose proportionality from our study within their study. What will happen if the dose is not within the range of minimum and maximum doses?

In a strict sense, dose proportionality may not be held if the given dose is out of the range. Based on the power function, $PK = e^{\beta_0} dose^{\beta_1}$, the derivative of PK in terms of β_1 can be obtained as $\frac{dPK}{ddose} = e^{\beta_0}\beta_1 dose^{\beta_1 - 1}$. Ideally, if $\beta_1 = 1$, and assuming the power function of PK is always held, the derivative will be e^{β_0} no matter how the dose changes. However, in practice, we are only interested in a range of doses that are meaningful to the human body, and the requirement for β_1 should not be that restricted. For example, if β_1 is near 1, and the change of dose is only slight, the derivative is still approximately a constant. This is why the FDA defined the criteria of dose proportionality based on the ratio of maximum and minimum dose, r. The smaller r is, the smaller change of dose, the less restriction for β_1 to hold the derivative approximately constant. Therefore, if the given dose is out of range, the current criteria should be adjusted, and a new experiment is needed to assess dose proportionality.

In practice, one may also wish to consider the extent of incorrect dosing. If this occurred for a single participant as a protocol deviation, it may be that a modified intention-to-treat analysis or per protocol analysis may be most appropriate to only include those adhering to the intended dose range. If an intention-to-treat design was used, one might consider using the same criteria but clearly noting the protocol deviation to assist in contextualizing any results.

4 One Supplementary Source Scenario Results

This section presents extended simulation results from the main manuscript for one supplementary source.

Figure 2 illustrates the operating characteristics for scenarios with one supplementary source for the three approaches: Constrained MEM (MEM-C), Unconstrained MEM (MEM-U), and group sequential method with Pocock boundary (Pocock). Numeric summaries are presented in Table 1. The operating characteristics evaluated include the bias, mean square error (MSE), power/type I error rate, and the average interim analysis stopping point. Results are stratified by different slopes from the primary and supplementary studies. Each figure is further stratified by small sample size and large sample size scenarios as well as the true status of dose proportionality (i.e., proportional or not).

4.1 Bias and MSE

Figure 2a presents the bias of β_p from each approach. For the two MEM methods, the green bars (1,1), which represent primary slope = supplementary slope = 1, are the shortest among all bars in the scenario of dose proportionality. Light blue (0.95, 1) and steel blue (1.05, 1) are two scenarios where dose proportionality is also achieved since 0.95 and 1.05 both in the equivalence interval of (0.84, 1.16). Those bars show symmetry around the green (1, 1) bar, and have higher bias than (1,1).

The bias in scenarios where the primary trial is not dose proportional, represented by yellow (0.84, 1), orange (1.16, 1), coral (0.75, 1) and pink (1.25, 1), all show rotational symmetry. Those patterns indicate that, for MEM-C and MEM-U, the estimation of the slope is positively biased when the supplementary slope is 1 and the primary slope is less than 1, and negatively biased when the primary slope is greater than 1. This tendency illustrated the influence of the supplementary data source on primary data, therefore the estimation is biased towards 1. However, the bias of MEM-U is more severe than MEM-C since MEM-C restricted the weight put on supplementary data sources when information borrowing exceeded some threshold. The frequentist Pocock boundary, however, does not have an obvious pattern since its bias estimates are all extremely small (all biases are less than 0.01), because the frequentist approach is not influenced by the supplementary data source since no borrowing is allowed.

MEMs also have much larger biases when the primary slope is at the boundary of dose proportionality (i.e. 0.84 or 1.16), and the supplementary slope is 1 (yellow and orange bars). This situation will not happen if both the primary slope and supplementary slope are both at the same boundary of 0.84 (grey bars) since the supplementary source will not bias the primary data towards 1.

In addition, for MSE, Figure 2b demonstrates that the MSE are small when slopes are (1,1), and are maximized when slopes are at (0.84, 1) and (1.16, 1), indicating that MEMs have more variability in estimation of β_1 when the primary source β_1 is on the boundary and the supplementary source β_1 is at 1. The MSE from MEM is often greater than the MSE from Pocock, but this is likely attributable to the difference in sample size caused by more frequent early stopping and incorporation of non-exchangeable supplemental information with MEMs.

4.2 Power and Type I Error Rates

Figure 2c visualizes the power and type I error rates across different scenarios. For the scenarios with slopes (0.84, 1) or (1.16, 1), the type I error rate is inflated, especially when the sample size is large for MEM-U and MEM-C methods. This indicates that the bias introduced from the supplementary source is not necessarily eliminated when both primary and supplementary sample sizes get larger, but these results may change for different sample size combinations (e.g., the same size for both). For other scenarios with different slope combinations away from the 0.84 or 1.16 equivalence boundary, all type I error rates are controlled under 0.05.

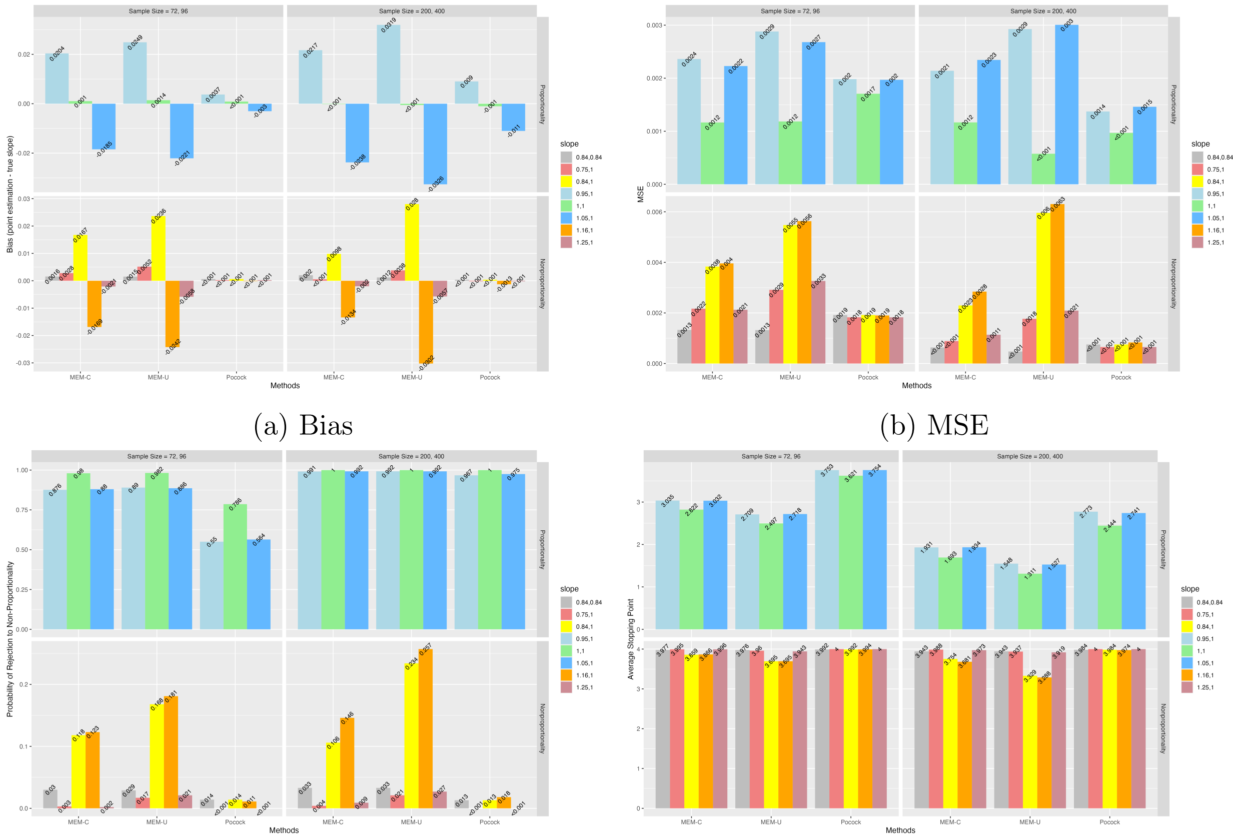
The power increases as the sample size increases for all approaches. For MEM-C, power increases to above 0.99 for an increase in sample size relative to a power of 0.88 with (0.95, 1), 0.98 of (1,1) and 0.88 of (1.05,1). For MEM-U, power also increases above 0.99 for the increase in sample size relative to a power of 0.89 for (0.95, 1), 0.98 for (1, 1), and 0.89 for (1.05, 1). For the Pocock approach without information sharing, the power increases above 0.96 for the larger sample size relative to a power of 0.55 with (0.95, 1), 0.79 of (1,1), and 0.56 of (1.05,1). With respect to power, the MEM-U and MEM-C approaches have much high power when the sample sizes are smaller (e.g., 36 for the primary study and 48 for the supplemental study) and all methods have high power in the larger sample size scenarios (e.g., 100 for the primary study and 200 for the supplemental study).

4.3 Averaging Stopping Point

In Figure 2d, the y-axis shows the Average Stopping Point for each primary and supplementary slope combination. Ideally, we want the light blue, green and steel blue bars in the proportionality scenarios to be lower, reflecting more early stopping for identification of dose proportionality prior to achieving full enrollment. Conversely, we want the grey, coral, yellow, orange, and pink bars in the nonproportionality scenarios to be equal to 4, indicating that they stop early for dose proportionality at an interim analysis.

In the situation of proportionality, the three average stopping points of MEM-U (2.7, 2.5, 2.7) are the lowest compared to the other methods. The three average stopping points of MEM-C (3.0, 2.8, 3.0) are slightly higher than MEM-U, but still lower than Pocock (3.8, 3.6, 3.7). Increasing primary and supplementary sample size from (36, 48) to (100, 200) decreased average stopping points for situation of dose proportionality. Larger average stopping points indicate that the study did not stop early for dose proportionality though dose proportionality exists.

Although MEM-U and MEM-C showed the highest power when does proportionality exists, they also showed a larger number of incorrect early stops for slopes in simulation scenarios reflecting the boundary cases (i.e., (0.84, 1) and (1.16, 1)), which in turn indicate an inflated type I error rate when nonproportionality exists. When the primary sample size is 36 and the supplementary sample size is 48, the average stopping points for the yellow and orange bars for MEM-U are both 3.7 and for MEM-C are both 3.9. The number of



(a) Bias

(b) MSE

(c) Power/Type I Error Rate

(d) Average Stopping Point

Figure 2: Bias, MSE, rejection rates, and average stopping point by method with $\pi_e = 0.05$ for MEM-U and MEM-C. Coloring represents different simulation scenarios for primary and supplemental source with the colored bars (coral, yellow, light blue, green, steel blue, orange, and pink) representing cases where the supplemental trial slope is 1 versus the grey bar representing a supplemental slope of 0.84. Figures are faceted by sample size (columns) and by proportionality of the primary source being near 1 (rows).

incorrect early stops becomes larger when sample sizes become (100, 200). The average stopping points for the yellow and orange bars for MEM-U are both 3.3 and for MEM-C are both 3.7, while for Pocock the average stopping point is always above 3.9.

Table 1: Comparison of Results among Unconstrained MEM, Constrained MEM, and Pocock under different π_e for MEM approaches

Sample Size	Slope	Method	Proportion of Reject Null, Stratified by π_e			Bias, Stratified by π_e			MSE, Stratified by π_e			Average Stopping Point, Stratified by π_e		
			0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
36,48	0.84,1	MEM-U	0.168	0.076	0.051	0.024	0.008	0.003	0.005	0.003	0.002	3.695	3.892	3.946
36,48	0.84,1	MEM-C	0.118	0.074	0.051	0.017	0.008	0.003	0.004	0.003	0.002	3.859	3.9	3.946
36,48	0.84,1	Pocock	0.014	0.014	0.014	0.001	0.001	0.001	0.002	0.002	0.002	3.992	3.992	3.992
36,48	1,1	MEM-U	*0.982*	*0.949*	*0.926*	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*	*0.002*	*2.497*	*2.929*	*3.08*
36,48	1,1	MEM-C	*0.98*	*0.949*	*0.926*	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*	*0.002*	*2.822*	*2.933*	*3.08*
36,48	1,1	Pocock	*0.786*	*0.786*	*0.786*	*0.001*	*0.001*	*0.001*	*0.002*	*0.002*	*0.002*	*3.621*	*3.621*	*3.621*
100,200	0.84,1	MEM-U	0.234	0.135	0.085	0.028	0.014	0.008	0.006	0.003	0.002	3.329	3.646	3.8
100,200	0.84,1	MEM-C	0.106	0.1	0.085	0.01	0.009	0.008	0.002	0.002	0.002	3.754	3.763	3.8
100,200	0.84,1	Pocock	0.013	0.013	0.013	0	0	0	0.001	0.001	0.001	3.984	3.984	3.984
100,200	1,1	MEM-U	*1*	*1*	*1*	*-0.001*	*-0.001*	*0*	*0.001*	*0.001*	*0.001*	*1.311*	*1.545*	*1.731*
100,200	1,1	MEM-C	*1*	*1*	*1*	*0*	*0*	*0*	*0.001*	*0.001*	*0.001*	*1.693*	*1.696*	*1.732*
100,200	1,1	Pocock	*1*	*1*	*1*	*-0.001*	*-0.001*	*-0.001*	*0.001*	*0.001*	*0.001*	*2.444*	*2.444*	*2.444*

Bias falling between -0.001 and 0.001 is rounded to 0.

Table 2: Comparison of Results from MEM-U ($\pi_e = 0.05$), MEM-C ($\pi_e = 0.05$), and Pocock under different slopes with one supplemental trial

Sample Size	Slope	Proportion of Reject Null, Stratified by Methods			Bias, Stratified by Methods			MSE, Stratified by Methods			Average Stopping Point, Stratified by Methods		
		MEM-U	MEM-C	Pocock	MEM-U	MEM-C	Pocock	MEM-U	MEM-C	Pocock	MEM-U	MEM-C	Pocock
36,48	0.84,0.84	0.029	0.03	0.014	0.002	0.002	0.001	0.001	0.001	0.002	3.976	3.977	3.992
36,48	0.84,1	0.168	0.118	0.014	0.024	0.017	0.001	0.005	0.004	0.002	3.695	3.859	3.992
36,48	0.75,1	0.017	0.003	0	0.005	0.003	0	0.003	0.002	0.002	3.96	3.995	4
36,48	*0.95,1*	*0.89*	*0.876*	*0.55*	*0.025*	*0.02*	*0.004*	*0.003*	*0.002*	*0.002*	*2.709*	*3.035*	*3.753*
36,48	*1,1*	*0.982*	*0.98*	*0.786*	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*	*0.002*	*2.497*	*2.822*	*3.621*
36,48	*1.05,1*	*0.886*	*0.88*	*0.564*	*-0.022*	*-0.018*	*-0.003*	*0.003*	*0.002*	*0.002*	*2.718*	*3.032*	*3.754*
36,48	1.16,1	0.181	0.123	0.011	-0.024	-0.017	0	0.006	0.004	0.002	3.695	3.866	3.994
36,48	1.25,1	0.021	0.002	0	-0.006	-0.002	0	0.003	0.002	0.002	3.943	3.996	4
36,48	1.5,1	0	0	0	0	0	0	0.002	0.002	0.002	4	4	4
100,200	0.84,0.84	0.033	0.033	0.013	0.001	0.002	0	<0.001	0.001	0.001	3.943	3.943	3.984
100,200	0.84,1	0.234	0.106	0.013	0.028	0.01	0	0.006	0.002	0.001	3.329	3.754	3.984
100,200	0.75,1	0.021	0.004	0	0.004	0.001	0	0.002	0.001	0.001	3.937	3.988	4
100,200	*0.95,1*	*0.992*	*0.991*	*0.967*	*0.032*	*0.022*	*0.009*	*0.003*	*0.002*	*0.001*	*1.548*	*1.931*	*2.773*
100,200	*1,1*	*1*	*1*	*1*	*-0.001*	*0*	*-0.001*	*0.001*	*0.001*	*0.001*	*1.311*	*1.693*	*2.444*
100,200	*1.05,1*	*0.992*	*0.992*	*0.975*	*-0.033*	*-0.024*	*-0.011*	*0.003*	*0.002*	*0.001*	*1.527*	*1.934*	*2.741*
100,200	1.16,1	0.257	0.146	0.018	-0.03	-0.013	-0.001	0.006	0.003	0.001	3.288	3.681	3.974
100,200	1.25,1	0.027	0.009	0	-0.006	-0.002	0	0.002	0.001	0.001	3.919	3.973	4
100,200	1.5,1	0	0	0	0	0	0	0.001	0.001	0.001	4	4	4

Bias falling between -0.001 and 0.001 is rounded to 0.

5 Two Supplementary Sources Scenario Results

In this section we provide the results for simulation scenarios with two supplemental sources through numeric summaries in Table 3.

Table 3: Comparison of Results among Unconstrained MEM, Constrained MEM, and Pocock under different slopes, Two Supplementary Trials, $\pi_e = 0.05$

Sample Size	Slope	Proportion of Reject Null, Stratified by Methods			Bias, Stratified by Methods			MSE, Stratified by Methods			Average Stopping Point, Stratified by Methods		
		MEM-U	MEM-C	Pocock	MEM-U	MEM-C	Pocock	MEM-U	MEM-C	Pocock	MEM-U	MEM-C	Pocock
36,36,48	1,1,1	*0.994*	*0.993*	*0.786*	*0.002*	*0*	*0.001*	*0.001*	*0.001*	*0.002*	*1.614*	*2.872*	*3.621*
36,36,48	1,1.16,1	*0.962*	*0.961*	*0.786*	*0.014*	*0.009*	*0.001*	*0.002*	*0.002*	*0.002*	*2.681*	*2.912*	*3.621*
36,36,48	1.16,1,1	0.488	0.205	0.011	-0.068	-0.032	0	0.016	0.006	0.002	2.757	3.888	3.994
36,36,48	1.16,1.16,1	0.092	0.073	0.011	-0.015	-0.013	0	0.003	0.003	0.002	3.834	3.91	3.994

Bias falling between -0.001 and 0.001 is rounded to 0.

6 Sensitivity Analysis: Use of Linear Regression without Random Effects when Modeling Correlated Data

In this section we evaluate the operating characteristics when using the linear mixed model (LMM) or a simpler linear regression model that ignores the repeated doses of a participant as part of the crossover nature of the trial. We compare Pocock and MEM in terms of large or small sample size & dose proportionality or lack of dose proportionality, under two modeling approaches: (1) **Full:** Linear mixed model is applied to account for the correlated nature of the data; (2) **IC:** linear regression model is applied and ignored the correlated nature of the data. In this section, we only compare MEM-U and Pocock. Results are shown in Table 4, and results from dose proportionality are shown in *Italic font*.

When simpler models are applied (IC), the type I error rate from both MEM-U and Pocock increased very slightly for most scenarios, though MEM-U still has higher type I error rate than Pocock. However, for the primary and supplementary slope (0.84, 1), the type I error rate is severely inflated in LMM. We discussed previously that MEM is not stable when the primary slope is at the boundary while the supplementary slope is misleading. However, when a simpler linear regression model is applied, the type I error rate is reduced from 0.168 to 0.064 for the small sample size and 0.234 to 0.074 for the large sample size. This happened since the weight of the information borrowing model decreased from 0.90 to 0.74 when using the linear regression model instead of the LMM. The BIC of the information borrowing model is larger than the BIC of the no borrowing model in the linear regression model than LMM, therefore the weights of the information borrowing model become lower. In other words, applying the "incorrect" linear regression model on correlated data reduced the inflated type

I error rate when the primary slope is at the boundary and the supplementary slope indicates dose proportionality.

In the IC setting, the power from both MEM-U and Pocock is reduced, especially for a small sample size. Both type I error rate and power of MEM-U are higher than Pocock. What is interesting is that for the IC approach with a small sample size, the power of Pocock is much lower (0.444 for primary slope 1 and 0.305 for primary slope 1.05) because ignoring the correlation inflates the estimated variance from observations. But the MEM-U still maintains good power (0.757 for primary slope 1 and 0.617 for primary slope 1.05) because it borrows information from the single supplementary data source with slope=1. For the large sample size scenario, the power of Pocock for IC setting increased from around 0.35 to around 0.85. These results suggest that ignoring the correlated nature of data may lead to a lower type I error rate with a trade-off of lower statistical power. However, depending on the motivation of a study, this may be a tolerable trade-off if sample sizes cannot be increased to lead to larger power.

The bias from MEM-U also reduced as we applied the linear regression model to correlated data instead of LMM for (0.84, 1) and (1.05, 1) scenarios in both small and large sample sizes. For example, when sample sizes are (36, 48), and slopes are (0.84, 1), the bias of MEM-U dropped from 0.024 to 0.009, and when slopes are (1.05, 1), bias dropped from -0.022 to 0.009. The same trend is also observed in large sample sizes. When using the linear regression model, the bias is reduced relative to the LMM since more weight is put on the no information-sharing configuration of exchangeability in the MEMs.

When fitting the linear regression model, the average stopping points often increased relative to the LMM. For instance, for MEM-U at the slope (0.84, 1) and (1.05, 1), since the type I error rate is much more lower after applying a simpler model, the average stopped points increased from 3.7 to 3.9 for slope (0.84, 1), and from 2.7 to 3.5 for slope (1.05, 1) in small sample size. The same increase is observed in large sample sizes.

Table 4: Comparison of Results between Linear Mixed Model and Linear Regression Model for Correlated Data under Different Slopes, $\pi_e = 0.05$

Sample Size	Slope	Method	CI contains		Bias		MSE		Stoplook	
			Full	IC	Full	IC	Full	IC	Full	IC
36, 48	0.84, 0.84	MEM-Unconstrained	0.029	0.034	0.002	0.003	0.001	0.003	3.976	3.97
36, 48	0.84, 0.84	Pocock	0.014	0.016	0.001	0.002	0.002	0.003	3.992	3.988
36, 48	0.84, 1	MEM-Unconstrained	0.168	0.064	0.024	0.009	0.005	0.004	3.695	3.932
36, 48	0.84, 1	Pocock	0.014	0.016	0.001	0.002	0.002	0.003	3.992	3.988
36, 48	1.5, 1	MEM-Unconstrained	0	0	(-0.001,0.001)	0.002	0.002	0.003	4	4
36, 48	1.5, 1	Pocock	0	0	(-0.001,0.001)	0.002	0.002	0.003	4	4
36, 48	1, 1	MEM-Unconstrained	*0.982*	*0.757*	*0.001*	*0.001*	*0.001*	*0.002*	*2.497*	*3.375*
36, 48	1, 1	Pocock	*0.786*	*0.444*	*0.001*	*0.002*	*0.002*	*0.003*	*3.621*	*3.855*
36, 48	1.05, 1	MEM-Unconstrained	*0.886*	*0.617*	*-0.022*	*-0.009*	*0.003*	*0.003*	*2.718*	*3.474*
36, 48	1.05, 1	Pocock	*0.564*	*0.305*	*-0.003*	*(-0.001,0.001)*	*0.002*	*0.003*	*3.754*	*3.885*
100, 200	0.84, 0.84	MEM-Unconstrained	0.033	0.035	0.001	0.002	<0.001	0.001	3.943	3.954
100, 200	0.84, 0.84	Pocock	0.013	0.021	(-0.001,0.001)	0.003	0.001	0.001	3.984	3.972
100, 200	0.84, 1	MEM-Unconstrained	0.234	0.074	0.028	0.008	0.006	0.002	3.329	3.843
100, 200	0.84, 1	Pocock	0.013	0.021	(-0.001,0.001)	0.003	0.001	0.001	3.984	3.972
100, 200	1.5, 1	MEM-Unconstrained	0	0	(-0.001,0.001)	0.002	0.001	0.001	4	4
100, 200	1.5, 1	Pocock	0	0	(-0.001,0.001)	0.002	0.001	0.001	4	4
100, 200	1, 1	MEM-Unconstrained	*1*	*1*	*-0.001*	*0.001*	*0.001*	*0.001*	*1.311*	*1.913*
100, 200	1, 1	Pocock	*1*	*0.992*	*-0.001*	*0.001*	*0.001*	*0.001*	*2.444*	*2.483*
100, 200	1.05, 1	MEM-Unconstrained	*0.992*	*0.932*	*-0.033*	*-0.016*	*0.003*	*0.002*	*1.527*	*2.307*
100, 200	1.05, 1	Pocock	*0.975*	*0.841*	*-0.011*	*-0.006*	*0.001*	*0.002*	*2.741*	*2.936*

Bias falling between -0.001 and 0.001 is rounded to 0.
IC: Ignore the correlated nature of the data and use a linear regression model.
Full: Use the appropriate linear mixed model on the correlated data.

7 Sensitivity Analysis: Non-Correlated Data from a Study without Crossover

The previous sections considered scenarios that simulated correlated data by means of a cross-over trial and analyzed the data with and without the appropriate mixed effects model approaches. However, one may be interested in simpler scenarios where there is no correlated data. This represents an additional situation we denote as **NC** for no correlation. In this section, data is not correlated and one can fit the simpler linear regression model instead of the more complex linear mixed effects model.

This context has no correlation within a participant since they only have one allocated dose instead of two different doses in a cross-over. The simulation is based on the relationship:

$$ln(C_{ss,i}) = \beta_0 + \beta_1 ln(Dose) + \epsilon_i \tag{7}$$

where $i = 1, 2, \ldots, N$; $\epsilon_i \sim N(0, \sigma_e^2)$; and $\sigma_e = 0.15$. Again, we set $\beta_0 = 0$ since it is not important in dose proportionality assessment.

Table 5 displays the result from the linear regression model on uncorrelated data from parallel studies. The type I error rate is 0.082 for small sample size and 0.058 for large sample size for MEM-U when slope is at the 0.84 boundary for the primary study and is 1 for the secondary study. No obvious bias and MSE are observed. While the overall power and type I error rate for MEM-U is higher than the approach without information sharing and using Pocock boundaries, one can calibrate the MEM-U interim monitoring boundary to get more comparable type I error control under a given scenario (e.g., (0.84, 1)). Although an inflated type I error rate and MSE were identified at the boundary of 0.84 in the linear mixed model simulations, this drawback does not happen in the linear regression model without correlation since the linear regression model tends to put more weight on the nonexchangeable model.

Table 5: Results from Linear Model Applied to Uncorrelated Data under Different Slopes, $\pi_e = 0.05$

Sample Size	Slope	Method	CI contains 1	Bias	MSE	Stop look
36, 48	0.84, 0.84	MEM-Unconstrained	0.035	0.002	0.001	3.939
36, 48	0.84, 0.84	Pocock	0.021	0.002	0.001	3.97
36, 48	0.84, 1	MEM-Unconstrained	0.082	0.006	0.002	3.856
36, 48	0.84, 1	Pocock	0.021	0.002	0.001	3.97
36, 48	1.5, 1	MEM-Unconstrained	0	0.001	0.001	4
36, 48	1.5, 1	Pocock	0	0.001	0.001	4
36, 48	1, 1	MEM-Unconstrained	*0.995*	*0.001*	*0.001*	*1.866*
36, 48	1, 1	Pocock	*0.977*	*0.001*	*0.001*	*2.449*
36, 48	1.05, 1	MEM-Unconstrained	*0.919*	*-0.015*	*0.002*	*2.258*
36, 48	1.05, 1	Pocock	*0.812*	*-0.008*	*0.002*	*2.888*
100, 200	0.84, 0.84	MEM-Unconstrained	0.032	0.002	<0.001	3.942
100, 200	0.84, 0.84	Pocock	0.023	0.002	0.001	3.949
100, 200	0.84, 1	MEM-Unconstrained	0.058	0.004	0.001	3.872
100, 200	0.84, 1	Pocock	0.023	0.002	0.001	3.949
100, 200	1.5, 1	MEM-Unconstrained	0	0.001	<0.001	4
100, 200	1.5, 1	Pocock	0	0.001	<0.001	4
100, 200	1, 1	MEM-Unconstrained	*1*	*(-0.001,0.001)*	*0.001*	*1.049*
100, 200	1, 1	Pocock	*1*	*0.001*	*0.001*	*1.174*
100, 200	1.05, 1	MEM-Unconstrained	*1*	*-0.014*	*0.001*	*1.233*
100, 200	1.05, 1	Pocock	*1*	*-0.009*	*0.001*	*1.475*

8 Sensitivity Analysis: No Interim Monitoring

In this section sensitivity analyses based on designs that do not include interim monitoring are presented. Without interim monitoring, there are essentially two approaches to consider: one that allows information sharing with MEMs at the end of the primary study and the standard frequentist approach that does not include information sharing. Without interim monitoring, both MEM-C and MEM-U are identical, since the constrained MEM only applies to limiting information sharing during an interim analysis.

Table 6: Results when no interim monitoring is incorporated

Sample Size	Slope	Information Sharing via MEMs	CI Contains 1	Bias	MSE
36, 48	1, 1	Yes	0.986	0.001	0.001
36, 48	1, 1	No	0.968	-0.000	0.002
36, 48	0.95, 1	Yes	0.880	0.013	0.002
36, 48	0.95, 1	No	0.791	-0.000	0.002
36, 48	0.84, 1	Yes	0.064	0.008	0.002
36, 48	0.84, 1	No	0.025	-0.000	0.002

Table 6 presents the results for two scenarios where the primary source may be considered similar to the historic source (i.e., (1,1) and (0.95,1)) and for one scenario where the primary source is on the boundary of equivalence (i.e., (0.84,1)). Information sharing with MEMs results in higher rejection rates for declaring equivalence for all scenarios. There is higher power for MEMs in the (1,1) scenario of 97.5% versus 96.8% and in the (0.95,1) scenario of 88.0% versus 79.1%, but also corresponds to an inflated type I error rate of 6.4% versus 2.5% when the primary source is on the equivalence boundary. Bias 0.001 to 0.013 higher across scenarios with information sharing while MSE values were similar across all methods.

These results suggest that information sharing in dose equivalence studies may still be beneficial, but the typical consideration with incorporating external information for increased bias and inflated type I error rates need to be carefully taken into account when choosing the primary analytic strategy. However, in the scenario of similar slopes with (0.95,1) do represent an 8.1% increase in power by incorporating the past trial of data with MEMs.

References

[1] Kaizer, A. M., Koopmeiners, J. S. Hobbs, B. P. (2018). Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics* **19**(2) 169–184.

[2] Kotalik, A., Vock, D. M., Hobbs, B. P. Koopmeiners, J. S. (2022). A group-sequential randomized trial design utilizing supplemental trial data. *Statistics in Medicine* **41**(4) 698–718.