On sample size determination based on a Bayesian index of superiority for two independent binomial proportions

Yohei Kawasaki1*

Abstract: In clinical study, calculation for sample size is often done using the frequentist framework. However, several methods of calculation using the Bayesian framework have been proposed. In this paper, we propose a method for the determination of sample sizes using the index $\theta = P(\pi_1 > \pi_2 | X_1, X_2)$, where $X_1$ and $X_2$ denote the binomial random variables for trials $n_1$ and $n_2$ and parameters $\pi_1$ and $\pi_2$, respectively. In this paper, we propose a new method of calculation, by which we determine in advance the estimated difference between the posterior proportions and the value to be derived at the end of the test to obtain the necessary sample size. We provide the lists of necessary sample sizes with various assumptions and the calculation program using Mathematica.

Keywords: binominal proportions; Bayesian inference; sample size

1. Introduction
In clinical trial tests, the difference between two independent binomial proportions is frequently examined. The comparative result is used to determine the propriety of a new treatment. Biostatisticians determine the main analysis method, estimated difference between binomial

ABOUT THE AUTHOR
Yohei Kawasaki received the BS degree in mathematics in 2005 and the MS and PhD. Degrees in mathematical science from Tokyo University of Science, Japan in 2007 and 2012, respectively. He joined the Mitsubishi Tanabe Pharma Corporation, the National Center for Global Health and Medicine, Tokyo University of Science, University of Shizuoka, in 2005, 2013, 2014, and 2016, respectively.

He has published over 80 research articles in reputed international journals. His research interests include the classical analysis, Bayesian analysis, Biostatistics, and mathematical education.

He is currently a senior lecturer in the Department of Pharmacoeconomics, Graduate School of Medicine and Public Health, Kyoto University, Japan. He also is a guest researcher of University of Shizuoka and Shizuoka General Hospital as Biostatistician.

PUBLIC INTEREST STATEMENT
In clinical study, calculation for sample size is often done using the frequentist framework. In this paper, we propose a new method of calculation, by which we determine in advance the estimated difference between the posterior proportions and the value to be derived at the end of the test to obtain the necessary sample size. We provide the lists of necessary sample sizes with various assumptions.
proportions, significance level, and power before a clinical study is conducted. This method is used to calculate the sample size within the frequentist framework.

A large body of literature describes methods for calculating sample size, including both frequentist (Lachin, 1981; Lemeshow, Hosmer, Klar, & Lwango, 1990) and Bayesian approaches. Several criteria have been proposed for Bayesian sample size determination. The specific application to the binomial parameter was examined in detail by Pham-Gia and Turkkan (1992), while Adcock (1987) considered multinomial experiments, which include the binomial as a special case. Adcock (1992) compared the various approaches presented in the above two studies. Joseph, Wolfson, and Berger (1995) proposed three different Bayesian approaches to calculate sample size. These proposed methods are based on highest posterior density-credible intervals and are discussed and illustrated in the context of a binomial experiment. Decision-theoretic criteria (Lindley, 1997) and sample sizes based on average power of hypothesis tests (Spiegelhalter & Freedman, 1986) have also been considered. Chaloner and Verdinelli (1995) recently published a review of Bayesian optimal design, while Adcock (1997) reviewed both frequentist and Bayesian sample size criteria. More recently, M’Lan, Joseph, and Wolfson (2008) investigated the binomial sample size problem using generalized versions of the average length and average coverage criteria, median length and median coverage criteria, and worst outcome criterion and its modified version. These methods of calculating sample size are based on a highest posterior density-credible interval.


\[
\theta = P(\pi_1 > \pi_2 | X_1, X_2)
\]  

(1.1)

where \(X_1\) and \(X_2\) denote the binomial random variables for trials \(n_1\) and \(n_2\) and parameters \(\pi_1\) and \(\pi_2\), respectively. They provided approximate and exact expressions for index \(\theta\) using a beta prior and presented the results of actual clinical trials to demonstrate the utility of this index. This index can be used to determine the probability of the binomial proportion of a treatment group to be superior to that of a control group. Thus, this index determines the inferiority between two independent binomial proportions. The proposed index \(\theta\) is advantageous in that it can be easily and intuitively understood and applied.

In this paper, we propose a new method for calculating sample size using index \(\theta\) in (1.1). For this method, the difference between two independent posterior binomial proportions and index \(\theta\) are decided before clinical research, and the calculation result of the sample size required in the case is shown. Moreover, we compare sample sizes using approximate and exact methods for index \(\theta\).

The remainder of this paper is organized as follows. We obtain approximate and exact expressions for index \(\theta\) in Section 2. In Section 3, we propose a method for calculating sample size using \(\theta\) in (1.1), and we provide lists of necessary sample sizes from simulation results with various assumptions in Section 4. In Section 5, we consider an imaginary clinical trial of a new treatment and calculate the value of sample sizes using the lists of necessary sample size provided in Section 4. Finally, we conclude the paper with a brief summary in Section 6.

2. Methodology

Let \(X_i\) denote binomial random variables for \(n_i\) trials with parameters \(\pi_i\), respectively. The conjugate prior density for \(\pi_i\) is a beta distribution with parameters \(\alpha_i\) and \(\beta_i\), where \(\alpha_i > 0, \beta_i > 0,\) and \(i = 1, 2\). The proposed posterior density for \(\pi_i\) is
g_i(\pi_i|x_i) = \frac{1}{B(a_i, b_i)} \pi_i^{a_i-1}(1-\pi_i)^{b_i-1},

where a_i = n_i + x_i, b_i = n_i - x_i + \beta_i, and B(a, b) is the proposed beta function.

Approximate expression for \( \theta \)

One method of calculation for the index \( \theta \) in (1.1) is an approximation using the standard normal table. Let \( \pi_{1, \text{post}} \) denote the binominal proportion in the posterior density. We assume that the \( a_i \) and \( b_i \) of the posterior density are large. We need to find a \( Z \)-test statistic. We can show the expectation of difference for the posterior density and the variance of difference for the posterior density as follows:

\[
E(\pi_{1, \text{post}} - \pi_{2, \text{post}}) = \mu_{1, \text{post}} - \mu_{2, \text{post}},
\]

\[
V(\pi_{1, \text{post}} - \pi_{2, \text{post}}) = \frac{\mu_{1, \text{post}}(1-\mu_{1, \text{post}})}{a_1 + b_1 + 1} + \frac{\mu_{2, \text{post}}(1-\mu_{2, \text{post}})}{a_2 + b_2 + 1},
\]

where \( \mu_{i, \text{post}} = a_i/(a_i + b_i) \) denotes the posterior mean of \( \pi_i \). The \( Z_g \)-test statistic is

\[
Z_g = \frac{(\pi_{1, \text{post}} - \pi_{2, \text{post}}) - E(\pi_{1, \text{post}} - \pi_{2, \text{post}})}{\sqrt{V(\pi_{1, \text{post}} - \pi_{2, \text{post}})}}.
\]

The \( Z_g \)-test statistic is approximately distributed as the standard normal distribution. Therefore, the approximate probability of index \( \theta \) is

\[
\theta = P(\pi_1 > \pi_2 | X_1, X_2)
\]

\[
\approx 1 - \Phi\left( \frac{-(\mu_{1, \text{post}} - \mu_{2, \text{post}})}{\sqrt{\frac{\mu_{1, \text{post}}(1-\mu_{1, \text{post}})}{a_1 + b_1 + 1} + \frac{\mu_{2, \text{post}}(1-\mu_{2, \text{post}})}{a_2 + b_2 + 1}} \right)
\]

where \( \Phi(\cdot) \) is the cumulative distribution function (CDF) of the standard normal distribution. We can easily calculate the approximate probability. In this paper, we show the difference between an exact probability and the approximate probability in Section 4.

Exact expression for \( \theta \).

In contrast, Kawasaki and Miyakoa (2012) derived an exact expression for index \( \theta \) in (1.1) using the posterior density. The exact expression for index \( \theta \) is

\[
\theta = P(\pi_1 > \pi_2 | X_1, X_2)
\]

\[
= \frac{B(a_1 + a_2, b_1)}{B(a_1, b_1)B(a_2, b_2)} \cdot \text{F}_2(a_2, 1 - b_2, a_1 + a_2; 1 + a_2, a_1 + a_2 + b_1; 1),
\]

where

\[
\text{F}_2(k_1, k_2; l_1, l_2; 1) = \sum_{t=0}^{\infty} \frac{(k_1)_t(k_2)_t(k_3)_t}{(l_1)_t(l_2)_t} \frac{1}{t!}, \quad k_1 + k_2 + k_3 < l_1 + l_2
\]
is the hypergeometric series, and \((k)_t\) is the Pochhammer symbol.

**Method of Calculation for Sample Size**

We propose a method of calculation for sample size using \(\theta\) in (1.1) in this section. The detailed flow of the method is as follows:

1. Determine the value of hyperparameters \(\alpha_1, \beta_1, \alpha_2, \text{ and } \beta_2\) in beta distributions that are prior distributions.
2. Determine the value of \(\delta\), which is the difference in the posterior densities of the two groups:

\[
\delta = E(\pi_{1,\text{post}} - \pi_{2,\text{post}}) = \frac{a_1}{a_1 + b_1} - \frac{a_2}{a_2 + b_2}
\]

3. Determine the value of \(\theta_{\text{min}}\), which is the lower limit of \(\theta\).
4. Determine the ratio \(r\) between the sample size of the two groups (where \(n_1 = r \times n_2\)).
5. Based on the above assumptions (1 to 4), set to \(n_2 = 1\) and calculate the following.
6. Find the combinations of the realized values \(X_1\) and \(X_2\) of the random variables \(X_1\) and \(X_2\), which are satisfied by the condition of \(\delta\) in step 2.
7. Calculate the value of \(\theta\) for all combinations found in step 6.
8. Update the value of \(n_2\) by \(n_2 + 1\) and go back to step 6 (iterate this calculation an appropriate number of times).
9. Find the value of the minimum \(n_1\) for which all obtained values of \(\theta\) in step 7 exceed \(\theta_{\text{post}}\), and determine that the necessary sample size for the treatment group is this value. The necessary sample size for the control group can be defined by the relative expression \(n_1 = r \times n_2\).

**3. Results**

**3.1. Simulation**

In this section, we simulate the necessary sample size using \(\theta\) in (1.1), which was defined in Section 2. Each assumption and setting is as follows: the noninformative prior is \((\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1)\), the expected difference between the posterior densities of the two groups is \(\delta = 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15\), the lower limit of \(\theta\) is \(\theta_{\text{min}} = 0.7, 0.75, 0.8, 0.85, 0.9, 0.95\), and the ratio of sample sizes between the treatment and control groups is \(r = 1, 2, 3\). In addition, the necessary sample size under other conditions is easily calculated by simulation, although we do not include this in the present paper.

Simulation Results when \(n_1 = n_2\).

In this section, we display the result when the sample sizes of groups 1 and 2 are equal. Figure 1(a)–(d) feature \(\theta\) on the vertical axis and sample size on the horizontal axis. In Figure 1(a)–(d), we set the estimated difference between the posterior binomial proportions as 0.05, 0.10, 0.15, and 0.2, respectively. Each horizontal line indicates the value when \(\theta = 0.7, 0.75, 0.8, 0.85, 0.9, \text{ or } 0.95\). Moreover, we displayed two sections in Table 1; one is calculated based on the exact probability of \(\theta\) and the other, on the approximate probability.

Figure 1(a)–(d) shows that when sample size becomes bigger, \(\theta\) increases, but not monotonously. This is a trait of discrete distribution in the frequentist framework. We also find that the difference between the posterior proportions becomes larger as the \(\theta\) value become higher. In accordance with this, the necessary sample size decreases.
We display the detailed results for sample size in Table 1. The vertical column shows the value of $\theta$, and the horizontal row shows the difference between posterior proportions. We can see at several points that the necessary sample size for exact calculation is slightly higher than for approximation.

**Simulation Results when $n_1 \neq n_2$**

Figure 2(a)–(d) shows the results when sample sizes are $n_1 = 2 \times n_2$. Table 2 shows the detailed results as above. These figures and tables show that the necessary sample size $n_1 + n_2$ is larger in this
case than when sample sizes are equal among groups. We find from the following result that this trend increases as the imbalance between \( n_1 \) and \( n_2 \) becomes larger. In addition, Figure 3(a)–(d) and Table 3 show the result when \( n_1 = 3 \times n_2 \). Those also show that when sample sizes among groups are imbalanced the necessary sample size becomes larger. We also find that as the imbalance between \( n_1 \) and \( n_2 \) becomes larger, the difference between the necessary sample sizes for exact and approximate calculation tends to become larger.

![Figure 2. The exact probability of \( \theta \) when \( n_1 = 2 \times n_2 \).](image)

![Table 2. The result of necessary sample size \( n_1 \) when \( n_1 = 2 \times n_2 \).](table)

<table>
<thead>
<tr>
<th>The poster binomial proportions</th>
<th>Calculation based on exact probability of ( \theta )</th>
<th>Calculation bases on approximate probability of ( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05 0.06 0.07 0.08 0.09 0.1 0.11 0.12 0.13 0.14 0.15</td>
<td>0.7</td>
<td>92 58 44 26 20 14 10 10 8 12 6</td>
</tr>
<tr>
<td>0.75</td>
<td>158 108 76 54 36 32 22 18 16 14 10</td>
<td>0.75</td>
</tr>
<tr>
<td>0.8</td>
<td>246 166 122 82 68 54 40 34 24 22 20</td>
<td>0.8</td>
</tr>
<tr>
<td>0.85</td>
<td>398 256 184 134 106 84 66 52 46 36 34</td>
<td>0.85</td>
</tr>
<tr>
<td>0.9</td>
<td>600 400 278 212 164 126 105 86 72 58 54</td>
<td>0.9</td>
</tr>
<tr>
<td>0.95</td>
<td>1,000 656 476 346 270 220 180 146 120 104 88</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Example In this section, we provide examples of the setting method for sample size for a case in which we consider an imaginary clinical trial of a new treatment based on the chart in section 3. Suppose the following situation. Assume a non-informative prior as the prior distribution ($\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 1$). Set the same sample size for the treatment and control groups ($n_1 = n_2$). Assume that the expected value of difference of the posterior density between the two groups is 0.08 ($\delta = 0.08$). Under the above condition, from Table 1, if the lower limit of $\theta$ is set to 0.9 ($\theta_{\text{min}} = 0.9$), the necessary sample size for the treatment and new groups is 132 by both exact and approximate calculation. If the lower limit is set to 0.95 ($\theta_{\text{min}} = 0.95$), the necessary sample size becomes 239 by exact calculation, but 225 by approximation.
4. Conclusion
In this paper, we suggested a new calculation method to obtain the sample size for the index of binomial proportion $\theta = P(\pi_1 > \pi_2 | X_1, X_2)$ suggested by Kawasaki and Miyaoka (2012) and simulated the actual value. We also defined a method of calculation for $\theta$ based on both approximation and exact methods and compared the results.

We used Mathematica, which is standard calculation software for simulations. The main reason we used Mathematica is because hypergeometric series calculation is necessary for exact calculation of $\theta$, and its formula is implemented in Mathematica. Detailed programming code is provided in Appendix A.

The method to define sample size suggested in this paper begins with setting the estimated difference between the posterior proportions and the lower limit of $\theta$. In addition, by setting the proportion of the sample sizes of the treatment and control groups and adding the information (if it is available), the sample size can be derived.

We found three important results regarding $\theta$ derived by simulation for both exact calculation and approximation. First, when the sample size becomes larger, $\theta$ increases, but not monotonously. Second, as the estimated difference between the posterior proportions becomes larger, the necessary sample size decreases. Third, when the difference between the proportions of the treatment and control groups becomes larger, the necessary sample size for both groups becomes larger.

In addition, comparison of the exact and approximate values in the simulation yields the following result. Exact calculation and approximation require almost the same sample size, and the trend for $\theta$ to increase when the sample size becomes larger is almost the same. The sample size for exact calculation is slightly larger than that for approximation. This difference tends to increase as the imbalance between the treatment group and the control group increases.

Finally, index $\theta$, which has been extensively investigated in this research, is derived using the Bayesian framework rather than the frequentist framework as before. One of the merits of using the Bayesian framework is that it is easy to understand. Therefore, the index is expected to deliver very important information on clinical trials, where setting the sample size is considered very important. The setting method suggested in this paper may serve as a useful guideline in this regard.

Acknowledgements
We are very grateful to the editor and two referees for their invaluable comments.

Funding
The author received no direct funding for this research.

Author details
Yohei Kawasaki
E-mail: ykawasaki@u-shizuoka-ken.ac.jp
1 Department of Drug Evaluation & Informatics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan.

Citation information
Cite this article as: On sample size determination based on a Bayesian index of superiority for two independent binomial proportions, Yohei Kawasaki, Cogent Mathematics (2017), 4: 1284972.

References
Appendix A

We provide the sample size calculation program using Mathematica. Each parameter is defined as follows: “\(a_1\)”, “\(b_1\)”, are hyperparameters of treatment group; “\(a_2\)”, “\(b_2\)”, are hyperparameters of control group; “\(\delta\)”, expected difference in the posterior densities of two groups; “\(r\)”, is \(n_1 = r \times n_2\), which is the ratio of sample size between treatment group and control group. Program is as follows:

\[
\text{(* setting of parameters*)}
\]
\[
a_1 = 1;
b_1 = 1;
a_2 = 1;
b_2 = 1;
delta = 0.05;
r = 1;
nmin = 1;
nmax = 250;
\text{(* Exact method *)}
\]
\[
\text{list } = \{0\};
\text{Do[Do[Do[}
\]
\[
\text{If[Round[(a1 + x1)/((a1 + x1) + (n1 - x1 + b1)) - (a2 + x2)/((a2 + x2) + (n1/r - x2 + b2)), 0.01] == delta,}
\]


list = Append[
    list, {n1, Beta[(x1 + a1) + (x2 + a2), (n1 - x1 + b1)]/((x2 + a2)*Beta[(x2 + a2), (n1/r - x2 + b2)]*Beta[(x1 + a1), (n1 - x1 + b1)])*HypergeometricPFQ[((x2 + a2), 1 - (n1/r - x2 + b2), (x1 + a1) + (x2 + a2)), (1 + (x2 + a2), (x1 + a1) + (x2 + a2) + (n1 - x1 + b1)), 1]], list = list], {x2, 0, n1/r, 1}], {x1, 0, n1, 1}], {n1, nmin, nmax, 2}]
list = Delete[list, 1];
ListPlot[{list, ConstantArray[0.7, nmax], ConstantArray[0.75, nmax],
    ConstantArray[0.8, nmax], ConstantArray[0.85, nmax],
    ConstantArray[0.9, nmax], ConstantArray[0.95, nmax]},
    Joined -> {False, True, True, True, True, True, True},
    AxesOrigin -> {1, 0.5},
    PlotRange -> {{nmin - 1, nmax + 1}, {0.5, 1}},
    AxesLabel -> {n1, Prob}, Ticks -> {5*Range[nmax]}]

(* Approximate method *)
list = {0};
Do[Do[Do[If[Round[(a1 + x1)/((a1 + x1) + (n1 - x1 + b1)) - (a2 + x2)/((a2 + x2) + (n1/r - x2 + b2)), 0.01] == delta, list = Append[list, {n1, 1 - N[CDF[NormalDistribution[0, 1], -((a1 + x1)/((a1 + x1) + (n1 - x1 + b1)) - (a2 + x2)/((a2 + x2) + (n1/r - x2 + b2)))/Sqrt[((a1 + x1)*(n1 - x1 + b1))/(((a1 + x1) + (n1 - x1 + b1)))^2 + ((a2 + x2)*(n1/r - x2 + b2))/(((a2 + x2) + (n1/r - x2 + b2)))^2]]}], list = list], {x2, 0, n1/r, 1}], {x1, 0, n1, 1}], {n1, nmin, nmax, 2}]
list = Delete[list, 1];
ListPlot[{list, ConstantArray[0.7, nmax], ConstantArray[0.75, nmax],
    ConstantArray[0.8, nmax], ConstantArray[0.85, nmax],
    ConstantArray[0.9, nmax], ConstantArray[0.95, nmax]},
    Joined -> {False, True, True, True, True, True, True},
    AxesOrigin -> {1, 0.5},
    PlotRange -> {{nmin - 1, nmax + 1}, {0.5, 1}},
    AxesLabel -> {n1, Prob}, Ticks -> {5*Range[nmax]}]