Open Access | CrossMark

Bayesian sample size determination in two-sample poisson models

Abstract

Sample size determination is a vital part of clinical studies where cost and safety concerns lead to greater importance of not using more subjects and resources than are required. The Bayesian approach to sample size determination has the advantages of being able to use prior data and expert opinion to possibly reduce the total sample size while also acknowledging all uncertainty at the design stage. We apply a Bayesian decision theoretic approach to the problem of comparing two Poisson rates and find the required sample size to obtain a desired power while controlling the Type I error rate.

Keywords: bayes factor, poisson rate, power, type i error

Ryan Sides, David Kahle, James Stamey
Statistical Science, Baylor University, USA

Correspondence: David Kahle, Department of Statistical Science, Baylor University, Waco, TX, USA, Tel 254-710-6102, E-mail david_kahle@baylor.edu

Introduction

Sample size determination continues to be an important research area of statistics. Perhaps nowhere is this truer than pharmaceutical statistics, where cost and time constraints have made finding the appropriate sample size before conducting a study of the utmost importance. The problem is quite simple: too small a sample can lead to under-powered studies, while too large a sample size wastes precious resources. In this article we consider the sample size determination problem as it pertains to the two-sample testing of Poisson rates from a Bayesian perspective subject to operating characteristics constraints.

There are several advantages to the Bayesian perspective when trying to determine a study's requisite sample size, a topic that is expounded in Adcock.[1] Their construction does not depend on asymptotic approximations or bounds. Classical solutions to the sample size determination problem typically hinge on asymptotic arguments that require the researcher to specify one parameter value (perhaps vector valued) as representative for the entire parameter space, a process that is typically done using bounding arguments. This, for example, is what is done when determining the requisite sample size for a confidence interval of a fixed level and given length. The resulting sample sizes are consequently conservative. On the other hand, the Bayesian approach provides the statistician with the ability to model his indecision about the parameter through expert knowledge or previous studies. As noted by Bayarri and Berger,[2] this can allow the Bayesian approach to have better operating characteristics, such as a smaller required sample sizes or better Type I and II error rates.

Various Bayesian sample size determination methods have been studied for binomial and Poisson data. Stamey et al.,[3] considered one and two sample Poisson rates from the perspective of interval based criteria such as coverage and width. Hand et al.,[4] extend those ideas by considering both interval-based and test-based criteria, albeit without considering power. Katsis and Toman[5] used more decision theoretic criteria for the two sample binomial case, but only to the extent of controlling the posterior risk with a pre specified bound. Zhao et al.,[6] extend those ideas by using computational methods to consider expected Bayesian power of the test. In this article, we extend these results to the Poisson data model. We also consider the problem the subject to Type I and Type II constraints. This is thus an important extension of,[6] because it

1) Extends the ideas to the Poisson case

2) Enables the incorporation of operating characteristics.

A subtle difference between the classical and Bayesian methods of sample size determination merits discussion before proceeding. One of the novel contributions of this article is an algorithmic solution to the sample size determination problem subject to operating characteristics constraints for Poisson data. However, since the entire problem is treated in a Bayesian context, the concept of Type I and Type II error rates is understood in an average, or expected, sense; see.[7] For example, the ``power of a test'' retains the interpretation of the probability the decision rule rejects when the null hypothesis when it is false; but rather than being a function defined over the alternative space, here it is averaged over that space and weighted by the prior distribution specified on the alternative hypothesis. To make the distinction more clear, we refer to this as the expected Bayesian power (EBP), as is done in,[8] alternatively, it may be referred to as the probability of a successful test. These ideas, though apparently not considered in the literature previously, can also apply to the concept of the significance level of a test. While frequentist methods typically report one value for significance level, what they are really doing (in non point null hypotheses) is taking the largest possible significance level; thus, taking an expectation of a significance level curve could be done as well. Consequently, in this article we also consider the expected Bayesian significance level (EBSL), defined as the expected value of the test under the prior distribution given on the null space. In both cases, any particular instance of the actual Type I and Type II error rates can be greater than or less than nominal.

This article proceeds as follows. In Section 2 we introduce the theoretical formulation of the sample size determination problem for two Poisson variates including consideration of operating characteristics. In Section 3, we present an algorithmic solution to the sample size determination problem posed in Section 4. Section 5 contains an application of the method in the area of pharmaceutical statistics. We then conclude with a discussion.

Problem specification and the bayes rule

We now follow the general framework of[6] in the development of this problem, adapting the binomial case to fit the Poisson data model. Suppose $Y_1 \sim Pois(t\lambda_1)$ and $Y_2 \sim Pois(t\lambda_2)$, independently, where λ_1 and λ_2 represent the rate parameters of interest and t

represents a common sample (or "opportunity") size. Together, we write these $Y = (Y_1, Y_2)'$ with observations $y = (y_1, y_2)'$. The sample size determination problem is to calculate the necessary sample size required to test the hypotheses $H_0 : \lambda_1 = \lambda_2$ vs $H_1 : \lambda_1 \neq \lambda_2$

using a given decision rule; here we use the Bayes rule. Denoting the parameter pair $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$, the associated null and alternative spaces are therefore $\Lambda_0 = \{\boldsymbol{\lambda} \in R_+^2 : \lambda_1 = \lambda_2\}$, which we identify with R_+ with elements generically denoted λ, and $\Lambda_1 = \{\boldsymbol{\lambda} \in R_+^2 : \lambda_1 \neq \lambda_2\}$.

As this problem is being considered from the Bayesian perspective, we place prior probabilities of π_0 and $\pi_1 = 1 - \pi_0$ on H_0 and H_1, respectively. Conditional on the null being true, we represent the expert opinion regarding λ as $p_0(\lambda)$, defined over the set Λ_0. Alternatively, conditionally on H_1 being true, we represent the belief concerning $\lambda_1 \neq \lambda_2$ with a joint prior $p(\lambda_1, \lambda_2)$ with support Λ_1. Marginalizing over the hypotheses, we have the unconditional prior

$$p(\lambda_1, \lambda_2) = p_0(\lambda)\pi_0 I_{H_0} + p_1(\lambda_1, \lambda_2)\pi_1 I_{H_1},$$

Where I_{H_0} and I_{H_1} are the indicator function of H_0 and H_1, respectively.

In practice, the prior distributions on λ, λ_1, and λ_2 summarize expert opinion concerning the parameters in each of the two scenarios H_0 and H_1. This information can be obtained in a number of ways including past data, prior elicitation of expert opinion (see especially[9]), or based on uninformative criteria. For simplicity, we consider conjugate priors for all three λ's, so that under H_0, $\lambda \sim \text{Gamma}(\alpha, \beta)$, and under H_1, $\lambda_i \sim \text{Gamma}(\alpha_i, \beta_i)$ independently. This assumption is not very restrictive and allows us to specify parameters of prior distribution sin stead of distributions themselves.

We now derive the optimal Bayes (decision) rule in deciding between the hypotheses presented in (1). In solving the related problem between two binomial proportions,[6] use the classical decision theoretic setup using the 0-1 loss function.[10] Here we use the more general unequal loss function

$$L(H, a) = \begin{cases} c_1 & \text{if } H_0 \text{ is true and } a = 1, \\ c_2 & \text{if } H_1 \text{ is true and } a = 0, \\ 0, & \text{otherwise} \end{cases}$$

Where $a = \delta(y)$ is the decision rule with $a = 0$ representing selection of H_0 and $a = 1$ selection of H_1. Thus, c_1 represents the loss associated with a Type I error, and c_2 that of a Type II error.

The Bayes action is simply the one that minimizes posterior expected loss.[11] Since the posterior expected loss associated with an action $a \in \{0,1\}$ is

$$\rho(a) = E_{\boldsymbol{\lambda}|y}[L(\varphi, a)]$$
$$= \begin{cases} c_1 P(H_0 | Y = y), & \text{if } a = 1, \\ c_2 P(H_1 | Y = y), & \text{if } a = 0, \end{cases}$$

Setting $c = c_1 / c_2$ we can express the optimal decision rule, a^*, as

$$a^*(y) = \begin{cases} 0, & \text{if } P(H_1 | Y = y) < cP(H_0 | Y = y), \\ 1, & \text{if } P(H_1 | Y = y) \geq cP(H_0 | Y = y). \end{cases}$$

The rejection region, W, is therefore

$$W = \{y : P(H_1 | Y = y) \geq cP(H_0 | Y = y)\}.$$

The optimal rule in (2) can be nicely represented in terms of the Bayes factor. The Bayes factor is defined as the ratio of the

posterior odds of H_1 to the prior odds of H_1, so that a large Bayes factor is evidence for rejecting the null hypothesis.[12] Specifically, the Bayes factor is defined

$$B = \frac{P(H_1 | Y = y) / P(H_0 | Y = y)}{\pi_1 / \pi_0} = \frac{P(H_1 | Y = y)\pi_0}{P(H_0 | Y = y)\pi_1}.$$

This ratio is useful in Bayesian inference because it is often interpreted as partially eliminating the influence of the prior on the posterior, instead emphasizing the role of the data. Moreover, the decision rule is a function of a Bayes factor:

$$W = \{y : P(H_1 | Y = y) \geq cP(H_0 | Y = y)\}$$
$$= \left\{y : P(H_1 | Y = y)\frac{\pi_0}{\pi_1} \geq cP(H_0 | Y = y)\frac{\pi_0}{\pi_1}\right\}$$
$$= \left\{y : B \geq c\frac{\pi_0}{\pi_1}\right\}.$$

This is particularly useful because it allows for the interpretation of the Bayes factor B as the test statistic for the decision rule in (2); this is the condition in (4).

We now derive closed-form expression for the posterior probabilities of the null and alternative hypotheses. Using Bayes' theorem, we have

$$P(H_0 | Y = y) = \frac{P(Y = y | H_0)\pi_0}{P(Y = y | H_0)\pi_0 + P(Y = y | H_1)\pi_1}$$

and

$$P(H_1 | Y = y) = \frac{P(Y = y | H_1)\pi_1}{P(Y = y | H_0)\pi_0 + P(Y = y | H_1)\pi_1}.$$

Consequently, the posterior probabilities have closed-form expressions if the likelihoods do. Computing these, we have

$$P(Y = y | H_0) = \int_0^\infty f(y | \lambda, H_0) p_0(\lambda | H_0) d\lambda$$
$$= \int_0^\infty \frac{(\lambda t)^{y_1} e^{-\lambda t}}{y_1!} \frac{(\lambda t)^{y_2} e^{-\lambda t}}{y_2!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda$$
$$= \frac{t^{y_1+y_2} \beta^\alpha \Gamma(y_1 + y_2 + \alpha)}{y_1! y_2! \Gamma(\alpha)(2t + \beta)^{y_1+y_2+\alpha}},$$

and

$$P(Y = y | H_1) = \int_0^\infty f(y | \boldsymbol{\lambda}, H_1) p_1(\boldsymbol{\lambda} | H_1) d\boldsymbol{\lambda}$$
$$= \prod_{i=1}^2 \int_0^\infty \frac{(\lambda_i t)^{y_i} e^{-\lambda_i t}}{y_i!} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i-1} e^{-\beta_i\lambda_i} d\lambda_i$$
$$= \prod_{i=1}^2 \frac{t^{y_i} \beta_i^{\alpha_i} \Gamma(y_i + \alpha_i)}{y_i! \Gamma(\alpha_i)(t + \beta_i)^{y_i+\alpha_i}}.$$

Note that the probability of the data under the null hypothesis is the product of two independent negative binomial likelihoods.

Combining (3) with (5) and (6) allows us to represent W in terms of the null and alternative likelihoods as follows:

$$W = \{y : P(H_1 | Y = y) \geq cP(H_0 | Y = y)\}$$
$$= \{y : P(Y = y | H_1)\pi_1 \geq cP(Y = y | H_0)\pi_0\}.$$

Consequently, (7) and (8) give explicit conditions for the rejection of the optimal decision rule:

$$W = \left\{y : B = \frac{\Gamma(\alpha)(2t + \beta)^{y_1+y_2+\alpha}}{\beta^\alpha \Gamma(y_1 + y_2 + \alpha)} \prod_{i=1}^2 \frac{\beta_i^{\alpha_i} \Gamma(y_i + \alpha_i)}{\Gamma(\alpha_i)(t + \beta_i)^{y_i+\alpha_i}} \geq c\frac{\pi_0}{\pi_1}\right\}$$

Note that the left side of (9) is our test statistic and Bayes factor, B, so that (9) is an explicit formulation of the condition presented in (4).

Sample size determination for the bayes Rule

The explicit description of the decision rule in (9) allows us to compute all sorts of quantities of interest. For given prior parameters π_0, π_1, α, β, α_1, β_1, α_2, and β_2 and loss penalties c_1 and c_2 (or simply c), the Expected Bayesian Power (EBP) ω_t is defined

$$\omega_t = P(Y \in W \mid H_1) = \sum_{y \in W} P(Y = y \mid H_1) = \sum_{y \in W} \prod_{i=1}^{2} \frac{t^{y_i} \beta_i^{\alpha_i} \Gamma(y_i + \alpha_i)}{y_i ! \Gamma(\alpha_i)(t + \beta_i)^{y_i + \alpha_i}},$$

and the Expected Bayesian Significance Level (EBSL) α_t is

$$\alpha_t = P(Y \in W \mid H_0) = \sum_{y \in W} P(Y = y \mid H_0) = \sum_{y \in W} \frac{t^{y_1 + y_2} \beta^\alpha \Gamma(y_1 + y_2 + \alpha)}{y_1 ! y_2 ! \Gamma(\alpha)(2t + \beta)^{y_1 + y_2 + \alpha}}.$$

Note three things. First the inclusion of the t subscripts highlights the fact that these quantities depend on t. Second, both ω_t and α_t marginalize over the corresponding alternative and null spaces Λ_1 and Λ_0, respectively, this is the sense in which the power and significance level are expectations. Third, the constant c (or c_1 and c_2) is represented in the expressions through W_t, which is itself dependent on t.

In their articles,[5,12] demonstrate that as the sample size tends to infinity, the Bayes factor B converges to either 0 or 1. As a consequence, in the current context as the sample size t tends to infinity the Bayes factor B converges to either 0 or 1, so that[12] implies that $\omega_t \xrightarrow{a.s.} 1$ and $\alpha_t \xrightarrow{a.s.} 0$. Thus, for any pre-specified power ω and significance level α, there exists a t such that for all $t' \geq t^*$, $\omega_{t'} \geq \omega$ and $\alpha_{t'} \geq \alpha$. We define $t_{\alpha,\omega}^*$ to be the infimum of this collection of lower bounds, i.e.

$$t_{\alpha,\omega}^* = \inf_{t \in R_+} \{ t : \alpha_t \leq \alpha \text{ and } \omega_t \geq \omega \}.$$

$t_{\alpha,\omega}^*$ is said to be the optimal sample size for ω and α, and computing t^* is called the sample size determination problem. If only a power is specified, or if only a significance level is specified, the other quantity is left off of the subscript and out of the definition. We often write simply t^* for $t_{\alpha,\omega}^*$.

Were (10) and (11) monotonic and continuous in t, the sample size determination problem would be quite straightforward. Simply run a numerical root-finder (e.g. Newton's method) on $\omega_t - \omega$ and $\alpha_t - \alpha$ take the larger t. Unfortunately, however, as a function of t both the power and significance level are discontinuous functions that are not monotonic. As a consequence, it is possible, for example, for there to be two such sample sizes t_1 and t_2 with $t_1 < t_2$ such that $\omega_{t_1} \geq \omega$ and yet $\omega_{t_2} < \omega$. This is a consequence of the dependence of W_t on t: as t grows, the rejection region changes, and these changes result in discrete jumps in the rejection region.

Practically speaking, in our experience the appearance of these jumps is monotonically decreasing in magnitude and dissipate quite quickly in t so that, while there are jumps, they become relatively minor even for quite small t. Thus, while out-of-the-box numerical routines are insufficient for the task, a straight-forward heuristic algorithm suffices to certify t^* to a reasonable level of accuracy.

Initialize $t = t_0$ with a value small enough to satisfy $\omega_t < \omega$ and $\alpha_t > \alpha$, typically a value like .1 will suffice. Then, double t until both conditions are met. Once both conditions are met, decrement t by 1 until the conditions are no longer satisfied; then decrement by .1, and so on to achieve the desired precision. Alternatively, one may move in

a binary search manner; this method is faster but loses the certificate of the solution up to the highest evaluated point. By contrast, the first heuristic certifies that the sample size achieved is optimal up to the largest t observed, which is of the form $2^k t_0$.

One computational detail is relevant for implementing this procedure. By definition, since Y_1 and Y_2 are Poisson variates, their sample spaces are infinitely large, and thus computing ω_t and α_t is not numerically possible - one cannot check every y to determine whether or not it is included in W_t. There is, however, a very reasonable work-around for this problem using prior predictive distributions. Under H_1, the prior predictive distributions on Y_1 and Y_2 are both negative binomial. Specifically,

$$Y_1 \sim \text{NegBin}\left(\alpha_1, \frac{\beta_1}{t + \beta_1}\right) \text{ and } Y_2 \sim \text{NegBin}\left(\alpha_2, \frac{\beta_2}{t + \beta_2}\right).$$

While one cannot enumerate the entire sample space of Y, one can be confident they have a satisfactory approximation by computing small (e.g. .00001) and large (e.g. .99999) quintiles of these distributions and then simply taking every combination of the ranges from low to high. This is the approach we take to computing the sums listed in (10) and (11) above.

An example from cancer therapeutics

1) An example of the proposed methodology is readily available from the field of cancer therapeutics. Suppose (1) Drug A is an industry standard therapy for a certain type of cancer that is known to have the common side effect of mild seizures every hour of infusion

2) (2) Drug B is a novel compound believed to have the same side effect but at a lessened rate

3) (3) The goal is to design a design a clinical trial that compares the two using the minimum resources required to meet 5% significance level (EBSL) and 80% power (EBP). Moreover, suppose that the losses associated with Type I and Type II errors have a ratio of 1:1.

From past studies, it is known that the uncertainty in the rate of seizures with Drug A (per hour) is well-represented by a Gamma (4, 4) distribution so that $\lambda_A \sim \text{Gamma}(4,4)$. Drug B, by contrast, is believed to be a bit worse, with perhaps a rate that is double that of Drug A with experts 90% sure the value is less than about 3. This translates into roughly $\lambda_B \sim \text{Gamma}(4,8)$. Figure 1 shows both of these priors graphically.
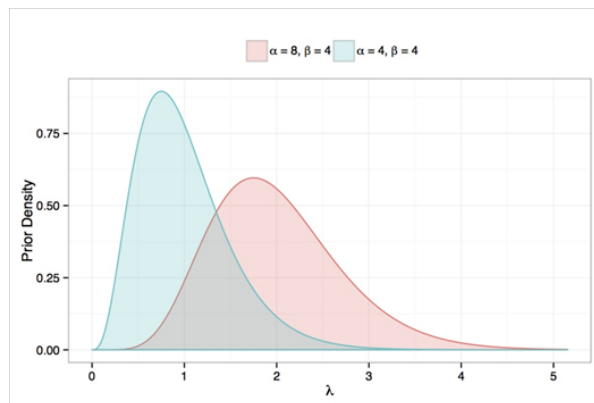

Figure 1 Prior structures used in Poisson sample size determination example.

Assuming that the rates are the same, the past indicators of Drug A supersede the lack of evidence for Drug B, so that under the null hypothesis $\lambda = \lambda_1 = \lambda_2 \sim \mathrm{Gamma}(4,4)$ is most appropriate. Assuming that the null and alternative hypotheses are given the same belief (50%), the rejection rule from (9) is therefore

$$W = \left\{ \boldsymbol{y} : \frac{\Gamma(4)(2t+4)^{y_1+y_2+4}}{4^4\Gamma(y_1+y_2+4)} \frac{4^4\Gamma(y_1+4)}{\Gamma(4)(t+4)^{y_1+4}} \frac{4^8\Gamma(y_2+8)}{\Gamma(8)(t+4)^{y_2+8}} \geq 3 \right\}$$

$$= \left\{ \boldsymbol{y} : \frac{(2t+4)^{y_1+y_2+4}}{(t+4)^{y_1+4}} \geq \frac{15120}{65536} \frac{(y_1+y_2+3)!}{(y_1+3)!(y_2+7)!} \right\}$$

To design a test with 80% power and at the 5% significance level, we need but to run the algorithm described in Section 3. To illustrate the scenario, we plot the significance level and power for every t from 1 to 80; these are included in Figures 2 and 3, respectively. Note how quickly (in t) the functions become smooth, nullifying any concern about jumps. The 80% level is achieved at $t = 37$, with a power of 80.1%, and the 5% significance level is achieved at $t = 57$, where the significance level is 4.9%. Thus, to achieve both, we select the higher sample size, $t = 57$.
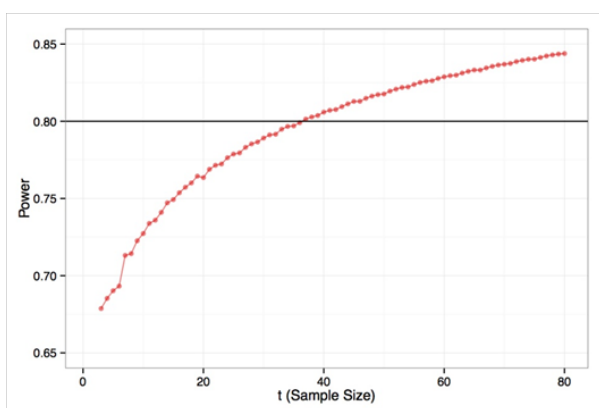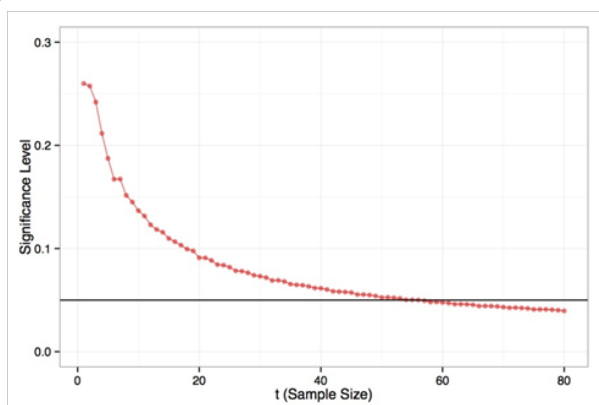

Figure 2 EBP curve

Figure 3 EBSL curve

We can also verify these results via simulation by generating one million random values of λ_1 and λ_2 from the priors given above. To validate the significance level result, we simply (1) sample from the prior, (2) generate two Poisson observations with mean equal to the sample in (1) times 57, and (3) determine whether the test rejects (1) or not (0). Averaging the million test results, the value is confirmed to Monte Carlo error (code available upon request). To validate the power result, we (1) sample one number from a Gamma(4, 4)and one number from a Gamma(8, 4), (2) sample two Poisson observations from distributions with mean 37 times the two variates generated in

(1), and (3) determine whether the test rejects or not. Averaging the million results validates the theoretical result.

Discussion

In this paper we have used conjugate prior structures in order to assess our beliefs about a rate parameter in a two sample Poisson trial a priori in order to find the minimal sample size needed to reach certain operating characteristics. By the use of a loss function constant, we are able to control for either the desired expected Bayesian significance level or the desired expected Bayesian power or both. This type of analysis had not been considered previously, specifically, simultaneously controlling for both Type I and Type II error.

Future work includes the consideration of analysis priors in this research. Ideally, we would be able to adapt this research to account for the fact that researchers often times use one set of priors when conducting sample size analyses, but a more vague or non-informative set of priors when actually analyzing the experiment. Though we used only mildly informative priors in our example, it may be the case where a substantially informative prior is required for the design stage but a less informative prior would be used in the analysis stage. Further, this process could be expanded to consider non-conjugate priors as well. However, the analytical tractability of conjugate priors made the Poisson/gamma model ideal, and modeling prior beliefs of a Poisson rate with a gamma distribution is not an unreasonable thing to do.

It also should be noted that time considerations, while already improved throughout the process, can always continue to improve. One improvement to current methods involves replacing the bi-sectional approaches with one that approximates the EBP and EBSL curves with some logarithmic function; this improvement should get us in the ballpark of a candidate solution much quicker. Future work also includes a more in depth look at how expected Bayesian error rates compare to typical frequentist ones, and potentially an in-depth look at the different sample size determination and testing methods in order to determine the relative advantages and disadvantages of each. Further, we could generalize the algorithm such that we are not looking at a common sample size $t = t_1 = t_2$, but rather two different sample sizes t_1 and t_2 such that they do not need to be equal. Lastly, code is available upon request.

Acknowledgements

Conflict of interest

None.

References

1. Adcock CJ. Sample Size Determination: A Review. *The Statistician*. 1997;46(2):261–283.

2. Bayarri MJ, Berger JO. The Interplay of Bayesian and Frequentist Analysis. *Statistical Science*. 2004;19(1):58–80.

3. Stamey JD, Young DM, Bratcher TL. Bayesian Sample-Size Determination for One and Two Poisson Rate Parameters with Applications to Quality Control. *Journal of Applied Statistics*. 2006;33(6):583–594.

4. Hand A, Stamey JD, Young DM. Bayesian Sample-Size Determination for Two Independent Poisson Rates. *Computer Methods and Programs in Biomedicine*. 2011;104(2):271–277.

5. Katsis A, Toman B. Bayesian Sample Size Calculations for Binomial Experiments. *Journal of Statistical Planning and Inference.* 1992;81(2):349–362.

6. Zhao Z, Tang N, Li Y. Sample-Size Determination for Two Independent Binomial Experiments. *Journal of Systems Science and Complexity.* 2011;24(5):981–990.

7. O'Hagan A, Stevens J, Campbell M. Assurance in Clinical Design. *Pharmaceutical Statistics.* 2005;4(3):187–201.

8. Speighelhalter DJ, Abrams K R, Myles JP. Bayesian Approaches to Clinical Trials and Health Care Evaluation. *Wiley.* 2004.

9. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association.* 2005;100:680–701.

10. Shao J. *Mathematical statistics.* 2nd ed. Springer; 2003.

11. Berger J. *Statistical decision theory and Bayesian analysis.* Springer; 1985.

12. Kass R, Raftery A. Bayes Factors. *Journal of the American Statistical Association.* 1995;90(430):773–795.