Published in final edited form as:

J Biopharm Stat. 2016 ; 26(4): 672–685. doi:10.1080/10543406.2015.1052473.

ASSESSING THE IMPACT OF SAFETY MONITORING ON THE EFFICACY ANALYSIS IN LARGE PHASE III GROUP SEQUENTIAL TRIALS WITH NON-TRIVIAL SAFETY EVENT RATE

Yanqiu Weng¹, Yuko Y. Palesch², Stacia M. DeSantis³, and Wenle Zhao²

¹Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, USA

²Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

³Division of Biostatistics, University of Texas Health Sciences Center, Houston, Texas, USA

Abstract

In Phase III clinical trials for life-threatening conditions, some serious but expected adverse events, such as early deaths or congestive heart failure, are often treated as the secondary or co-primary endpoint, and are closely monitored by the data and safety monitoring committee. A naïve group sequential design for such a study is to specify univariate statistical boundaries for the efficacy and safety endpoints separately, and then implement the two boundaries during the study, even though the two endpoints are typically correlated. One problem with this naïve design, which has been noted in the statistical literature, is the potential loss of power. In this paper, we develop an analytical tool to evaluate this negative impact for trials with non-trivial safety event rates, particularly when the safety monitoring is informal. Using a bivariate binary power function for the group sequential design with a random effect component to account for subjective decision making in safety monitoring, we demonstrate how, under common conditions, the power loss in the naïve design can be substantial. This tool may be helpful to entities such as the data and safety monitoring committees when they wish to deviate from the pre-specified stopping boundaries based on safety measures.

Keywords

Data and safety monitoring committee; Efficacy; Group sequential design; Power; Safety; Type I error

1. Introduction

Group sequential design (GSD) is the most commonly used statistical approach for data monitoring in Phase III clinical trials with a univariate endpoint (Lan and DeMets, 1983; Friedman et al., 1998; Jennison and Turnbull, 2000; Food and Drug Administration, 2006). For trials with two or more endpoints, such as one primary efficacy endpoint and one

primary or secondary safety endpoint, a naïve GSD specifies separate univariate statistical boundaries for each endpoint, and then implements the two boundaries – propose stopping the trial as soon as one of the boundaries is crossed – even if the two endpoints are correlated (Whitehead, 1999). From a statistical point of view, this may be problematic because the pre-specified statistical operating characteristics for the efficacy boundary from the univariate design can be altered if study termination is triggered by the safety monitoring process. Nevertheless, this is one of the most commonly implemented study designs for monitoring two or more responses in confirmatory Phase III trials (Wittes et al., 2007; Hedenmalm et al., 2008; Ginsberg et al., 2011). The popularity of this design is not only due to its ease of implementation and a lack of available sample size and power calculation software for the bivariate GSDs, but also due to the fact that the use of the naïve method can well protect the study-wise Type I error probability (Whitehead, 1999) irrespective of efficacy-safety correlation and even if data monitoring is flexible or “informal”.

Even though the naïve method protects the Type I error probabilities for the efficacy and safety outcome analyses, the potential cost in power of using this method is appreciable. Whitehead (1999) has noted that using such a naïve method could lead to an underpowered design, but the extent of the loss of power remains an unanswered question (Prentice, 1997). This study aims to assess this negative impact of safety monitoring on efficacy outcome analysis in Phase III trials with non-trivial life-threatening safety event settings, such as in acute neurological trials. In such trials, serious adverse events such as death or congestive heart failure may not be rare or unexpected and are often treated as the co-primary endpoint or the critical secondary endpoint (Torp-Pedersen et al., 1999; Hill et al., 2007). To monitor safety, a guideline consisting of fixed size repeated significance tests or repeated confidence intervals is often proposed, and the safety monitoring may be more frequent, liberal and informal than the efficacy monitoring (Prentice, 1997). One example is the Albumin in Acute Stroke (ALIAS) Part 1 trial (Ginsberg et al., 2011) for assessing the efficacy of human serum albumin versus saline in acute stroke. In the ALIAS Part 1 Trial, correlated efficacy endpoint (*i.e.*, good stroke function recovery vs. bad recovery including death) and safety endpoint (*i.e.*, death within 30 days) were monitored using two separate monitoring boundaries (safety using repeated 95% confidence intervals). The trial was suspended by the Data and Safety Monitoring Committee (DSMC) after one fourth patients were enrolled due to safety concerns. However, it was noted that the frequent and liberal safety monitoring may not only increase the chance of stopping for safety even if there are no really safety issues, but also reduce the chance of claiming success for efficacy. Thus, it was of interest to assess the true power for efficacy when safety monitoring is frequent, liberal and flexible.

Cook and Farewell (1994) proposed a bivariate power function for the GSD with two continuous responses (efficacy and safety) (Cook and Farewell, 1994; Cook, 1996). Weng, Zhao and Palesch (2012) generalized Cook and Farewell’s power function for bivariate binary responses (Weng et al., 2012). However, both methods were based on the assumption that statistical boundaries be strictly adhered to during the trial. To our knowledge, there is no analytical procedure available to calculate the power when data monitoring is informal. For example, the DSMC does not always stop the trial when the stopping criterion for safety is met or their stopping decision for safety is dependent on the observed efficacy evidence. Thus, this paper is to develop an analytical method to estimate power for bivariate GSD that

can account for flexible and partly subjective decision-making by the DSMC. In Section 2, we introduce the proposed approach. In Section 3, we quantify the true Type I error probabilities and power for a hypothetical naïve design similar to the ALIAS Trial. In Section 4, we discuss the application and limitations of the new approach.

2. A general approach to estimate power for a bivariate binary response in GSD

2.1 Notation and formulation of problem

Consider a Phase III group sequential (GS) trial comparing a new treatment (A) with a standard treatment (B). A binary efficacy outcome (E) and a binary safety outcome (S) are of primary interest. The outcome E is defined as treatment success or failure, and the outcome S is whether a subject experiences the pre-specified safety event during the study. Let ρ_j be the correlation of the binary efficacy and safety outcomes in treatment j ($j = A, B$); let P_{jm} be the true event rate for outcome m ($m = E, S$) in treatment j ; let $\delta_m = P_{Am} - P_{Bm}$ represent the treatment effect on outcome m . Let K_m be the number of interim plus final analyses for outcome m , and $t_m(k)$ be the calendar time of the k th analysis, $1 \leq k \leq K_m$. In addition, at some analysis stages, the efficacy and safety responses can both be tested. Thus, let K_{ES} be the number of these stages, $0 \leq K_{ES} \leq \min(K_E, K_S)$. Assume balanced allocation across all the interim analyses, and let $2n_{km}$ denote the number of the randomized patients at the k th analysis of outcome m , $k = 1, \dots, K_m$. To set up a power function for the bivariate GSD, we need the joint distribution of the test statistics across all interim analyses and a decision rule on how to implement the statistical boundaries.

2.2 Joint distribution of the interim test statistics

For large Phase III trials with event rates that are non-trivial, the number of events at the first interim analysis is usually no less than 5 to justify normal approximation of binomial distribution in tests. Suppose at each interim analysis, accumulating efficacy and/or safety data are tested by the one-sided two-proportion Z test. Let Z_{km} be the standardized test statistic for outcome m at stage k , $k = 1, \dots, K_m$, and $\mathbf{Z} = (Z_{1E}, \dots, Z_{K_{EE}}, Z_{1S}, \dots, Z_{K_{SS}})'$. When n_{1m} is large, $\mathbf{Z}$ is approximately distributed as multivariate normal (MVN), $\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z)$, with

$$E(Z_{km}) = \frac{\sqrt{n_{km}}(P_{Am} - P_{Bm})}{\sqrt{P_{Am}Q_{Am} + P_{Bm}Q_{Bm}}},$$

And

$$\text{cov}(Z_{km}, Z_{k'm'}) = \sqrt{\frac{\min(n_{km}, n_{k'm'})}{\max(n_{km}, n_{k'm'})}} \sigma_{mm'}, \quad k(k') = 1, \dots, K_m(K_{m'}), \quad m, m' = E, S,$$

Where $\sigma_{EE} = \sigma_{SS} = 1$, $\sigma_{ES} = \sigma_{SE} = \frac{\rho_A \sqrt{P_{AE} Q_{AE} P_{AS} Q_{AS}} + \rho_B \sqrt{P_{BE} Q_{BE} P_{BS} Q_{BS}}}{\sqrt{P_{AE} Q_{AE} + P_{BE} Q_{BE}} \sqrt{P_{AS} Q_{AS} + P_{BS} Q_{BS}}}$, $Q_{Am} = 1 - P_{Am}$, $Q_{Bm} = 1 - P_{Bm}$, see Appendix 1 or Weng et al., 2012 for detail.

2.3 Inclusion of random effects to account for the subjective safety monitoring

In addition to the statistical evidence, the knowledge/expertise of DSMC members, information external to the trial and/or unexpected events can sometimes influence DSMCs' recommendation. To mimic the subjectivity in the DSMC's decision, we include a random variable, $D_k \sim N(\mu_{Dk}, 1)$ in the power function. Functionally, the random effect D_k models uncertainty from the evidence other than statistical finding in trial conduct and determines whether or not a safety boundary crossing will be adhered to by personnel monitoring the trial at time k using the following rules:

- If $D_k \geq 0$, adhere to the safety guideline and stop the study;
- If $D_k < 0$, override the safety guideline and continue the study.

Let $\mathbf{D} = (D_1, \dots, D_{K_S})'$ and assume that $\mathbf{D}$ is distributed as MVN, $\mathbf{D} \sim \text{MVN}(\boldsymbol{\mu}_D, \mathbf{I}_{K_S})$, where $\boldsymbol{\mu}_D = (\mu_{D1}, \dots, \mu_{DK_S})'$ and $\mathbf{I}_{K_S}$ is an identity matrix. We can specify μ_{Dk} s based on various assumptions of the subjectivity. If the subjectivity in decision-making is assumed arising purely from external information or clinical consideration, μ_{Dk} can be specified as a real number, then the probability of stopping a trial when the safety boundary is crossed is $\Phi(\mu_{Dk})$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. For example, when $\mu_{Dk} = 0$ the stopping probability is 0.5. The choice of μ_{Dk} is based on investigators' assumption or historical reference. On the other hand, the subjective decision of DSMCs can sometimes be biased by the statistical evidence in a real trial. To model this bias, we can let μ_{Dk} depend on test statistics, say, be a function of random variables $Z_{kS} = z_{kS}$ and/or $Z_{kE} = z_{kE}$. For example, let $\mu_{Dk} = z_{kS} - c_{kS}$ for $k = 1, \dots, K_S - 1$, where c_{kS} is the critical value for the safety test statistics at k , and let $\mu_{DK_S} = \infty$ so that the safety rule is always adhered to in the final analysis. This allows us to model the situation that the further Z_{kS} deviates from the boundary in an interim analysis, the more likely the DSMC will stop the trial for safety concern.

Since the decision-making process is assumed to be driven by both objective and subjective considerations, the stopping probabilities are determined by the joint distribution of $\mathbf{Z}$ and $\mathbf{D}$. If $\boldsymbol{\mu}_D$ is specified as a vector of real numbers, $(\mathbf{Z}, \mathbf{D})$ is distributed as multivariate normal,

$$(\mathbf{Z}, \mathbf{D}) \sim \text{MVN} \left(\begin{pmatrix} \boldsymbol{\mu}_Z \\ \boldsymbol{\mu}_D \end{pmatrix}, \begin{pmatrix} \Sigma_Z & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{K_S} \end{pmatrix} \right).$$

If μ_{Dk} depends on $\mathbf{Z}$ only through the observed Z value at the k th interim analysis, say, $Z_{kS} = z_{kS}$, since D_k s are conditionally independent given $\mathbf{Z}$, we can write the joint distribution of $(\mathbf{Z}, \mathbf{D})$, $f(\mathbf{z}, \mathbf{d})$ as,

$$f(\mathbf{z}, \mathbf{d}) = f(\mathbf{z}) \left[\prod_{k=1}^{K_S-1} g_k(d_k | z_{kS}) \right] g_{K_S}(d_{K_S} | z_{K_S S}), \quad (1)$$

where $f(\mathbf{z})$ is the distribution of $\mathbf{Z}$, $g_k(d_k | z_{kS})$ is the distribution of D_k given $Z_{kS} = z_{kS}$, and $g_{K_S}(d_{K_S} | z_{K_S S})$ is the distribution of D_{K_S} given $Z_{K_S S} = z_{K_S S}$ in the final analysis. If we assume that the DSMC strictly adheres to the safety rule in the final analysis, we can simply drop D_{K_S} from the above derivation because there is no random effect on the decision making then. Given the joint distribution of $(\mathbf{Z}, \mathbf{D})$, the Type I error probability and power are calculated using numeric or Monte Carlo integration based on the space defined by statistical boundaries and DSMC's decision making rules.

2.4 Set up the power functions for flexible decision-making in the GS study

When a safety boundary is crossed in a Phase III efficacy trial, several plausible decision-making scenarios might be:

- Scenario 1: The DSMC follows the statistical guideline, and stops the study immediately.
- Scenario 2: The DSMC does not stop the study for safety unless the safety boundary has been crossed in a previous interim analysis.
- Scenario 3: Due to clinical uncertainty or external information, the probability of stopping study is $\Phi(\mu_{Dk})$, which was introduced in Section 2.3.
- Scenario 4: The DSMC looks at the efficacy data. The study will be stopped only if there is a negative trend on efficacy.

Based on the joint distribution of $(\mathbf{Z}, \mathbf{D})$ in Section 2.3, we can assemble the power functions for above scenarios. Let T_{km} be the critical region for $\mathbf{Z}$ and $\mathbf{D}$ for stopping the trial because of outcome m at stage k ; let T_{km}^c be the continuation region; let S_{km} be the sample space of $\mathbf{Z}$ and $\mathbf{D}$ when the trial is stopped for outcome m at stage k (or time $t_m(k)$). S_{km} , T_{km} and T_{km}^c have the relationship

$$S_{km} = \left\{ \mathbf{Z}, \mathbf{D} : T_{km} \cap \bigcap_{C_E(k)} T_{\ell E}^c \cap \bigcap_{C_S(k)} T_{\ell S}^c \right\}, \quad (2)$$

Where $C_E(k) = \{\ell: t_E(\ell) < t_m(k)\}$, $1 \leq k \leq K_E$, and $C_S(k) = \{\ell: t_S(\ell) < t_m(k)\}$, $1 \leq k \leq K_S$.

The marginal power function for outcome m ($m=E, S$) can be expressed as

$$\text{Marginal Power}(m) = \sum_{k=1}^{K_m} \Pr(S_{km}). \quad (3)$$

Although the efficacy and safety responses can be monitored with different frequencies during the study, it is possible that, at some time points, the two responses are both tested, as long as there exists k and k' so that $t_E(k) = t_S(k')$, $1 \leq k \leq K_E$, $1 \leq k' \leq K_S$ (or $K_{ES} > 0$). Let

$S_{kE,k'S}$ be the sample space of $\mathbf{Z}$ and $\mathbf{D}$ for rejecting both hypotheses (efficacy and safety) at stage k for efficacy (or stage k' for safety) and $t_E(k) = t_S(k')$, we have

$$S_{kE,k'S} = \left\{ \mathbf{Z}, \mathbf{D}: T_{kE} \cap T_{k'S} \cap C_E(k) \cap T_{\ell E}^c \cap C_S(k') \cap T_{\ell S}^c \right\}. \quad (4)$$

The probability of rejecting both hypotheses in the study can be express as

$\sum_{B(k,k')} \Pr(S_{kE,k'S})$, where $\mathcal{B}(k,k')$ is the collection of all (k,k') vectors that satisfy $t_E(k) = t_S(k')$, i.e., $\mathcal{B}(k,k') = \{k, k': t_E(k) = t_S(k')\}$. Following this idea, if one defines power as the probability of rejecting the hypothesis of outcome m only during the study, the power for m can be expressed as

$$\text{Power}(m \text{ only}) = \sum_{k=1}^{K_m} \Pr(S_{km}) - \sum_{B(k,k')} \Pr(S_{kE,k'S}), \quad (5)$$

which we call a joint power function. In addition, when the random effect assumption for the decision-making is not required for outcome m , we can simply remove $\mathbf{D}$ from S_{km} and $S_{kE,k'S}$.

As shown in Equations (2) to (5), a key step for assembling the power function under a particular decision-making assumption is to define T_{km} and T_{km}^c . Suppose $\mathbf{c} = (c_{1E}, \dots, c_{KEE}, c_{1S}, \dots, c_{KSS})$ is a one-sided statistical boundary for $\mathbf{Z}$. We present T_{km} and T_{km}^c for Scenarios 1–4 in Table 1. For all scenarios we assume that the DSMC will adhere to the efficacy guideline throughout the study, but they may have certain flexibility in regard to the safety monitoring.

2.5 Power calculation

As presented in Equations (3) and (5) and Table 1, the power calculation is based on S_{km} , a set involving both unions and intersections. For a trial with a modest number of interim looks, $\Pr(S_{km})$ is calculated in a straightforward way composed of two steps: 1) Partition S_{km} into R disjoint subsets, P_{kmr} s, which only involve intersections ($r=1, \dots, R$); 2) Calculate the probabilities of P_{kmr} s by multiple integration and sum these probabilities together,

$$\Pr(S_{km}) = \sum_{r=1}^R \Pr(P_{kmr}).$$

An illustrative example of calculation is provided in Appendix 2. All statistical programming was performed in R version 2.15.1. Monte Carlo simulations were also conducted to ensure that the R function for each scenario is correctly coded (result not provided).

3. An illustrative example

We use a hypothetical example developed from the ALIAS Trial to illustrate the practical significance of our method. A Phase III trial aims to investigate the effectiveness of a neuroprotective drug (human serum albumin or ALB) in acute ischemic stroke. Eligible subjects are randomized to two arms, the ALB treatment arm and the saline (control) arm, with 1:1 allocation. The primary efficacy endpoint is a binary response, *i.e.*, treatment success or failure at 3 months from randomization. Assuming the response rate in the control group as 40%, a sample size of 1,200 is derived to adequately detect a 10% increase ($\delta_E = 0.10$) in the response rate of efficacy on the new treatment with 0.934 power. The calculation of these error probabilities is exclusively based on a one-sided *univariate* O'Brien-Fleming type boundary (overall Type I error 0.025) that includes three interim analyses and one final analysis, and the efficacy analyses are performed following every 300 subjects are assessed. A specific expected serious adverse event is considered as the primary measure of safety. Although it does not govern the sample size, the safety outcome is sequentially monitored by the DSMC and 12 safety analyses are planned after every 100 subjects are assessed using 0.01 level repeated one-sided tests.

The question of interest is, what are the actual Type I error probability and power for the efficacy analysis if the safety monitoring is informal but still can trigger stopping of the study? To show the impact of safety monitoring, we outline six commonly assumed decision-making models of DSMCs when they are facing a crossed safety boundary at k th analysis. They are:

- Scenario 1: stop the study immediately.
- Scenario 2: if it is the first time that the safety boundary is crossed, continue the study, otherwise, stop the study.
- Scenario 3a: stop the study with the probability of 0.50.
- Scenario 3b: stop the study with the probability of $\pi(k)$, where $\pi(k) = \Phi(z_{kS} - c_{kS})$, so that $\pi(k)$ depends on the magnitude of z_{kS} through an increasing function, *e.g.*, $\pi(k) = 0.50$ when $z_{kS} - c_{kS} = 0$ and $\pi(k) \approx 0.98$ when $z_{kS} - c_{kS} = 2.0$.
- Scenario 4a: look at the efficacy data, stop the study only if the standardized test statistic for efficacy at that time point is negative.
- Scenario 4b: look at the efficacy data, continue the study only if the standardized test statistic for efficacy at that time point is greater than $\Phi^{-1}(0.9)$, where $\Phi^{-1}(\cdot)$ is the inverse function of $\Phi(\cdot)$.

We adopt the proposed method to solve these problems. Conceptually, the Type I error probability for efficacy is the probability that one rejects the null hypothesis for efficacy when $\delta_E = 0$; the power for efficacy is the probability that one rejects the null hypothesis of efficacy, when $\delta_E = 0.10$. When calculating these error probabilities, we do not fix δ_S , and instead let it change from 0 to 0.05. The reason we allow such small variation in δ_S is that the exact value of δ_S is usually unknown in advance in most Phase III trials, even if the test drug has been examined in previous early phase studies. Furthermore, the correlation

between the efficacy and safety response is also an unknown parameter in most studies, thus we assess the actual Type I error probability and power for $\delta_S \in [0, 0.05]$ and $\rho_j \in [-0.3, 0.3]$. We use the proposed analytical method to calculate Type I error and power. From Section 2.3, it can be shown that $(\mathbf{Z}, \mathbf{D})$ s in the Scenarios 1, 2, 3a, 4a and 4b are distributed as MVN. Their integral can be computed numerically using the R function, `pmvnorm`, from the `mvtnorm` package (R code is available upon request). The distribution of $(\mathbf{Z}, \mathbf{D})$ in the Scenario 3b is not MVN, see Equation 1. A multi-dimensional Monte Carlo integration is performed to obtain its integral.

The actual Type I error probabilities for the Scenarios 1, 2, 3a and 4a are presented in Table 2, which can be summarized as following:

- The true Type I error probabilities for the efficacy analysis are all preserved at the 0.025 level, which means that ignoring the multiplicity issues between the efficacy and safety boundaries will not inflate the Type I error probability.
- The true Type I error is sensitive to δ_S and decreases significantly as δ_S increases.
- The true Type I error is robust to variation of ρ_j when $\delta_S = 0$, but the robustness is weakened as δ_S increases. The Type I error decreases as ρ increases.
- Different decision-making rules lead to different Type I error probabilities. The Type I error in the Scenario 4a is more robust to variation of ρ_j and δ_S than in other scenarios.

The true power for all six scenarios are presented in Table 3, which can be summarized as following:

- The use of univariate GS boundaries for the bivariate response will cause an overestimate of power (all entries in Table 3 are less than 0.934), whether or not the safety monitoring is formal or informal.
- The true power for efficacy is sensitive to δ_S . For example, when $\delta_E = 0.10$ the true power for efficacy is approximately 0.90 if $\delta_S = 0$, but declines to 0.75 if $\delta_S = 0.03$.
- The power does not change significantly as ρ_j changes, meaning that power is generally robust to variation of ρ_j .
- The decision-making pattern of the DSMC can substantial influence the power; this is illustrated by Figures 1 and 2 which plot the power curves for efficacy and safety under six decision-making scenarios, when $\rho_A = \rho_B = 0$.

In Figures 1 and 2, we find that the formal safety monitoring strategy leads to the lowest power for efficacy, but the highest power for safety. This is because when the safety monitoring is informal, the study is less likely to be stopped for safety (*i.e.*, reduces the power for safety), and thus is more likely to be stopped for efficacy (*i.e.*, increases the power for efficacy). We also note that a higher power curve for efficacy is always paired with a lower power curve for safety. The actual heights of these two curves are related to the boundaries that the investigators specify before the study and the decision-making process that the DSMC follows during the study. Actually, given a pair of statistical boundaries for efficacy and safety plus a decision rule, we can always plot these two curves using our

method. The power curve serves two practical functions: first, to help the investigators choose the appropriate boundaries according to their pre-defined risk-benefit ratio for the new treatment; and second, to alert the DSMC of the impact of their “informal” monitoring process on trial’s operating characteristics.

4. Discussion

We have introduced a unified approach that is able to provide power functions for common safety monitoring scenarios when the event rate is expected to be non-trivial. By a recursive procedure (Armitage et al., 1969), we can use the resultant marginal (*i.e.*, Equation (3)) or joint (*i.e.*, Equation (5)) power function to appropriately power a new trial based on a pre-defined partition of power for efficacy and safety under a specific testing framework (*i.e.*, a specific decision-making rule). Depending on clinical objectives and the disease of interest, the role that a safety outcome plays in Phase III trials differs drastically from trial to trial. For instance, if the event rate is non-trivial, the safety response can be specified as either a co-primary endpoint (Pitt et al., 2003) or a secondary endpoint (Hill et al., 2007; Wittes et al., 2007). Safety monitoring can proceed very formally (Bolland and Whitehead, 2000), but more often, proceeds “informally” as compared to the efficacy monitoring (Wittes et al., 2007; Hedenmalm et al., 2008; Ginsberg et al., 2011). The safety and efficacy hypotheses can be tested simultaneously (Jennison and Turnbull, 1993; Cook and Farewell, 1994; Cook, 1996; Kosorok et al., 2004), or tested under a hierarchical framework (Tang and Geller, 1999; Glimm et al., 2010; Hung and Wang, 2010; Tamhane et al., 2010; Maurer et al., 2011). Although we considered only 6 decision-making models in this paper, by changing the definition of T_{km} and T_{km}^c as introduced in Section 2.4, the method can be extended to more testing frameworks as discussed above.

This study focuses on the bivariate binary response, but the idea easily can be applied to bivariate continuous responses as discussed by Cook and Farewell (1994) and Cook (1996). We use a normal approximation to the binomial distribution to find the joint distribution of the interim test statistics for the bivariate binary efficacy-safety response. Since the normal approximation is only valid when the product of sample size and response rate is large, the proposed method requires that the expected number of events for the first interim analysis is not too small (Tamhane et al., 2012). This requirement is usually met in large confirmatory trials when safety event rate is not too rare. However, when the sample size is small or the response rate is low, the proposed method may not be suitable. An exact method would be needed in this situation to compute the power.

As demonstrated in Figures 1 and 2, the use of two univariate GS boundaries for one primary efficacy outcome and one critical secondary safety outcome will substantially over-estimate power for efficacy as well as safety analyses. This result emphasizes the necessity of bivariate GS designs, even though the safety outcome is a secondary endpoint and the safety monitoring is “informal”. However, one major concern for the application of bivariate GSDs is the robustness of the error probabilities for the efficacy analysis to mis-specification of unknown safety profiles, such as ρ and δ_S . Todd (Todd, 2003) has developed an adaptive bivariate GSD for unknown ρ to preserve the pre-specified type I error probability, but both her paper and this paper have shown that power is generally robust to changes in ρ . In

contrast, it has been demonstrated in this paper that power is extremely sensitive to δ_S . Therefore, an adaptive design that can dynamically learn the information on both δ_S and ρ from data collected as part of the trial is desired. In addition, we have shown that the flexible decision-making process of the DSMC could substantially alter the pre-planned statistical operating characteristics of the study. When interim looks, particularly for safety, are frequent in the trial, setting a prior distribution and then updating the posterior distribution of the mean of the random decision-making component, $\mathbf{D}$, under the Bayesian framework might also be considered for a sample size re-estimation design. Last but not least, since the proposed power calculation method can be used under various decision-making assumptions, it is highly recommended for the DSMC members to use the proposed method to assess the possible impact of their decision-making practice when they are going to override a pre-specified guideline.

Acknowledgments

The authors thank the Editor and three referees for their helpful comments and suggestion.

Funding

This work was supported by a National Institute of Neurological Disorders and Stroke (NINDS) grant, U01 NS054630 and U01 NS059041.

References

- Armitage P, McPherson CK, Rowe BC. Repeated Significance Tests on Accumulating Data. *Journal of the Royal Statistical Society Series A*. 1969; 132:235–244.
- Bolland K, Whitehead J. Formal approaches to safety monitoring of clinical trials in life-threatening conditions. *Statistics in Medicine*. 2000; 19(21):2899–2917. [PubMed: 11042622]
- Cook RJ. Coupled error spending functions for parallel bivariate sequential tests. *Biometrics*. 1996; 52(2):442–450. [PubMed: 8672698]
- Cook RJ, Farewell VT. Guidelines for monitoring efficacy and toxicity responses in clinical trials. *Biometrics*. 1994; 50(4):1146–1152. [PubMed: 7786995]
- Food and Drug Administration. Guidance for Clinical Trial Sponsors: Establishment and Operation of Clinical Trial Data Monitoring Committee. 2006
- Friedman, LM.; Furberg, C.; DeMets, DL. Fundamentals of clinical trials. 3. New York: Springer; 1998.
- Ginsberg MD, Palesch YY, Martin RH, Hill MD, Moy CS, Waldman BD, Yeatts SD, Tamariz D, Ryckborst K. The albumin in acute stroke (ALIAS) multicenter clinical trial: safety analysis of part 1 and rationale and design of part 2. *Stroke*. 2011; 42(1):119–127. [PubMed: 21164127]
- Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. *Stat Med*. 2010; 29(2):219–228. DOI: 10.1002/sim.3748 [PubMed: 19827011]
- Hedenmalm K, Melander H, Alvan G. The conscientious judgement of a DSMB--statistical stopping rules re-examined. *European Journal of Clinical Pharmacology*. 2008; 64(1):69–72. [PubMed: 18000658]
- Hill MD, Moy CS, Palesch YY, Martin R, Dillon CR, Waldman BD, Patterson L, Mendez IM, Ryckborst KJ, Tamariz D, Ginsberg MD. The albumin in acute stroke trial (ALIAS); design and methodology. *International Journal of Stroke*. 2007; 2(3):214–219. [PubMed: 18705947]
- Hung HMJ, Wang SJ. Challenges to multiple testing in clinical trials. *Biometrical Journal*. 2010; 52(6): 747–756. [PubMed: 20589856]
- Jennison C, Turnbull BW. Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety endpoints. *Biometrics*. 1993; 49(3):741–752. [PubMed: 8241370]

- Jennison, C.; Turnbull, BW. Group sequential methods with applications to clinical trials. Boca Raton: Chapman & Hall/CRC; 2000.
- Kosorok MR, Yuanjun S, DeMets DL. Design and analysis of group sequential clinical trials with multiple primary endpoints. *Biometrics*. 2004; 60(1):134–145. [PubMed: 15032783]
- Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983; 70(3):659–663.
- Maurer W, Glimm E, Bretz F. Multiple and Repeated Testing of Primary, Coprimary, and Secondary Hypotheses. *Statistics in Biopharmaceutical Research*. 2011; 3(2):336–352.
- Pitt B, Remme W, Zannad F, Neaton J, Martinez F, Roniker B, Bittman R, Hurley S, Kleiman J, Gatlin M. Eplerenone Post-Acute Myocardial Infarction Heart Failure E Survival Study I. Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. *New England journal of medicine*. 2003; 348(14):1309–1321. [PubMed: 12668699]
- Prentice RL. Discussion: On the role and analysis of secondary outcomes in clinical trials. *Controlled Clinical Trials*. 1997; 18(6):561–567.
- Tamhane AC, Mehta CR, Liu L. Testing a primary and a secondary endpoint in a group sequential design. *Biometrics*. 2010; 66(4):1174–1184. [PubMed: 20337631]
- Tamhane AC, Wu Y, Mehta CR. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. *Statistics in Medicine*. 2012; 31(19):2027–2040. [PubMed: 22729929]
- Tang DI, Geller NL. Closed testing procedures for group sequential clinical trials with multiple endpoints. *Biometrics*. 1999; 55(4):1188–1192. [PubMed: 11315066]
- Todd S. An adaptive approach to implementing bivariate group sequential clinical trial designs. *Journal of Biopharmaceutical Statistics*. 2003; 13(4):605–619. [PubMed: 14584711]
- Torp-Pedersen C, Moller M, Bloch-Thomsen PE, Kober L, Sandoe E, Egstrup K, Agner E, Carlsen J, Videbaek J, Marchant B, Camm AJ. Dofetilide in patients with congestive heart failure and left ventricular dysfunction. Danish Investigations of Arrhythmia and Mortality on Dofetilide Study Group. *New England Journal of Medicine*. 1999; 341(12):857–865. [PubMed: 10486417]
- Weng Y, Zhao W, Palesch Y. Impact of safety monitoring on error probabilities of binary efficacy outcome analyses in large phase III group sequential trials. *Pharm Stat*. 2012; 11(4):310–317. [PubMed: 22589042]
- Whitehead J. On being the statistician on a Data and Safety Monitoring Board. *Statistics in Medicine*. 1999; 18(24):3425–3434. [PubMed: 10611616]
- Wittes J, Barrett-Connor E, Braunwald E, Chesney M, Cohen HJ, Demets D, Dunn L, Dwyer J, Heaney RP, Vogel V, Walters L, Yusuf S. Monitoring the randomized trials of the Women's Health Initiative: the experience of the Data and Safety Monitoring Board. *Clinical Trials*. 2007; 4(3): 218–234. [PubMed: 17715247]

Appendix 1. The joint distribution of the interim test statistics for a bivariate binary response

We will first develop the joint distribution of the interim test statistics when the binary efficacy and safety responses are tested simultaneously at each stage. This can be easily generalized to the situation where the safety and efficacy responses are tested with different frequencies, as shown at the end of this appendix.

Suppose there are K interim stages. At each stage, both the efficacy and safety responses are tested. Let X_{ijm} be the observed number of events for outcome m in treatment j at stage i , $m = E, S$, $j = A, B$, $i = 1, 2, \dots, K$, and Y_{im} be their difference between the two treatment arms, $Y_{im} = X_{iAm} - X_{iBm}$. Assume the binary efficacy and safety endpoints follow the Bernoulli distribution with probabilities P_{jE} and P_{jS} in treatment j , and let $2n_i$ denote the number of the randomized patients at stage i . When the condition of normal approximation to binomial

distribution is met at the first interim analysis, Y_{im} is approximately distributed as normal (Tamhane et al., 2012),

$$Y_{im} \sim N(n_i \delta_m, n_i (P_{Am} Q_{Am} + P_{Bm} Q_{Bm})),$$

where $\delta_m = P_{Am} - P_{Bm}$ and $Q_{jm} = 1 - P_{jm}$. Furthermore, given the correlation coefficient of the efficacy and safety responses, ρ_j , in treatment j , we can obtain the covariance of Y_{jE} and Y_{jS} ,

$$\text{cov}(Y_{jE}, Y_{jS}) = \rho_j n_j \sqrt{P_{jE} Q_{jE} P_{jS} Q_{jS}} + \rho_j n_j \sqrt{P_{jE} Q_{jE} P_{jS} Q_{jS}}.$$

Let $\mathbf{Y}_j = (Y_{jE}, Y_{jS})'$, then $\mathbf{Y}_j$ is distributed as bivariate normal, $\mathbf{Y}_j \sim \text{BVN}(n_j \boldsymbol{\delta}, n_j \boldsymbol{\Sigma}_0)$, where

$$\boldsymbol{\delta} = \begin{pmatrix} \delta_E \\ \delta_S \end{pmatrix}, \quad (6)$$

and

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \frac{P_{jE} Q_{jE} + P_{jS} Q_{jS}}{\rho_j \sqrt{P_{jE} Q_{jE} P_{jS} Q_{jS}} + \rho_j \sqrt{P_{jE} Q_{jE} P_{jS} Q_{jS}}} & \frac{\rho_j \sqrt{P_{jE} Q_{jE} P_{jS} Q_{jS}} + \rho_j \sqrt{P_{jE} Q_{jE} P_{jS} Q_{jS}}}{P_{jE} Q_{jE} + P_{jS} Q_{jS}} \end{pmatrix}. \quad (7)$$

Because of the independent increment structure of $\{\mathbf{Y}_j\}$, it can be shown that

$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)' \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} n_1 \boldsymbol{\delta} \\ n_2 \boldsymbol{\delta} \\ \vdots \\ n_K \boldsymbol{\delta} \end{pmatrix} \quad (8)$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} n_1 \boldsymbol{\Sigma}_0 & n_1 \boldsymbol{\Sigma}_0 & \cdots & n_1 \boldsymbol{\Sigma}_0 \\ n_1 \boldsymbol{\Sigma}_0 & n_2 \boldsymbol{\Sigma}_0 & \cdots & n_2 \boldsymbol{\Sigma}_0 \\ \vdots & \vdots & \ddots & \vdots \\ n_1 \boldsymbol{\Sigma}_0 & n_2 \boldsymbol{\Sigma}_0 & \cdots & n_K \boldsymbol{\Sigma}_0 \end{pmatrix}. \quad (9)$$

Using Equations (6) to (9), it is straightforward to standardize $\mathbf{Y}$ to $\mathbf{Z}$:

$$\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z),$$

with

$$E(Z_{im}) = \frac{\sqrt{n_i}(P_{Am} - P_{Bm})}{\sqrt{P_{Am}Q_{Am} + P_{Bm}Q_{Bm}}}, \quad (10)$$

and

$$\text{cov}(Z_{im}, Z_{i'm'}) = \sqrt{\frac{\min(n_i, n_{i'})}{\max(n_i, n_{i'})}} \sigma_{mm'}, \quad i(i')=1, \dots, K, \quad m, m'=E, S, \quad (11)$$

where $\sigma_{EE} = \sigma_{SS} = 1$, and $\sigma_{ES} = \sigma_{SE} = \frac{\rho_A \sqrt{P_{AE}Q_{AE}P_{AS}Q_{AS}} + \rho_B \sqrt{P_{BE}Q_{BE}P_{BS}Q_{BS}}}{\sqrt{P_{AE}Q_{AE} + P_{BE}Q_{BE}} \sqrt{P_{AS}Q_{AS} + P_{BS}Q_{BS}}}$. Furthermore, when the numbers of the interim analyses on the efficacy and safety responses are different, say, there are K_E analyses on the efficacy response and K_S analyses on the safety response, the expected value of Z_{km} for the k th analyses of outcome m become

$$E(Z_{km}) = \frac{\sqrt{n_{km}}(P_{Am} - P_{Bm})}{\sqrt{P_{Am}Q_{Am} + P_{Bm}Q_{Bm}}}, \quad k=1, 2, \dots, K_m \quad (12)$$

and the covariance of $Z_{km}, Z_{k'm'}$ for $k(k')=1, \dots, K_E(K_S)$ is

$$\text{cov}(Z_{km}, Z_{k'm'}) = \sqrt{\frac{\min(n_{km}, n_{k'm'})}{\max(n_{km}, n_{k'm'})}} \sigma_{mm'}. \quad (13)$$

It can be found that Equation (12) and (13) are directly developed from Equation (10) and (11) by substituting n_i and $n_{i'}$ with n_{km} and $n_{k'm'}$.

Appendix 2. A power calculation example

Suppose a study have three interim analyses and one final analysis for the efficacy and safety responses, and assume at each interim stage, efficacy and safety hypotheses are tested simultaneously. We now demonstrate the calculation of $\Pr(S_{3E})$, the probability for rejecting the null hypothesis of efficacy at the third interim analysis, under Scenarios 1–4.

1. Scenario 1

The DSMC strictly follows the statistical guideline.

When the DSMC strictly follows the guideline, S_{3E} itself is an intersection set,

$$S_{3E} = P_{3E1} = \{\mathbf{Z}: Z_{3E} \geq c_{3E}, Z_{2E} < c_{2E}, Z_{1E} < c_{1E}, Z_{2S} < c_{2S}, Z_{1S} < c_{1S}\}.$$

By multiple integration, we can directly derive $\Pr(P_{3E1})$, which is equivalent to $\Pr(S_{3E})$.

2. Scenario 2

The DSMC does not stop the study for safety unless the safety boundary is crossed twice.

To allow the study to continue to the third interim analysis, no efficacy boundary crossing and at most one safety boundary crossing could occur in the previous two interim analyses. Therefore, we can enumerate 3 mutually exclusive and exhaustive intersection subsets for $\mathbf{Z}$, they are:

- a. No safety boundary crossing

$$P_{3E1} = \{\mathbf{Z}: Z_{3E} \geq c_{3E}, Z_{2E} < c_{2E}, Z_{1E} < c_{1E}, Z_{2S} < c_{2S}, Z_{1S} < c_{1S}\};$$

- b. The safety boundary is crossed at the second interim look

$$P_{3E2} = \{\mathbf{Z}: Z_{3E} \geq c_{3E}, Z_{2E} < c_{2E}, Z_{1E} < c_{1E}, Z_{2S} \geq c_{2S}, Z_{1S} < c_{1S}\};$$

- c. The safety boundary is crossed at the first interim look

$$P_{3E3} = \{\mathbf{Z}: Z_{3E} \geq c_{3E}, Z_{2E} < c_{2E}, Z_{1E} < c_{1E}, Z_{2S} < c_{2S}, Z_{1S} \geq c_{1S}\}.$$

We can calculate $\Pr(P_{3Er})$ for $r=1, 2, 3$ by numeric integration, and derive S_{3E} by

$$\Pr(S_{3E}) = \sum_{r=1}^3 \Pr(P_{3Er}).$$

3. Scenario 3

When the safety boundary is crossed, the probability of study stopping is $\Phi(\mu_{Dk})$.

This scenario mimics the situation when there is uncertainty on clinical evidences, which makes the decision-making of the DSMC partly subjective. As mentioned previously, the power in this case is jointly determined by $\mathbf{Z}$ and $\mathbf{D}$. From Table 1 we can see that, the study will be stopped at stage i unless $\{Z_{iE} < c_{iE}, Z_{iS} < c_{iS}\}$ or $\{Z_{iE} < c_{iE}, Z_{iS} \geq c_{iS}, D_i < 0\}$, for $i = \dots, 1, k-1$. Since at each stage there exist two options for (Z_{iS}, D_i) to allow the trial to continue, we can enumerate 2^{k-1} mutually exclusive and exhaustive intersection subsets of $(\mathbf{Z}, \mathbf{D})$ for the study stopped at stage k . For $k=3$, we have 4 such subsets:

- $P_{3E1} = \{\mathbf{Z}, \mathbf{D} : Z_{3E} \geq c_{3E}, Z_{2E} < c_{2E}, Z_{1E} < c_{1E}, Z_{2S} < c_{2S}, Z_{1S} < c_{1S}\},$
- $P_{3E2} = \{\mathbf{Z}, \mathbf{D} : Z_{3E} \geq c_{3E}, Z_{2E} < c_{2E}, Z_{1E} < c_{1E}, Z_{2S} \geq c_{2S}, D_2 < 0, Z_{1S} < c_{1S}\},$

- $P_{3E3} = \{\mathbf{Z}, \mathbf{D} : Z_{3E} \quad c_{3E} Z_{2E} < c_{2E} Z_{1E} < c_{1E} Z_{2S} < c_{2S} Z_{1S} \quad c_{1S}, D_1 < 0\},$
- $P_{3E4} = \{\mathbf{Z}, \mathbf{D} : Z_{3E} \quad c_{3E} Z_{2E} < c_{2E} Z_{1E} < c_{1E} Z_{2S} \quad c_{2S}, D_2 < 0, Z_{1S} \quad c_{1S}, D_1 < 0\}.$

The calculation of $\Pr(S_{3E})$ is similar as that in Scenario 2.

4. Scenario 4

When the safety boundary is crossed, the DSMC will first look at the efficacy data. The study will be stopped if there is a negative trend for the test statistic of efficacy.

As suggested in Table 1, when the stopping decision for safety concern relies on efficacy, the study will be stopped at stage i unless $\{Z_{iE} < c_{iE}, Z_{iS} < c_{iS}\}$ or $\{a_{iE} < Z_{iE} < c_{iE}, Z_{iS} \quad c_{iS}\}$ for $i = \dots, 1, k-1$ and $a_{iE} < c_{iE}$. Since there are two options for (Z_{iE}, Z_{iS}) at each stage to allow continuation, we can enumerate 2^{k-1} mutually exclusive and exhaustive intersection subsets of (Z_{iE}, Z_{iS}) for a study stopped at stage k . For $k=3$, we have four such subsets:

- $P_{3E1} = \{\mathbf{Z} : Z_{3E} \quad c_{3E} Z_{2E} < c_{2E} Z_{1E} < c_{1E} Z_{2S} < c_{2S} Z_{1S} < c_{1S}\},$
- $P_{3E2} = \{\mathbf{Z} : Z_{3E} \quad c_{3E}, a_{2E} < Z_{2E} < c_{2E}, Z_{2S} \quad c_{2S}, Z_{1E} < c_{1E}, Z_{1S} < c_{1S}\},$
- $P_{3E3} = \{\mathbf{Z} : Z_{3E} \quad c_{3E}, Z_{2E} < c_{1E}, Z_{1S} < c_{2S}, a_{1E} < Z_{1E} < c_{1E}, Z_{1S} \quad c_{1S}\},$
- $P_{3E4} = \{\mathbf{Z} : Z_{3E} \quad c_{3E}, a_{2E} < Z_{2E} < c_{2E}, Z_{2S} \quad c_{2S}, a_{1E} < Z_{1E} < c_{1E}, Z_{1S} \quad c_{1S}\}.$

The calculation procedure is similar as those in Scenarios 2 and 3.

Figure 1.
Marginal power curves of the efficacy analysis

Figure 2.
Marginal power curves of the safety analysis

Table 1

The stop and continuation regions of Z and D for Scenarios 1–4

Scn	T_{kE}	T_{kE}^c	T_{kS}	T_{kS}^c
1	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kS}; Z_{kS} < c_{KS}\}$	$\{Z_{kS}; Z_{kS} < c_{KS}\}$
2	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kS}; Z_{gS}; Z_{kS} < c_{KS}; Z_{gS} < c_{gS}; \exists g < k\}$	$\{Z_{kS}; Z_{kS} < c_{KS}\} \cup \{Z_{kS}; Z_{gS}; Z_{kS} < c_{KS}; Z_{gS} < c_{gS}; \forall g < k\}$
3	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kS}; D_{kS}; Z_{kS} < c_{KS}; D_{kS} < 0\}$	$\{Z_{kS}; Z_{kS} < c_{KS}\} \cup \{Z_{kS}; D_{kS}; Z_{kS} < c_{KS}; D_{kS} < 0\}$
4 [‡]	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kE}; Z_{kE} < c_{KE}\}$	$\{Z_{kS}; Z_{k'E}^*; Z_{kS} \geq c_{kS}, Z_{k'E}^* < a_{k'E}\}$	$\{Z_{kS}; Z_{kS} < c_{kS}\} \cup \{Z_{kS}; Z_{k'E}^*; Z_{kS} \geq c_{kS}, Z_{k'E}^* \geq a_{k'E}\}$

 T_{kE} : the stop region for efficacy at stage k ; T_{kE}^c : the continuation region for efficacy at stage k ; T_{kS} : the stop region for safety at stage k ; T_{kS}^c : the continuation region for safety at stage k ;[‡]: $Z_{k'E}^*$ is the test statistics for efficacy at k th safety analysis, $a_{k'E}$ is a cut-off value to define the negative trend for $Z_{k'E}^*$

Table 2

Actual Type I error for efficacy

ρ^*	Scenario	δ_0						
		0	0.01	0.02	0.03	0.04	0.05	
-0.3	1	0.025	0.024	0.024	0.022	0.021	0.018	
	2	0.025	0.025	0.024	0.023	0.022	0.020	
	3a	0.025	0.025	0.024	0.023	0.022	0.020	
	4a	0.025	0.025	0.025	0.025	0.024	0.024	
0	1	0.024	0.023	0.021	0.019	0.016	0.013	
	2	0.024	0.024	0.023	0.021	0.018	0.015	
	3a	0.024	0.024	0.022	0.020	0.018	0.015	
	4a	0.025	0.025	0.025	0.025	0.024	0.024	
0.3	1	0.022	0.02	0.017	0.014	0.01	0.007	
	2	0.023	0.022	0.019	0.016	0.013	0.009	
	3a	0.023	0.021	0.019	0.016	0.012	0.009	
	4a	0.025	0.025	0.025	0.025	0.025	0.024	

* $\rho A = \rho B = \rho$

Table 3Actual marginal power for efficacy given $\delta_E = 0.10$

ρ^*	Scenario	δ_S						
		0	0.01	0.02	0.03	0.04	0.05	
-0.3	1	0.899	0.869	0.821	0.752	0.663	0.562	
	2	0.918	0.899	0.864	0.811	0.737	0.647	
	3a	0.911	0.889	0.851	0.795	0.720	0.631	
	3b	0.908	0.886	0.843	0.784	0.705	0.609	
	4a	0.928	0.925	0.922	0.917	0.911	0.904	
	4b	0.914	0.902	0.883	0.859	0.829	0.795	
	1	0.899	0.870	0.823	0.754	0.665	0.560	
	2	0.918	0.900	0.868	0.817	0.745	0.654	
0	3a	0.911	0.890	0.854	0.800	0.726	0.635	
	3b	0.909	0.885	0.847	0.783	0.709	0.615	
	4a	0.932	0.931	0.929	0.927	0.924	0.920	
	4b	0.922	0.914	0.902	0.885	0.863	0.834	
	1	0.901	0.873	0.828	0.760	0.669	0.558	
	2	0.920	0.905	0.876	0.828	0.756	0.663	
	3a	0.913	0.893	0.860	0.808	0.735	0.642	
	3b	0.911	0.889	0.851	0.796	0.715	0.616	
0.3	4a	0.934	0.934	0.933	0.933	0.932	0.930	
	4b	0.930	0.926	0.920	0.910	0.897	0.877	

* $\rho_A = \rho_B = \rho$