

# PERFORMANCE AND SAMPLE SIZE REQUIREMENTS OF BAYESIAN METHODS FOR BINARY OUTCOMES IN FIXED-DOSE COMBINATION DRUG STUDIES

Melinda M. Holt<sup>\*1</sup>

James D. Stamey<sup>2</sup>

John W. Seaman, Jr.<sup>2</sup>

Dean M. Young<sup>2</sup>

<sup>1</sup>*Department of Mathematics and Statistics, Sam Houston State University, Huntsville, TX 77341-2206, USA*

<sup>2</sup>*Department of Statistical Science, Baylor University, Waco, TX, 76798-7140, USA*

*\*mholt@shsu.edu*

*We use a Bayesian analysis to address two common issues in clinical trials designed to study fixed-dose drug combinations. One is to determine efficacy, which is achieved by demonstrating superiority of the combination over its components. The other is to identify synergistic joint action within the combination. The approach described here does not require knowledge of the dose-response relationships of the components or large sample approximations. We provide a procedure to estimate sample size in this context. In addition, we explore the performance of the Bayesian procedure in situations where existing methods are known to perform poorly.*

**Key words:** Superiority; synergy; Markov chain Monte Carlo; coverage; power; sample size; beta-binomial

## 1. INTRODUCTION

In fixed-dose clinical trials, one frequent goal is to determine whether or not the combination of two or more drugs is efficacious, that is, superior to its components. Another is to determine whether or not drugs interact in combination. We will propose a simple Bayesian analysis that can be applied to either question. Thus we will provide a Bayesian analog to use of the popular Min test, as applied to the question of superiority by Wang and Hung (1997) and to synergy by Laska, Meisner, and Siegel (1994).

Although we will describe the case of two component drugs, each of the results presented here can easily be extended to combinations of three or more drugs.

Here we assume that a safe dose of interest has been determined for the component drugs and the mixture. The proposed method does not require knowledge of the dose-response relationship but can utilize it, if available, in the formation of a joint prior distribution. Furthermore, the Bayesian approach formally incorporates other available prior information, such as estimates of relative potency of the two drugs, and allows direct estimation of the degree of either superiority or synergy. As Bayesian analyses become more widely accepted in the study of clinical trials, such an approach to this problem is timely. For further discussion of the role of Bayesian procedures in drug development, we refer the reader to the recent workshop co-sponsored by the U.S. Food and Drug Administration and Johns Hopkins University (2004) and the *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials—Draft Guidance for Industry and FDA Staff* (FDA, 2006). We provide a procedure for Bayesian sample-size estimation in this context and explore operational characteristics using various prior structures and sample sizes.

The remainder of the paper is organized as follows. In Section 2, we discuss the difference between superiority and synergy, as well as existing procedures to test for them. In Section 3, we describe and illustrate selected prior structures, as well as associated tests and interval estimates. We present methods of sample-size determination in Section 4 and performance studies in Section 5. We conclude with a brief discussion in Section 6.

## 2. EXISTING TESTS

### 2.1 The Problem

The basic problem can be described as follows. Subjects receive one of two drugs or a combination of the two. Suppose that  $n_i$  subjects receive drug  $i$ , with the  $j^{\text{th}}$  subject having response  $y_{ij}$  for  $i = 1, 2, 3$  and  $j = 1, 2, \dots, n_i$ . One treatment, say drug 3, is a combination of the other two. Here we consider trials that result in binary responses so that  $y_{ij} = 1$  if a subject responds and  $y_{ij} = 0$  otherwise. Let  $p_i$  denote the proportion of the population responding to drug  $i$ ,  $i = 1, 2, 3$ . Typical designs in fixed-dose clinical trials yield responses that are distributed binomially. Thus, we define  $r_i \equiv \sum_{j=1}^{n_i} y_{ij}$ ,  $i = 1, 2, 3$  and assume that the  $r_i$ 's are independent so that the likelihood is

$$f(\mathbf{r}/\mathbf{p}) = \prod_{i=1}^3 \binom{n_i}{r_i} p_i^{r_i} (1-p_i)^{n_i-r_i},$$

where  $\mathbf{p} = (p_1, p_2, p_3)$  and  $\mathbf{r} = (r_1, r_2, r_3)$ .

### 2.2 Superiority

According to the FDA (21 CFR 300.50), to be approved, a combination drug must be efficacious, or *superior*, to each of its components. In this situation, let  $(x_1, 0)$  represent the applied dose of drug 1, while drug 2 is applied at dose  $(0, x_2)$  and the new drug is applied at dose  $(x_1, x_2)$ . Wang and Hung (1997) develop several large-sample tests for superiority based on the following hypotheses:

$$H_0: \theta \leq 0 \text{ vs. } H_1: \theta > 0, \quad (2.1)$$

where  $\theta = p_3(x_1, x_2) - \max[p_1(x_1, 0), p_2(0, x_2)]$ . They do so by performing two separate two-sample  $Z$ -tests comparing  $p_3$  to  $p_1$  and  $p_3$  to  $p_2$ . If both tests reject at the chosen  $\alpha$

level, then  $H_0$  is rejected, yielding an  $\alpha$ -level test for the hypotheses in (2.1), often called the Min test. For more discussion of how the Min test preserves the  $\alpha$ -level, see Lehman (1952) or Berger (1982). Based on sample-size estimation and a power analysis, Wang and Hung recommend use of these procedures when group sample sizes are at least 20 and for response probabilities between 0.20 and 0.80. Thus, one goal of the Bayesian method proposed herein is to provide a viable alternative to the Min test for small samples and response probabilities below 0.20.

### 2.3 Synergy

A closely related question is that of joint action. The joint action of two drugs is said to be *synergistic* if the proportion of responses to the combination is greater than expected under some chosen model of non-interaction and *antagonistic* if the proportion is less than expected. The choice of a model of non-interaction has been a source of considerable debate. For a thorough discussion of this issue, see Greco, Bravo, and Parsons (1995). Here we consider dose additivity as the non-interaction model so that

$$\frac{x_1}{x_{1e}} + \frac{x_2}{x_{2e}} = 1,$$

where  $x_{ie}$  is the dose of drug  $i$ , for  $i = 1, 2$ , that results in a desired response level, e.g.  $ED_{50}$ . Laska, Meisner, and Siegel (1994) have considered the question of synergy based on this model by adapting the hypotheses in (2.1) so that  $\theta = p_3(x_1, x_2) - \max[p_1(z_1, 0), p_2(0, z_2)]$ ,  $z_1 = x_1 + x_2/\hat{\rho}$ ,  $z_2 = \hat{\rho}x_1 + x_2$ , and  $\hat{\rho}$  estimates the relative potency  $\rho = x_{2e}/x_{1e}$ . To test for antagonism, reverse the inequalities in (2.1). Note that to test for synergy simply implies comparing the combination drug to its components when the components are given at increased dose levels than those given when testing superiority. As discussed

by Laska, Meisner, and Siegel, the power of this procedure depends heavily on the *a priori* estimate of  $\rho$ . Thus, another goal of the Bayesian procedure is to model uncertainty with respect to  $\rho$  and to consider the effects of prior selection.

### 3. THE BAYESIAN APPROACH

#### 3.1 Prior Structures

In this section, we consider three approaches for the construction of a joint prior on  $(p_1, p_2, p_3)$ . Suppose priors  $\pi(p_1)$  and  $\pi(p_2)$  are obtained for  $p_1$  and  $p_2$ , respectively. Now, if the combination is thought to be superior to each of its components, then knowledge of  $p_3$  should be dependent in some way on knowledge of  $p_1$  and  $p_2$ . Therefore, the prior on  $p_3$ ,  $\pi(p_3 | p_1, p_2)$ , must be conditional on  $p_1$  and  $p_2$ . The joint prior for  $p_1, p_2$ , and  $p_3$  is then

$$\pi(\mathbf{p}) \equiv \pi(p_1, p_2, p_3) = \pi(p_1) \pi(p_2 | p_1) \pi(p_3 | p_1, p_2).$$

The explicit form of the dependency, however, will not likely be known (see Hung, 1996; Laska, Meisner, and Tang, 1997; and Miller and Seaman, 1998).

The beta family of distributions is commonly used for both non-informative and informative priors on probabilities in clinical trials. Denote by  $Beta(x, y)$  a beta distribution on  $[0, 1]$  with shape parameters  $x$  and  $y$  and density

$$b(u | x, y) = \frac{\Gamma(x + y)}{\Gamma(x)\Gamma(y)} u^{x-1} (1-u)^{y-1}, \quad 0 \leq u \leq 1. \quad \text{Let } \pi(p_i) = b(u | \alpha_i, \beta_i), \quad i = 1, 2, \text{ where}$$

$\alpha_i$  and  $\beta_i$  are determined by the experimenter. Given  $\alpha_i$  and  $\beta_i$ , the prior for  $p_3$ ,  $\pi(p_3 | \alpha_1, \beta_1, \alpha_2, \beta_2) \equiv \pi(p_3 | \alpha_3, \beta_3) = b(p_3 | \alpha_3, \beta_3)$ , can then be elicited. A variety of methods is available for eliciting beta distributions to represent the experimenter's knowledge—see, for example, Berry and Stangl (1996, pp. 12-16) and Thall, Millikan, Mueller, and Lee

(2003). In addition, Su (2006) has developed a software program called BetaBuster with a graphical user interface that greatly facilitates the elicitation of beta distributions.

One may require an analysis which uses no *a priori* belief in the combination's superiority. This may be necessary, for example, in the presentation of results to regulatory agencies. In that case, it might be reasonable to model  $p_1$ ,  $p_2$ , and  $p_3$  as exchangeable variates with diffuse priors, allowing maximal influence from the data. For example, one might use independent Jeffreys priors for the  $p_i$ 's; that is,  $\alpha_i = \beta_i = 0.5$ ,  $i = 1, 2, 3$ .

As an example of an informative joint prior, suppose our knowledge of  $p_3$  depends on  $p_1$  and  $p_2$  only through the elicited shape parameters of their priors. For example, one can model the supposed superiority of the combination by taking the mean of the prior for  $p_3$  to be larger than that of either  $p_1$  or  $p_2$ . Thus, one might choose  $\pi(p_3 | \alpha_1, \beta_1, \alpha_2, \beta_2) = b(p_3 | \alpha_3, \beta_3)$  so that  $\alpha_3/(\alpha_3 + \beta_3) > \max\{\alpha_1/(\alpha_1 + \beta_1), \alpha_2/(\alpha_2 + \beta_2)\}$ . Similar choices might be based on the mode or percentiles. Using this prior, the joint posterior is

$$\pi(\mathbf{p} | \mathbf{r}) = \prod_{i=1}^3 b(p_i | r_i + \alpha_i, n_i - r_i + \beta_i). \quad (3.1)$$

Finally, consider the situation in which prior knowledge of each component dose-response curve exists in the form of a dose-response model developed using frequentist methods from previous trials. For example, as in Simon and Freedman (1997), suppose  $p_1$  and  $p_2$  are represented by the logistic regression model

$$p_i = \frac{1}{1 + \exp[-(\gamma_{0i} + \gamma_i d_i)]} \quad (3.2)$$

for  $i = 1, 2$ , where  $d_i$  is a function of dose of drug  $i$ . Suppose further that we have maximum likelihood estimates  $\hat{\boldsymbol{\gamma}} \equiv (\hat{\gamma}_{01}, \hat{\gamma}_{02}, \hat{\gamma}_1, \hat{\gamma}_2)$ . Using these estimates, we may then place informative normal priors on  $\boldsymbol{\gamma} \equiv (\gamma_{01}, \gamma_{02}, \gamma_1, \gamma_2)$  thereby inducing priors on  $p_1$  and  $p_2$  for use in the combination drug study. Placing a relatively non-informative beta prior on  $p_3$  completes the joint prior distribution. We illustrate this approach in Section 6.2.

### 3.2 Computation

Our first goal is to compute  $\Pr(\theta > 0 \mid \mathbf{r})$ . For the structure described in (3.1), we can calculate

$$\Pr(\theta > 0 \mid \mathbf{r}) = \Pr(p_1 < p_2 < p_3 \mid \mathbf{r}) + \Pr(p_2 < p_1 < p_3 \mid \mathbf{r}). \quad (3.3)$$

By using the well-known relationship between the incomplete beta and the binomial (see, for example, Abramowitz and Stegan, 1972, p. 263), we determine that

$$\Pr(p_1 < p_2 < p_3 \mid \mathbf{r}) = \int_0^1 \int_0^1 \left( \text{Bin}(p_2 \mid a_1, b_1) \prod_{j=2}^3 b(p_j \mid \alpha_j + r_j, \beta_j + n_j - r_j) \right) dp_2 dp_3, \quad (3.4)$$

where  $\text{Bin}(p_2 \mid a_1, b_1) = \sum_{j=a_1}^{b_1} \binom{b_1}{j} p_2^j (1-p_2)^{b_1-j}$ ,  $a_1 = \alpha_1 + r_1$  and  $b_1 = \beta_1 + n_1 + \alpha_1 - 1$ .

Calculate  $\Pr(p_2 < p_1 < p_3 \mid \mathbf{r})$  similarly by simply reversing subscripts 1 and 2 in (3.4).

Thus, for the posterior distribution described in (3.1), calculation of  $\Pr(\theta > 0 \mid \mathbf{r})$  requires only standard numerical integration routines. For other prior structures in Section 3.1, calculation of  $\Pr(\theta > 0 \mid \mathbf{r})$  is accomplished via Markov chain Monte Carlo (MCMC) simulation.

Suppose that, in addition, researchers are interested in establishing a specified level of improvement of the combination over its components. The procedure described here can

provide an estimate of  $\theta$ , as well as an error measurement. This is accomplished using MCMC methods and will be demonstrated below.

### 3.3 Examples

To illustrate the procedures described above, we use the example presented by Wang and Hung (1997). Consider a fixed-dose randomized clinical trial designed to evaluate the combination of two antihypertensive drugs. The doses of drugs 1 and 2 are the same as those applied in combination; thus, this is a test for superiority of the combination. The binary response of interest is a decrease of 10 mmHg or greater from the baseline in sitting diastolic blood pressure or a post treatment blood pressure below 90 mmHg. Wang and Hung have stated that the sample sizes are approximately 34. We take them to be exactly 34 so that the data are  $\hat{p}_1 = 0.38$ ,  $\hat{p}_2 = 0.32$ , and  $\hat{p}_3 = 0.68$ .

Using  $Beta(0.5, 0.5)$  priors for all three parameters, the posterior mean of  $\theta = p_3 - \max(p_1, p_2)$  is 0.263 with a standard deviation of 0.106 and a 95% posterior interval of (0.048, 0.456). The posterior probability is  $\Pr(\theta > 0 | \mathbf{r}) = 0.992$ , which gives strong evidence to support superiority of the dose combination. Wang and Hung apply the Z-test twice, obtaining

$$Z_1 = \frac{0.68 - 0.38}{\sqrt{\frac{0.68 * 0.32}{34} + \frac{0.38 * 0.62}{34}}} = 2.54 \quad \text{and} \quad Z_2 = \frac{(0.68 - 0.32) - 0}{\sqrt{\frac{0.68 * 0.32}{34} + \frac{0.32 * 0.68}{34}}} = 3.11.$$

They report a two-sided  $p$ -value of 0.016 and conclude that the combination is superior to its components at the 5% significance level, but they do not offer an interval estimate.

Next, we consider an example provided by Laska, Meisner, and Siegel (1994). Suppose that prior experiments give estimates of the  $ED_{50}$ s for each of two drugs: 20 mg/kg for drug 1 and 10 mg/kg for drug 2. To test for synergy, a pharmacologist may

consider the drugs in combination at dose (10, 5) since the relative potency is estimated at 0.5. The associated hypotheses are

$$H_0: \theta \leq 0 \text{ vs. } H_1: \theta > 0,$$

where  $\theta = p_3(10, 5) - \max[p_1(20, 0), p_2(0, 10)]$ . Note that a researcher interested in testing for superiority in the combination of ED<sub>50</sub>s would test the hypotheses above by applying drug 1 at dose (10, 0) and drug 2 at dose (0, 5).

Now consider hypothetical results from this design. Suppose that  $n_1 = n_2 = n_3 = 30$  and 50% of patients respond to drugs 1 and 2 while 60% respond to the combination. Suppose we elicit mildly informative priors, letting  $p_1 \sim \text{Beta}(4, 4)$ ,  $p_2 \sim \text{Beta}(6, 6)$ , and  $p_3 \sim \text{Beta}(4, 2)$  as described above in Section 3.1. This prior structure implies that we know more *a priori* about drug 2 than drug 1 and that the combination is expected to produce a 67% response rate. This information yields the posterior

$$\pi(\mathbf{p} \mid \mathbf{r}) = b(p_1 \mid 19, 19) b(p_2 \mid 21, 21) b(p_3 \mid 22, 14).$$

Applying equation (3.3) yields a posterior probability of  $\Pr(\theta > 0 \mid \mathbf{r}) = 0.7435$ , which gives weak evidence to support synergy of the dose combination. Monte Carlo methods in *R* yield the posterior mean of 0.067 with a standard deviation of 0.102 and a 95% posterior interval of  $(-0.139, 0.262)$ . For this situation, the method of Laska, Meisner, and Siegel applies the Min test as described above, obtaining  $Z_1 = Z_2 = 0.7825$ , and thus does not reject at  $\alpha = 0.10$ .

Suppose instead that prior information for  $p_1$  and  $p_2$  comes from previous studies that produced dose response curves, such as the model in (3.2), so that  $\hat{\gamma} = \{-7.49, -3.70, 2.50, 1.60\}$ , where  $d_1 = \ln(20)$ ,  $d_2 = \ln(10)$ . Again consider the results  $\hat{p}_1 = 15/30$ ,  $\hat{p}_2 =$

15/30, and  $\hat{p}_3 = 18/30$ . Placing normal priors on  $\gamma$  with mean  $\mu$  and precision  $\tau$ , we consider mildly informative distributions and take  $\gamma_{01} \sim N(-7.49, 0.01)$ ,  $\gamma_{02} \sim N(-3.70, 0.01)$ ,  $\gamma_1 \sim N(2.50, 0.01)$ , and  $\gamma_2 \sim N(1.60, 0.01)$ . We also place a diffuse  $Beta(0.5, 0.5)$  prior on  $p_3$ . MCMC methods in Win-Bugs yield a posterior probability of  $\Pr(\theta > 0 | \mathbf{r}) = 0.6559$ , which gives weak evidence to support synergy in the dose combination. The posterior mean of  $\theta$  in this case is 0.045 with a standard deviation of 0.11 and a 95% posterior interval of  $(-0.181, 0.263)$ . The data does not change from the previous example, so the Min test would again obtain  $Z_1 = Z_2 = 0.7825$ , and thus not reject at  $\alpha = 0.10$ .

#### 4. SAMPLE-SIZE DETERMINATION

A vast literature exists for sample-size determination for comparing binomial proportions; see, for example, Joseph, du Berger, and Belisle (1997). Zou and Normand (2001) consider Bayesian sample-size determination for a hierarchical binomial model using MCMC and curve fitting. O'Malley, Normand, and Kuntz (2002) discuss a Bayesian approach to sample size determination for hypothesis testing in a logistic regression framework.

Interest lies in testing the hypothesis  $H_0: \theta \leq 0$  vs.  $H_1: \theta > 0$ , where, again,  $\theta = p_3 - \max(p_1, p_2)$ . Suppose one desires to detect a specified threshold value of  $\theta$ . That is, suppose we want to detect whether  $\theta > \delta$ , for some  $\delta > 0$ , with at least some specified probability, often 0.8 or 0.9. We follow Wang and Gelfand (2002) for this simulation-based sample-size determination procedure. Bayesian simulation-based sample-size

determination procedures begin when we specify a lower bound,  $b_1$ , on the probability that  $H_1$  is true. Next, for a particular sample size, a large number of sets of parameter values are randomly generated. We refer to the number of parameter sets as  $B$ . For each one of these sets of parameter values, data is generated, and posterior quantities of interest are calculated. These quantities are often credible sets for interval-based inference and posterior probabilities for hypothesis testing. For instance, if interest lies in hypothesis testing, the number of times the alternative hypothesis is concluded will be averaged over the  $B$  data sets, which is a type of *Bayesian power*. This procedure is repeated for several sample-sizes and the sample size that satisfies the criteria is found via either curve fitting or other search methods. Wang and Gelfand (2002) suggest eliciting two sets of prior distributions. One set is the sampling prior, and the other is the fitting prior. The sampling prior is analogous to the fixed values of parameters in frequentist sample-size determination and are used to generate the parameters for the simulation study. The fitting prior is the prior used in analyzing each generated data set. The two sets of priors can be the same, but often times due to regulatory agencies and other restrictions, the fitting priors are non-informative. We now describe an algorithm for the superiority model of Section 3.1. In this paper, we assume the sample sizes are all equal. A similar approach can be used for any of the models discussed in the paper. Here we seek to determine the required sample size for  $E_{\mathbf{r}}[P(\theta > 0 | \mathbf{r}) > 1 - \alpha] = 1 - \beta$ . A common choice for  $\alpha$  is 0.05, and a common choice for  $\beta$  is .2.

The estimation proceeds as follows:

1. Specify sampling priors for  $p_1$  and  $p_2$  and fit priors for  $p_1, p_2, p_3, \delta, \alpha$ , and  $\beta$ .

2. Create  $G$  grid values for sample sizes. That is, for  $g = 1, \dots, G$ , define sample sizes  $n_1, \dots, n_G$ .
3. For  $l = 1, \dots, B$  Monte Carlo iterations,
  - a) Generate  $p_1$  and  $p_2$  from their fitting priors and calculate  $p_3 = \delta + \max(p_1, p_2)$
  - b) For all the grid values,
    - (i) Generate  $r_{g,l,i} \sim \text{Bin}(n_{g,l}, p_i)$ ,  $i = 1, 2, 3$ ,  $g = 1, \dots, G$ ,  $l = 1, \dots, m$ .
    - (ii) Generate  $J$  values from the independent beta posteriors  $p_i \sim \text{Beta}(r_i + \alpha_i, n - r_i + \beta_i)$ .
    - (iii) For each of the  $j = 1, \dots, J$  generated values, approximate the posterior of  $\theta = p_3 - \max(p_1, p_2)$ .
    - (iv) Compute  $P(\theta > 0 | \mathbf{r}_j)$ .
    - (v) If the probability in (iv) is greater than  $1 - \alpha$ , then a 1 is recorded, otherwise a 0.
  - c) Calculate the Monte Carlo average of the 1's and 0's computed in 3. b. (v).
4. Repeating for all  $G$  grid values, find the minimum value of  $n$  such that the Monte Carlo probability in 3. c) is at least  $1 - \beta$ .

We now consider a specific example to illustrate the procedure. Suppose that sampling priors for  $p_1$  and  $p_2$  are elicited and found to be  $p_1 \sim \text{beta}(2, 8)$  and  $p_2 \sim \text{beta}(2.5, 7.5)$ . Using the procedure above, we determine the required sample size to detect effects of  $\delta = 0.1, 0.15, 0.2, 0.25$ , and  $0.3$  for  $\alpha = 0.05$  and  $\beta = 0.2$ . The sample sizes required for non-informative fitting priors are given in Table 1.

$\delta$	$n$
0.1	259
0.15	123
0.2	74
0.25	47
0.3	32

**Table 1** Required sample sizes for superiority example, where  $p_1 \sim \text{beta}(2, 8)$  and  $p_2 \sim \text{beta}(2.5, 7.5)$ ,  $\alpha = 0.05$ , and  $\beta = 0.2$ .

The program presented here for hypothesis testing can easily be altered for interval-based sample-size determination, such as the average coverage criterion (ACC) and average length criterion (ALC) discussed in Joseph, du Berger, and Bélisle (1997).

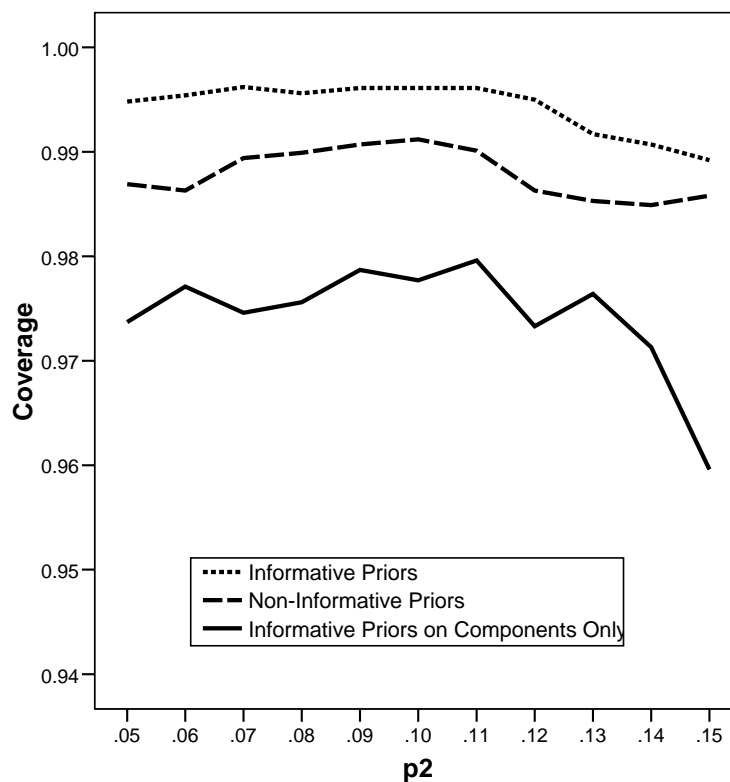
## 5. SIMULATION STUDY

### 5.1 Posterior Interval Coverage

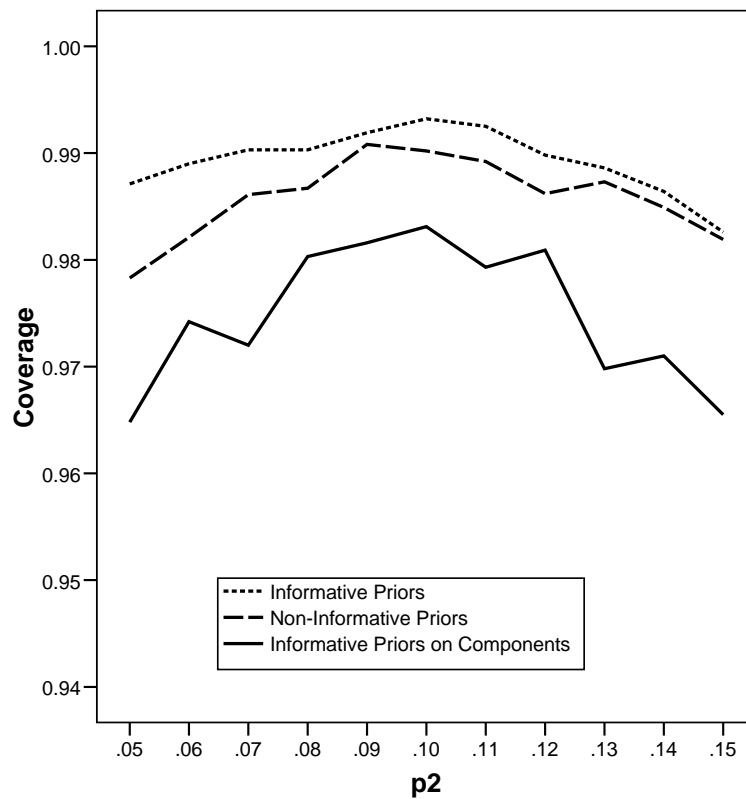
We now present a simulation study to illustrate the performance of the Bayesian procedure under various scenarios. The results represent the average for 5,000 replications. Note that the tests for both superiority and synergy consider the one-sided hypotheses in (2.1), with the difference between the two procedures existing at the dose application level. Once the doses have been selected, the hypothesis testing procedures are the same. Thus, we suppress dose notation and consider performance at various values of  $p_1$ ,  $p_2$ ,  $p_3$ , and hence  $\theta$ . For each set of parameters, we consider the effect of prior selection by first placing non-informative priors on all parameters, then placing mildly informative ones on all parameters, and finally placing informative priors on the component drugs with a non-informative prior on the combination. Likewise, we examine small, moderate, and large sample sizes.

In the first case, we set  $p_3 = 0.2$  and  $p_1 = 0.1$  and let  $p_2$  vary from 0.05 to 0.15 for samples of  $n_1 = n_2 = n_3 = 15$ ,  $n_1 = n_2 = n_3 = 30$ , and  $n_1 = n_2 = n_3 = 100$ . We also consider three sets of priors. First we use the non-informative Jeffreys prior,  $Beta(0.5, 0.5)$ , for all three parameters and then consider mildly informative priors taking  $p_1 \sim Beta(1, 9)$ ,  $p_2 \sim Beta(1, 9)$ , and  $p_3 \sim Beta(2, 8)$ . Note that in this situation the prior means for  $p_1$  and  $p_3$  are the true values while the prior mean for  $p_2$  is 0.1. Because  $p_2$  goes from 0.05 to 0.20 in the simulation, its prior distribution is only “correct” for  $p_2 = 0.1$ . Lastly, we let  $p_1 \sim Beta(1, 9)$ ,  $p_2 \sim Beta(1, 9)$ , and  $p_3 \sim Beta(0.5, 0.5)$ , modeling existing knowledge relative to the components but not the combination.

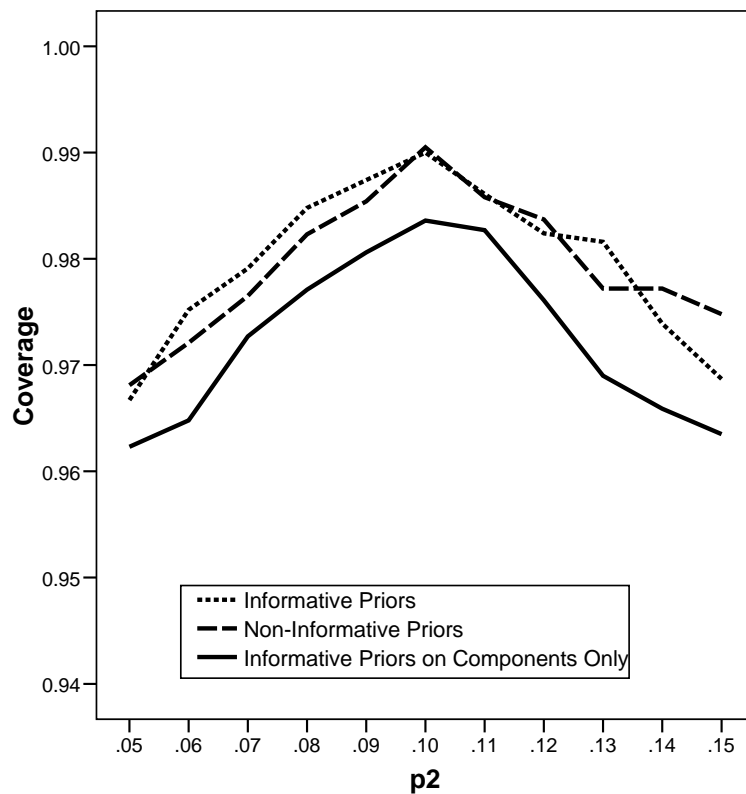
The one-sided interval coverages are provided in Figures 1-3. Using mildly informative priors for all three drugs produces the most conservative results. For the non-informative priors, as one would expect, results are also conservative. Most interesting is the performance in the case where information is available for the component drugs but not the combination. For this most likely situation, coverage is closest to nominal. In each case, coverage approaches nominal as the sample size increases. For moderate to large sample sizes, coverage is best for  $|p_1 - p_2| \geq 0.02$ .



**Figure 1** One-sided interval coverage as a function of  $p_2$  for  $n = 15$ .



**Figure 2** One-sided interval coverage as a function of  $p_2$ , for  $n = 30$ .

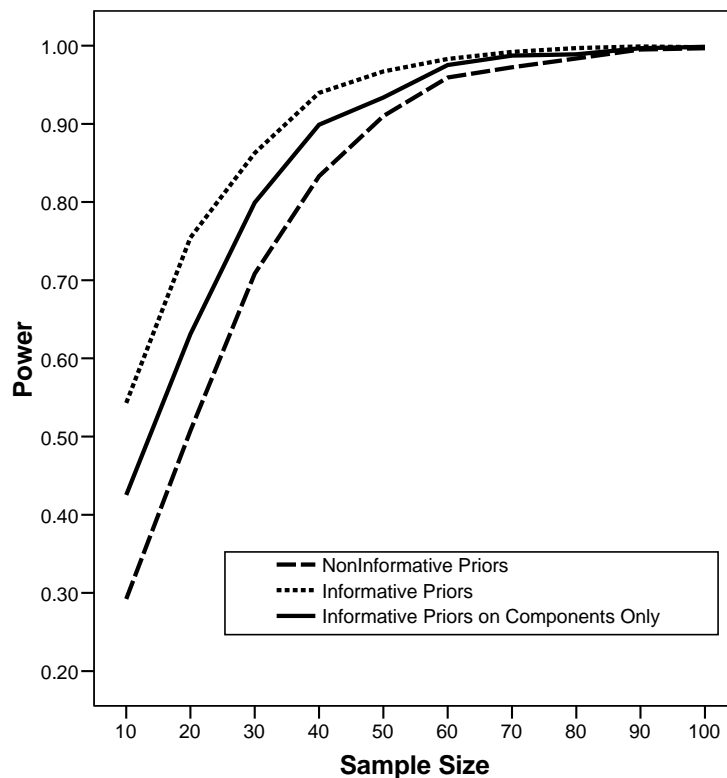


**Figure 3** One-sided interval coverage as a function of  $p_2$ , for  $n = 100$ .

For the second scenario, we let  $p_3 = 0.45$  and  $p_1 = 0.3$  and allow  $p_2$  to vary from 0.25 to 0.35 for the same three sample sizes. We first apply non-informative, then mildly informative priors, and also even more informative priors on the components with a non-informative one on the combination. For informative priors, we assume  $p_1 \sim \text{Beta}(2.5, 7.5)$  and  $p_2 \sim \text{Beta}(3, 7)$  and, for the combination,  $p_3 \sim \text{Beta}(5, 5)$ . The results are very similar to those in Figures 1 to 3. As expected, the widths for informative priors on all parameters are smaller than when informative priors are placed only on the components, and the widths are widest when non-informative priors are placed on all parameters. This difference does, of course, disappear with increased sample size. Likewise, exploring unequal sample sizes reveals only minimal changes in coverage.

## 5.2 Power

Next we turn our attention to hypothesis testing and power. We take  $p_1 = 0.2$ ,  $p_2 = 0.25$ , and  $p_3 = 0.55$ . As before, we consider three model structures: non-informative Jeffreys' priors for all three proportions, mildly informative priors for all three, and then informative priors for the components with a non-informative prior on the combination. As informative priors in this analysis, we model both  $p_1$  and  $p_2$  with a  $\text{Beta}(2.5, 7.5)$  distribution and  $p_3$  with a  $\text{Beta}(5.5, 4.5)$ . This prior structure implies a belief that  $p_1 \approx p_2$  when, in fact,  $p_1 < p_2$ , albeit only slightly. As Figure 4 illustrates, a power of 0.80 is achieved at a sample size of roughly 35 for the non-informative structure. For the structure that places informative priors on the components and a non-informative prior on the combination, the sample size reduces to approximately 30. For the informative structure, it is reduced to between 20 and 25.



**Figure 4** Power as a function of sample size for three prior structures.

### 5.3 Type I Error

To explore the Type I error of the proposed procedure, we take  $p_1 = 0.2$ ,  $p_2 = 0.25$ , and  $p_3 = 0.25$ . Here consider four model structures: non-informative Jeffreys priors for all three proportions, mildly informative pessimistic priors, mildly informative optimistic priors, and then informative priors for the components with Jeffreys prior on the combination. As in Section 5.2, we model both  $p_1$  and  $p_2$  with a  $Beta(2.5, 7.5)$  distribution. As a pessimistic prior structure, we model  $p_3$  with a  $Beta(2.5, 7.5)$  and then as an optimistic structure we model  $p_3$  with a  $Beta(3, 7)$ .

Figure 5 demonstrates that the procedure performs best when informative priors are placed on  $p_1$  and  $p_2$ , but not  $p_3$ . Both pessimistic and optimistic informative priors for  $p_1$

produce highly conservative results. In this case, performance becomes quite similar under all prior structures for sample sizes above 75.

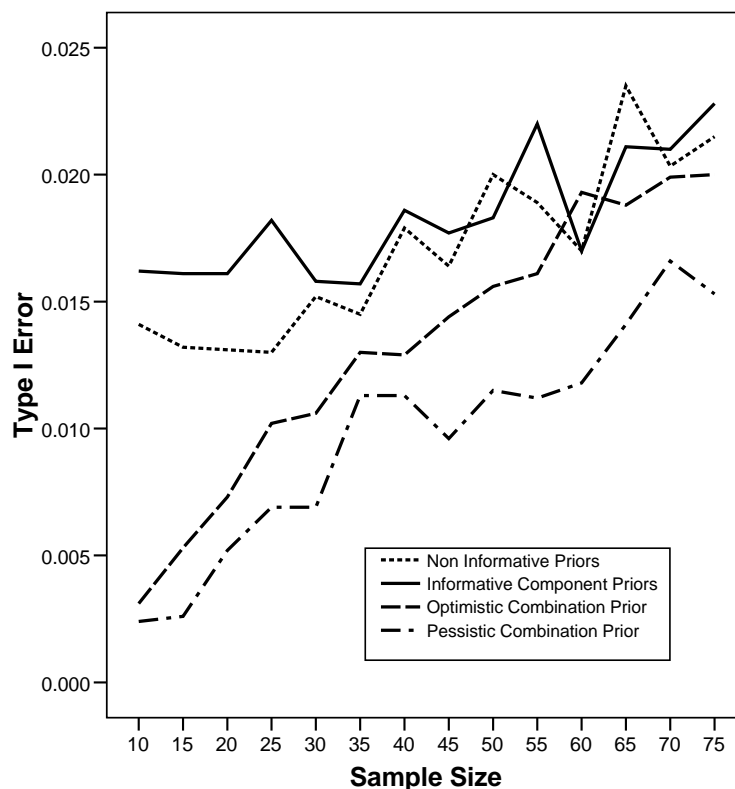


Figure 5 Type I error as a function of sample size for four prior structures.

## 7. DISCUSSION

The analyses presented herein provide a Bayesian alternative to existing procedures to test for either efficacy or synergy in a fixed-dose drug combination. These methods utilize the beta family of densities, affording flexible representation of prior information while permitting straight-forward calculation of posterior probabilities yielding direct assessment of drug combination efficacy or synergy. Just like existing procedures, the Bayesian method can be extended to combinations of three or more drugs by changing the hypotheses in (2.1) so that  $\theta = p_k - \max[p_1, p_2, \dots, p_{k-1}]$ . The associated prior

structure then becomes  $\pi(\mathbf{p}) \equiv \pi(p_1, p_2, p_3, \dots, p_k) = \pi(p_1) \pi(p_2 | p_1) \pi(p_3 | p_1, p_2) \dots \pi(p_k | p_1, p_2, \dots, p_{k-1})$ . The sample-size estimation procedure developed in Section 4 accommodates at any desired combination of prior structure, effect size, Type I error, and power.

Because the Min test for testing efficacy, as proposed by Wang and Hung (1997), is recommended at sample sizes of at least 20 and for response probabilities between 0.20 and 0.80, we consider performance of the proposed Bayesian analysis for both small and large sample sizes and for response probabilities both below and above 0.20. Likewise, although the Min test has been shown to be uniformly most powerful among monotone tests, Gibson and Overall (1989), Laska and Meisner (1989), and Hung, Chi, and Lipicky (1993) have remarked that it can be excessively conservative. This phenomenon is particularly true when the expected responses to drugs 1 and 2 are equal. Thus, we consider both Type I error and power of the new procedure when  $p_1 \approx p_2$ . In addition, Laska, Meisner, and Seigel (1994) point out that the Min test for synergy relies heavily on the *a priori* estimate of relative potency. Thus, we consider prior models with various levels of precision.

The examination of performance reveals that this procedure performs best under the most likely scenario. That is, interval coverage remains nearest the nominal level for both small and large sample sizes when mildly informative priors are placed on the component drugs and a non-informative prior is placed on the combination. Performance is similar for both small and moderate values of  $p$ . For samples of size 30 and above, the Bayesian procedure demonstrates the same conservative nature as the Min test at  $p_1 \approx p_2$ . This phenomenon is demonstrated by the higher coverage probabilities for  $|p_1 - p_2| \leq 0.02$  and

the low Type I error rate. The sample size determination procedure also reveals that the sample sizes necessary to apply the Bayesian analysis in fixed-dose studies are reasonable. The power study again shows that, although reasonable power can be achieved under each prior structure, taking advantage of existing information can reduce sample size considerably. Caution should be used, however, when one considers informative priors because both pessimistic and optimistic structures can significantly reduce the achieved Type I error rate. Again, optimal performance is achieved when one places informative priors on the component drugs and a non-informative prior on the combination.

## ACKNOWLEDGMENT

We wish to thank the anonymous referees for their insightful comments that significantly improved this paper.

## REFERENCES

- Abramowitz, M. and Stegun, I.A. (1965). *Handbook of Mathematical Functions*. New York: Dover.
- Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295-300.
- Berry, D.A. and Stangl, D.K., ed. (1996). *Bayesian Biostatistics*. New York: Marcel Dekker.
- De Santis, F., Pacifico, M.P., and Sambucini, V. (2004). Optimal predictive sample size for case-control studies. *J. R. Stat. Soc. C* **53**: 427-441.
- Gibson, J.M. and Overall, J.E. (1989). The superiority of a drug combination over each of its components. *Stat. Med.* **8**: 1479-1484.
- Greco, W. R., Bravo, G., and Parsons, J. C. (1995). The search for synergy: A critical review from a response surface perspective. *Pharmacological Reviews*, **47**: 331-385.
- Hung, H.M.J. (1996). Global tests for combination drug studies in factorial trials. *Stat. Med.* **15**: 233-247.
- Hung, H.M.J., Chi, G.Y.H., and Lipicky, R.J. (1993). Testing for the existence of a desirable dose combination. *Biometrics* **49**: 85-94.

- Hung, H.M.J., Chi, G.Y.H., and Lipicky, R.J. (1994). On some statistical methods for analysis of combination drug studies. *Communications in Statistics: Theory and Methods* 23: 361-376.
- Joseph L., du Berger R., Bélisle P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Stat. Med.* 16: 769-781.
- Laska, E.M. and Meisner, M.J. (1989). Testing whether an identified treatment is best. *Biometrics* 45: 1139-1151.
- Laska, E.M., Meisner, M., and Seigel, C. (1994). Simple designs and model-free tests for synergy. *Biometrics* 50: 834-841.
- Laska, E.M., Meisner, M., and Tang, D.I. (1997). Classification of the effectiveness of combination treatments. *Stat. Med.* 16: 2211-2228.
- Lehmann, E. L. (1952). Testing multiparameter hypotheses. *Annals of Mathematical Statistics* 23, 541-552.
- Miller, M.A. and Seaman, J.W., Jr. (1998). A Bayesian approach to assessing the superiority of a dose combination. *Biometrical Journal* 40: 43-55.
- Muller, P. and Parmigiani, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *J. Am. Stat. Assoc.* 90: 1322-1330.
- O'Malley, A.J., Normand, S.-L.T. and Kuntz, R.E. (2002). Sample size calculations for a historically controlled clinical trial with adjustments for covariates. *J. Biopharm. Stat.* 12: 227-247.
- Simon, R. and Freedman, L.S. (1997). Bayesian design and analysis of two  $\times$  two factorial clinical trials. *Biometrics* 53: 456-464.
- Su, C. (2006). BetaBuster, a software module made available by W. Johnson and I. Gardner (U. California, Davis) on the website Bayesian Epidemiologic Screening Techniques: <http://www.epi.ucdavis.edu/diagnostictests/betabuster.html#SumDesc>.
- Thall, P.F., Millikan, R.E., Mueller, P., and Lee, S.-J. (2003). Dose-finding with two agents in phase I oncology trials. *Biometrics* 59: 487-496.
- U.S. Food and Drug Administration (2004). <http://www.cfsan.fda.gov/~frf/bayesdl.html>.
- U.S. Food and Drug Administration (2006). <http://www.fda.gov/cdrh/osb/guidance/1601.html>.
- U.S. Food and Drug Administration (2006). <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=300.50>.
- Wang, F. and Gelfand, A.E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science* 17: 93-208.
- Wang, S.J. and Hung, H.M.J. (1997). Large sample tests for binary outcomes in fixed-dose combination drug studies. *Biometrics* 53: 498-503.
- Zou K.H., Normand S.-L.T. (2001). On determination of sample size in hierarchical binomial models. *Stat. Med.* 20:2163-2182.