RT&A, No 3 (79)
Volume 19, September 2024

BAYESIAN NON-INFERIORITY TEST BETWEEN TWO BINOMIAL PROPORTIONS

W. B. Yahya[1], C. P. Ezenweke[2], O. R. Olaniran[3], I. A. Adeniyi[4], K. Jimoh[5], R. B. Afolayan[6], M. K. Garba[7], I. Ahmed[8]

•

[1,3,6,7]Department of Statistics, University of Ilorin, Ilorin, Nigeria
[1]wbyahya@unilorin.edu.ng; [3]rid4stat@yahoo.com;
[6]afolayan.rb@unilorin.edu.ng ; [7]garba.mk@unilorin.edu.ng

[2,4]Department of Mathematics, Federal University Lokoja, Nigeria
[2]chinenye.ezenweke@fulokoja.edu.ng; [4]isaac.adeniyi@fulokoja.edu.ng

[5]Department of Physical Sciences, Al-Hikmah University, Ilorin, Nigeria
[5]jimkaminsha@alhikmah.edu.ng

[8]Department of Statistics, Nasarawa State University, Keffi, Nigeria
[8]ibrahimloko@nsuk.edu.ng

Abstract

The paper aimed to propose a new Bayesian test method for establishing a non-inferiority measure between an active treatment (drug) and a new (cheaper) treatment using two independent binomial samples. A Bayesian test statistic was developed for testing non-inferiority between two independent binomial proportions. Conjugate Beta prior was assumed for the binomial proportions to elicit posterior from the same Beta family of distributions. The efficiency of this test method was established via power analysis and its ability to yield the nominal Type I error rate (alpha) in a detailed Monte-Carlo study. Results from this study showed that the proposed test method yielded higher powers and good estimates of the Type I error rate at the chosen sample sizes and varying non-inferiority margins (effect sizes). Thus, the new Bayesian test method is very efficient at detecting the significance of the non-inferiority margin between two independent binomial proportions when such is not negligible at all sample sizes. Further results showed that the size of the two population proportions being tested influences the power and the estimated nominal Type I error rate with an increase in power and a good estimate of Type I error rate achieved when both population proportions being tested are less than 0.5. It is therefore concluded that the new Bayesian test method can be employed whenever it is desirable to establish the existence of non-inferiority or otherwise between a pair of (clinical) treatments (drugs). All the simulations and analyses were performed with the R statistical package.

Keywords Non-inferiority test, Test of proportion, Bayesian Inference, Conjugate prior, Power

1. Introduction

In pharmacological and drug discovery studies, the need for the development of new drugs to compete with some of the existing ones for a particular ill-health condition is often desirable for many reasons. Clinicians and end-users may prefer a new drug (or treatment) if significant advantages can be derived from its use such as having fewer side effects, being less toxic, being relatively cheaper, being relatively more convenient in its formulation and administration (e.g. tablets instead of infusion or infusion/drug instead of surgery) and so on (Yahya et al. [1]). Therefore, it is always a welcome development that improved, better, and (possibly) cheaper drugs or treatments are discovered and introduced for the treatment of various forms of ailments.

In clinical trials, a new drug (or treatment) for a particular medical condition can only be accepted as being good if its performance is at least as good as an existing one (the active control). In other words, a new drug may be preferred if its performance is equivalent to or better than (not inferior to) the active control (Chen and Peace [2], Ng [3], Norleans [4]). This is the scenario that brought about the concept of non-inferiority between two drugs or treatments in clinical trials.

There are different methods for determining if a new drug (or treatment), often referred to as the test product, is as good as an existing one, most of which involve the use of statistical tools. Some of the procedures for determining if the performance of a new drug is not inferior to an existing one include *superiority trials* and *non-inferiority trials* (Lesaffre [5], Yahya and Jolayemi [6]).

A superiority trial is used to prove that a new drug is better than an existing one, while a non-inferiority trial is used to show that the new drug is not (much) worse than the existing one (Kawasaki et al. [7]). Many clinical trials that compare a test product with an active comparator are designed as *non-inferiority* trials (Ng [3], EMA [8]). This is attributed to the fact that a placebo (controlled trial) is considered unethical or impractical (Temple [9]) among other reasons. In this study, our focus is much on *non-inferiority trials* based on nominal outcomes from two independent proportions, and a new test method to handle such a situation is proposed.

The term '*non-inferiority trial*' is commonly used to refer to a randomized clinical trial in which a new treatment is compared with a standard active treatment rather than a placebo or untreated control group (D'Agostino et al. [10], Pocock [11]). The non-inferiority trial is a trial to show how an experimental treatment is statistically and clinically not inferior to the active control, see D'Agostino et al. [10]. In clinical trials, non-inferiority tests are frequently used to demonstrate that the response for study drugs is not much worse than the response for reference drugs, see Kawasaki et al. [7], Kawasaki and Miyaoka [12].

1.1 Non-Inferiority Tests

A *non-inferiority test* is used to indicate whether the responses from new drugs are clinically not much worse than the response to active control. It is often conducted in clinical trials. Many clinical trials that compare a test product with an active comparator or control are designed as non-inferiority trials. A non-inferiority trial is sometimes stated as being to demonstrate that the test product is not inferior to the active control. However, only a superiority trial can demonstrate this. A non-inferiority trial aims to demonstrate that the test product is not worse than the active control by more than a pre-specified threshold, often called the *non-inferiority margin,* Δ, see Ng [3], EMA [8], D'Agostino et al. [10].

Non-inferiority studies are typically confirmatory trials that employ a randomized parallel-group design with an active control group. Some trials also include a placebo control group. The placebo is used to validate the study and to demonstrate the superiority of the test treatment to the

placebo (Garrett [13]). Non-inferiority methods are frequently used in more serious, acute, and sometimes life-threatening situations such as oncology and infectious diseases, see Garrett [13].

Without loss of generality, a non-inferiority test is used to indicate whether the responses from a new drug are clinically not much worse than the responses from the active control (an existing drug) as it is often conducted in clinical trials. Thus, a prior judgment is made that, for the new treatment to be of merit, it only needs to be as good as the active control regarding appropriate outcome measure(s) of response, see Pocock [11].

Non-inferiority tests are examples of directional (one-sided) tests. There are many forms of this test depending on the form of the response. The *non-inferiority* tests for two-sample designs in which the outcome is a continuous normal random variable were carried out using the two-sample t-test procedure and the analysis of variance (ANOVA) *k*-sample F-test procedure when there are more than two samples (Yahya and Jolayemi [6]). When the response or outcome is binary, the non-inferiority test procedure, which is the main focus of this study, involves determining if the difference between two binomial proportions is beyond a specified non-inferiority margin Δ.

1.2 Non-Inferiority Test Between Two Independent Proportions

The responses from non-inferiority studies can be binary, for example, cure or no cure, alive or dead, cancerous or non-cancerous, and so on. When this is the case, the test product group and active control group are referred to as two independent binominal samples. The non-inferiority tests for this case are carried out using the tests for the differences between two proportions. For instance, suppose that the current drug (treatment) for a given disease works 80% of the time. Suppose also that this treatment is either too expensive or occasionally exhibits serious side effects or that its administration is considered quite tedious (e.g. surgery) and a new promising treatment or drug has been developed to the point where it can be tested. One of the questions that must be answered is whether the new treatment is as good as the current treatment. In other words, do at least 80% of treated subjects respond positively to the new treatment?

Due to the many benefits (being cheaper, having relatively fewer side effects, and being very easy to administer) of the new treatment, clinicians would be willing to adopt the new treatment even if it is slightly less effective than the current treatment by some small tolerable margin. In order words, it is of interest to determine how much less effective the new treatment can be relative to the existing one (the active control) for which the new treatment can still be confidently adopted. This is called the *non-inferiority margin* of the new drug compared to the existing one (Ng [3], Yahya and Jolayemi [12]). The non-inferior margin represents how much worse the new treatment can be compared with the standard treatment, yet still, be considered 'similar' or 'not worse' than the standard treatment, Leung et al. [14]. In essence, there is a performance level of the new treatment as compared to the existing active control treatment below or above which is no longer considered ignorable.

In the above example, suppose it was found that the performance rate of the new drug is 75% among the population of users and this was considered acceptable by the clinicians and other users due to its numerous attendant benefits over the existing treatment. This simply shows that the *non-inferiority margin* of the new drug is 5%, i.e. $\Delta = 5\%$ and the new treatment is thus, considered non-inferior to the existing treatment. Thus, the drug developers need to design an experiment to test the hypothesis that the response rate of the new drug (treatment) is at least 0.75. In other words, the test hypothesis problem here is to establish whether the non-inferiority margin of 5% is exceeded by the new drug or not.

Different test methods have been proposed for establishing non-inferiority between a pair of two independent population proportions if such exists. Most of these methods were developed through the frequentist techniques while a few had appeared within the Bayesian framework. In

this work, a new Bayesian test method for testing the difference between two population proportions within the context of the non-inferiority formulation is proposed. In the development of this new test method, the emphasis is more on the effect sizes rather than the p-values of the statistical tests as opined by Leung et al. [14].

2. Methods

Let X_j be an independent Bernoulli random variable that indicates whether a clinician or patient prefers treatment/drug j with the associated probability of success π_j, while $j = 1$ for the existing active treatment and $j = 2$ for the new treatment. Therefore, in a total of n_j end-users of treatment/drug j, the random variable $\sum_{i=1}^{n_j} X_{ij}$ that represents the number of end-users that preferred treatment/drug j is distributed Binomial with parameters n_j and π_j. That is, $\sum_{i=1}^{n_j} X_{ij} \sim Bin(n_j \pi_j)$. The sample estimate of π_j is obtained as $\hat{\pi}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$, $j = 1,2$. For simplicity, the terms 'treatment' and 'drug' shall be used interchangeably in the subsequent discussions to represent the existing or the new clinical treatment adopted or preferred by the end-users.

To establish the efficiency of the new drug relative to the existing drug within the context of non-inferiority, the hypothesis of non-inferiority between the two drugs in terms of their relative preferences by end-users is constructed around the two independent population proportions π_1(for existing drug) and π_2 (for new drug), and the non-inferiority margin Δ_0 as follows:

$$H_0: \pi_1 \geq \pi_2 + \Delta_0 \text{ against } H_a: \pi_1 < \pi_2 + \Delta_0 \quad (1)$$

The inequality statement $\pi_1 \geq \pi_2 + \Delta_0$ in the null hypothesis H_0 simply indicates that the proportion (π_2) of users that preferred the new drug is only up to the proportion (π_1) of those that preferred the existing drug by a margin of the non-inferiority parameter value Δ_0.

Therefore, if the null hypothesis H_0 is not rejected by the test, it simply shows that the new drug is not inferior to the existing drug an indication that the non-inferiority margin Δ_0 is negligible. However, a rejection of the null hypothesis H_0 in favour of the alternative hypothesis H_a is a strong indication that the new drug is superior (more preferred by users) to the existing drug. Hence, the non-inferiority margin Δ_0, which has contributed significantly to the rejection of H_0 is not negligible.

Without loss of generality, the hypothesis set in (1) can be re-expressed as follows (with the null hypothesis H_0 indicating no difference between the treatment and the active control groups):

$$H_0: \pi_1 - \pi_2 = \Delta_0 \text{ against } H_a: \pi_1 - \pi_2 < \Delta_0 \quad (2)$$

Some of the existing methods for testing the non-inferiority hypothesis set in (2) include the z-test with a *pooled variance* given by

$$Z_p = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - \Delta_0}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\pi}_c(1-\hat{\pi}_c)}} \sim N(0,1) \quad (3)$$

where, $\hat{\pi}_c$ is computed as $\hat{\pi}_c = \frac{\sum_{i=1}^{n_1} x_{i1} + \sum_{i=1}^{n_2} x_{i2}}{n_1 + n_2}$.

Another test statistic for hypothesis in (2) is the Wald statistic, Ward and Ahlquist [15] given by:

$$Z_w = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - \Delta_0}{\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}} \sim N(0,1) \quad (4)$$

It has been indicated in some studies that the performance of the Wald statistic suffers when the sample size is small, Di-Caterina and Kosmidis [16]. However, Munzel and Hsuschke [17] have shown the framework of the non-inferiority test for ordered categorical data. When the number of categories is two, it can be regarded as a problem of the difference between two proportions. Hence, the Wald test statistic is derived by extending the method proposed by Munzel and Hsuschke [17] to the non-inferiority test for deriving the difference between two proportions.

In this paper, an attempt is made to extend the Wald-test procedures from a Bayesian perspective.

2.1. The Posterior Distribution of Bayesian Binomial Proportions

The posterior distribution $p(\pi|X)$is defined as:

$$p(\pi|X) = \frac{L(\pi/X)\times p(\pi)}{\int_{\pi} L(\pi/X)\times p\ (\pi)d\pi} \tag{5}$$

Given that $f(x) = \pi^{x}(1-\pi)^{1-x} x = 0,1$; then, the likelihood distribution function $L(\pi/X)$ is defined by

$$L(\pi/X) = \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i}$$

$$\rightarrow \quad L(\pi/X) = \pi^{\sum x_i}(1-\pi)^{n-\sum x_i} \tag{6}$$

Therefore, the Maximum Likelihood Estimator (MLE) of π is given by

$$\text{MLE}\ (\pi) = \hat{\pi} = \frac{\sum x_i}{n} \tag{7}$$

Assuming a conjugate beta prior for the likelihood in order to ascertain the same class of the posterior, the prior distribution, $p(\pi)$ is defined as;

$$p(\pi) = \frac{\pi^{a_0-1}(1-\pi)^{b_0-1}}{\Gamma(a_0)\times\Gamma(b_0)}\Gamma(a_0 + b_0)\ , 0 < \pi < 1 \tag{8}$$

Therefore, the posterior distribution $p(\pi|X)$ is determined using (5) as

$$p(\pi|X) = \frac{\pi^{\sum x_i}(1-\pi)^{n-\sum x_i} \times \frac{\pi^{a_0-1}(1-\pi)^{b_0-1}}{\Gamma(a_0)\times\Gamma(b_0)}\Gamma(a_0 + b_0)}{\int_0^1 \pi^{\sum x_i}(1-\pi)^{n-\sum x_i} \times \frac{\pi^{a_0-1}(1-\pi)^{b_0-1}}{\Gamma(a_0)\times\Gamma(b_0)}\Gamma(a_0 + b_0)d\pi}$$

$$p(\pi|X) = \frac{\pi^{\sum x_i}(1-\pi)^{n-\sum x_i} \times \pi^{a_0-1}(1-\pi)^{b_0-1}}{\int_0^1 \pi^{\sum x_i}(1-\pi)^{n-\sum x_i} \times \pi^{a_0-1}(1-\pi)^{b_0-1}d\pi}$$

$$p(\pi|X) = \frac{\pi^{a_0+\sum x_i-1}(1-\pi)^{b_0+n-\sum x_i-1}}{\int_0^1 \pi^{a_0+\sum x_i-1}(1-\pi)^{b_0+n-\sum x_i-1}d\pi} \tag{9}$$

Recall from the Beta function that,

$$\int_0^1 \pi^{a_0+\sum x_i-1}(1-\pi)^{b_0+n-\sum x_i-1}d\pi = \frac{\Gamma(a_0+\sum x_i-1)\times\Gamma(b_0+n-\sum x_i-1)}{\Gamma(a_0+b_0+n-2)} \tag{10}$$

Therefore, the prior distribution in (10) becomes;

$$p(\pi|X) = \pi^{a_0+\sum x_i-1}(1-\pi)^{b_0+n-\sum x_i-1} \times \frac{\Gamma(a_0+\sum x_i-1)\times\Gamma(b_0+n-\sum x_i-1)}{\Gamma(a_0+b_0+n-2)} \quad (11)$$

Thus,

$$p(\pi|X) \sim Beta\ (a_0+\sum x_i\ b_0+n-\sum x_i), 0<\pi<1 \quad (12)$$

If we let $a = a_0 + \sum x_i$; $b = b_0 + n - \sum x_i$ in (11) and (12), the posterior mean, $E(\pi|x)$ and posterior variance, $Var(\pi|x)$ can be obtained as;

$$E(\pi|x) = \frac{a}{a+b} \quad (13)$$

$$Var(\pi|x) = \frac{ab}{(a+b)^2(a+b+1)} \quad (14)$$

2.2 Prior Elicitation

2.2.1 *Choosing a Conjugate Prior by Matching Location and Scale Parameters*

Given that the $Beta(ab)$ family of distributions is the conjugate family for $Binomial(n\pi)$ distribution, the posterior will be a member of the same family, with the parameters updated by simple rules. Then, we can find the posterior without integration. Note that, the beta distribution can have many shapes. Therefore, the prior chosen should correspond to one's belief. We suggest choosing a $Beta(a_0 b_0)$ that matches one's prior belief about the mean (location) and standard deviation (scale).

Let π_0 be the prior mean for the proportion and let σ_0^2 be the prior variance for the proportion. But, we know that the mean of $Beta(ab)$ distribution is $\frac{a}{a+b}$. If this is set to equal to what the prior belief about the mean of the proportion is, we have;

$$\pi_0 = \frac{a_0}{a_0+b_0} \quad (16)$$

Also, the variance of the $Beta(ab)$ distribution is $\frac{ab}{(a+b)^2(a+b+1)}$. This shall be set to equal to the prior belief about the variance for the proportion to have:

$$\sigma_0^2 = \frac{a_0 b_0}{(a_0+b_0)^2(a_0+b_0+1)} \quad (17)$$

Note that $\pi_0 = \frac{a_0}{a_0+b_0}$ from (16), this implies that $1-\pi_0 = \frac{b_0}{a_0+b_0}$. Substituting these in (17) to have;

$$\sigma_0^2 = \frac{\pi_0(1-\pi_0)}{a_0+b_0+1} \quad (18)$$

Solving equations (16) and (18) for a_0 and b_0 gives the $Beta(a_0 b_0)$ prior parameters.

Proof: Given that $\pi \sim Beta(ab)$ then,

$$E(\pi) = \frac{a}{a+b};\ Var(\pi) = \frac{ab}{(a+b)^2(a+b+1)} \quad (19)$$

Solving for a_0 in terms of b_0 and π_0 from (16) gives

$$a_0 = \frac{b_0\pi_0}{1-\pi_0} \tag{20}$$

Substituting (20) in (18) and solving for b_0 gives

$$\sigma_0^2 = \frac{\pi_0(1-\pi_0)}{\frac{b_0\pi_0}{1-\pi_0} + b_0 + 1}$$

$$\sigma_0^2 = \frac{\pi_0(1-\pi_0)}{\frac{b_0\pi_0 + b_0 - b_0\pi_0 + 1 - \pi_0}{1-\pi_0}}$$

$$\sigma_0^2 = \frac{\pi_0(1-\pi_0)}{\frac{b_0 + 1 - \pi_0}{1-\pi_0}}$$

$$\sigma_0^2 = \pi_0(1-\pi_0) \times \frac{1-\pi_0}{b_0 + 1 - \pi_0}$$

$$\sigma_0^2 = \frac{\pi_0(1-\pi_0)^2}{b_0 + 1 - \pi_0}$$

$$\sigma_0^2(b_0 + 1 - \pi_0) = \pi_0(1-\pi_0)^2$$

$$b_0\sigma_0^2 + \sigma_0^2 - \sigma_0^2\pi_0 = \pi_0(1-\pi_0)^2$$

$$b_0\sigma_0^2 + \sigma_0^2(1-\pi_0) = \pi_0(1-\pi_0)^2$$

$$b_0\sigma_0^2 = \pi_0(1-\pi_0)^2 - \sigma_0^2(1-\pi_0)$$

$$b_0 = \frac{\pi_0(1-\pi_0)^2 - \sigma_0^2(1-\pi_0)}{\sigma_0^2}$$

$$b_0 = \frac{(1-\pi_0)[\pi_0(1-\pi_0) - \sigma_0^2]}{\sigma_0^2} \tag{21}$$

Substituting b_0 in (21) into (20) to have

$$a_0 = \frac{\frac{(1-\pi_0)[\pi_0(1-\pi_0) - \sigma_0^2]}{\sigma_0^2}\pi_0}{1-\pi_0}$$

$$a_0 = \frac{\pi_0(1-\pi_0)[\pi_0(1-\pi_0) - \sigma_0^2]}{(1-\pi_0)\sigma_0^2}$$

$$a_0 = \frac{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2]}{\sigma_0^2} \tag{22}$$

Therefore, the posterior mean, $\hat{\pi}$ and posterior variance, $\hat{\sigma}_\pi^2$ are determined as follows. We recall from (13) and (14) that

$$E(\pi|x) = \frac{a}{a+b} = \hat{\pi}$$

and

$$V(\pi|x) = \frac{ab}{(a+b)^2(a+b+1)}$$

respectively.

$$\rightarrow V(\pi|x) = \frac{a}{a+b} \times \frac{b}{a+b} \times \frac{1}{a+b+1} = \hat{\sigma}_\pi^2$$

Also, recall that $a = a_0 + \sum x_i; b = b_0 + n - \sum x_i$ then, by proper substitution, we have that

$$\hat{\pi} = \frac{a_0 + \sum x_i}{a_0 + \sum x_i + b_0 + n - \sum x_i}$$

$$\rightarrow \quad \hat{\pi} = \frac{a_0 + \sum x_i}{a_0 + b_0 + n} \tag{23}$$

$$\hat{\sigma}_{\pi}^2 = \frac{a_0 + \sum x_i}{a_0 + \sum x_i + b_0 + n - \sum x_i} \times \frac{b_0 + n - \sum x_i}{a_0 + \sum x_i + b_0 + n - \sum x_i} \times \frac{1}{a_0 + \sum x_i + b_0 + n - \sum x_i + 1}$$

$$\rightarrow \hat{\sigma}_{\pi}^2 = \frac{a_0 + \sum x_i}{a_0 + b_0 + n} \times \frac{b_0 + n - \sum x_i}{a_0 + b_0 + n} \times \frac{1}{a_0 + b_0 + n + 1} \tag{24}$$

By substituting (21) and (22) in (23) for a_0 and b_0 the Bayesian posterior mean of binomial proportion π is determined as follows;

$$\hat{\pi} = \frac{\frac{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2]}{\sigma_0^2} + \sum x_i}{\frac{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2]}{\sigma_0^2} + \frac{(1-\pi_0)[\pi_0(1-\pi_0) - \sigma_0^2]}{\sigma_0^2} + n}$$

$$\hat{\pi} = \frac{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2] + \sigma_0^2 \sum x_i}{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2] + \pi_0(1-\pi_0)^2 - \sigma_0^2(1-\pi_0) + \sigma_0^2 n}$$

$$\hat{\pi} = \frac{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2] + \sigma_0^2 \sum x_i}{[\pi_0(1-\pi_0) - \sigma_0^2][\pi_0 + (1-\pi_0)] + \sigma_0^2 n}$$

$$\rightarrow \quad \hat{\pi} = \frac{\pi_0[\pi_0(1-\pi_0) - \sigma_0^2] + \sigma_0^2 \sum x_i}{\pi_0(1-\pi_0) + \sigma_0^2(n-1)} \tag{25}$$

Also, the Bayesian posterior variance $\hat{\sigma}_{\pi}^2$ of binomial proportion π is derived by substituting (21) and (22) in (24) as

$$\hat{\sigma}_{\pi}^2 = \frac{\hat{\pi}(1-\hat{\pi})\sigma_0^2}{[\pi_0(1-\pi_0) - \sigma_0^2] + \sigma_0^2 n + \sigma_0^2}$$

$$\rightarrow \quad \hat{\sigma}_{\pi}^2 = \frac{\hat{\pi}(1-\hat{\pi})\sigma_0^2}{\pi_0(1-\pi_0) + n\sigma_0^2} \tag{26}$$

where $\hat{\pi} = \frac{\sum x_i}{n}$ as given in (7).

2.3 Proposed Bayesian Non-inferiority Test of Two Independent Binomial population proportions

Following the above Bayesian estimates of mean and variance of a binomial proportion, the proposed Bayesian non-inferiority test statistic for testing the hypothesis set

$$H_0: \pi_1 - \pi_2 = \Delta_0 \text{ against } H_a: \pi_1 - \pi_2 < \Delta_0$$

as earlier stated in (2) is;

$$Z_B = \frac{\hat{\pi}_1 - \hat{\pi}_2 - \Delta_0}{\sqrt{\sigma_{\hat{\pi}_1}^2 + \sigma_{\hat{\pi}_2}^2}} \sim N(0,1) \tag{27}$$

where

$$\hat{\pi}_1 = \frac{\pi_{01}[\pi_{01}(1-\pi_{01}) - \sigma_{01}^2] + \sigma_{01}^2 \sum x_{1i}}{\pi_{01}(1-\pi_{01}) + \sigma_{01}^2(n_1 - 1)} \tag{28}$$

$$\hat{\pi}_2 = \frac{\pi_{02}[\pi_{02}(1-\pi_{02}) - \sigma_{02}^2] + \sigma_{02}^2 \sum x_{2i}}{\pi_{02}(1-\pi_{02}) + \sigma_{02}^2(n_2 - 1)} \tag{29}$$

$$\hat{\sigma}^{\,2}_{\hat{\pi}_1} = \frac{\hat{\pi}_1(1-\hat{\pi}_1)\times \sigma^{\,2}_{01}}{\pi_{01}(1-\pi_{01}) + \sigma^2_{01} n_1} \tag{30}$$

$$\hat{\sigma}^{\,2}_{\hat{\pi}_2} = \frac{\hat{\pi}_2(1-\hat{\pi}_2)\times \sigma^{\,2}_{02}}{\pi_{02}(1-\pi_{02}) + \sigma^2_{02} n_2} \tag{31}$$

and $\hat{\pi}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}\, j = 12$.

2.4 Decision Rule

The decision rule for the test function is to reject H_0 if $\hat{Z}_B < -Z_{1-\alpha}$ where $Z_{1-\alpha}$ is the quantile of the standard normal distribution at Type I error rate α.

If the null hypothesis H_0 is not rejected by the test function (27), it simply shows that the new drug is not inferior to the existing drug an indication that the non-inferiority margin Δ_0 is negligible. However, a rejection of the null hypothesis H_0 in favour of the alternative hypothesis H_a by the test is a strong indication that the new drug is superior (more preferred by users) to the existing drug. Hence, the non-inferiority margin Δ_0, which has contributed significantly to the rejection of H_0 is not negligible.

2.5 The Power and Type I Error Rate (α) of the Proposed Bayesian Non-inferiority Test

In computing the power and nominal Type I error rate of the proposed Bayesian Non-inferiority test method, we consider the hypothesis of the non-inferiority test stated in (2), its test statistic as well as its decision rule. The hypothesis set to test is

$$H_0 \pi_{\ 1} - \pi_2 = \Delta_0 \text{ against } H_a \pi_{\ 1} - \pi_2 < \Delta_0$$

with the proposed test statistic for the above test according to (27) is of the form

$$Z_B = \frac{\hat{\Delta}_0 - \Delta_0}{SE(\hat{\Delta}_0)} \sim N(01) \tag{32}$$

where $\hat{\Delta}_0 = \hat{\pi}_1 - \hat{\pi}_2$ and $SE(\hat{\Delta}_0)$ is the standard error of $\hat{\Delta}_0$. The decision rule for the test function as earlier stated is to reject H_0 if $\hat{Z}_B < -Z_{1-\alpha}$ where $Z_{1-\alpha}$ is the quantile of the standard normal distribution at Type I error rate α. Note that, $SE(\hat{\Delta}_0) = \sqrt{\sigma^2_{\hat{\pi}_1} + \sigma^2_{\hat{\pi}_2}}$.

2.5.1 *Power Computation*

The power of the statistical test of size α for testing the null hypothesis H_0 against the alternative set H_a is the probability that the test rejects H_0 when H_a is true. That is;

$$\text{Power} = P[Testfunction Rejects H_{\ 0} | H_a is\ \ true]$$

Therefore, given the above decision rule for testing the hypothesis (2) above at some specified Type I error rate α, the power of the test is defined by

$$\text{Power} = P[Z_{cal} < -Z_{1-\alpha} | H_a istrue\ \ ]$$

$$\rightarrow \quad \text{Power} = P\left[\frac{\hat{\Delta}_0-\Delta_0}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha}|\pi_1-\pi_2 < \Delta_0\right] \tag{33}$$

Suppose Δ_1 is the true difference between the two proportions π_1 and π_2 for the alternative hypothesis H_a to be true, hence, the test hypothesis (2) can be re-expressed as

$$H_0: \pi_1 - \pi_2 = \Delta_0 \text{ against } H_a: \pi_1 - \pi_2 = \Delta_1, \text{ with } \Delta_1 < \Delta_0.$$

The expression for power of the test in (33) then becomes;

$$\text{Power} = P\left[\frac{\hat{\Delta}_0-\Delta_0}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha}|\Delta_1\right] \tag{34}$$

Thus, we have;

$$\text{Power} = P\left[\frac{\hat{\Delta}_0-\Delta_1}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha}\right]$$

$$\text{Power} = P\left[\frac{\hat{\Delta}_0}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha} + \frac{\Delta_1}{SE(\hat{\Delta}_0)}\right]$$

$$\rightarrow \quad \text{Power} = P\left[Z < -Z_{1-\alpha} + \frac{\Delta_1}{SE(\hat{\Delta}_0)}\right]$$

$$\therefore \quad \text{Power} = \Phi\left[-Z_{1-\alpha} + \frac{\Delta_1}{SE(\hat{\Delta}_0)}\right] \tag{35}$$

where Φ is the cumulative distribution function of the normal density and $SE(\hat{\Delta}_0)$ is the standard error of $\hat{\Delta}_0 = \hat{\pi}_1 - \hat{\pi}_2$. From (30) and (31), $SE(\hat{\Delta}_0)$ is computed by

$$SE(\hat{\Delta}_0) = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)\times\sigma_{01}^2}{\pi_{01}(1-\pi_{01}) + \sigma_{01}^2 n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)\times\sigma_{02}^2}{\pi_{02}(1-\pi_{02}) + \sigma_{02}^2 n_2}}$$

Therefore, the power of the proposed Bayesian non-inferiority test is given by (35).

2.5.2 *Nominal Type I Error Rate Computation*

The Type I error rate of a statistical test α is the probability that the test function rejects the null hypothesis H_0 when H_0 is actually true. Let the actual size alpha level of the test be denoted by $\hat{\alpha}$. This can be computed by;

$$\hat{\alpha} = P\left[\frac{\hat{\Delta}_0-\Delta_0}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha}|\Delta_0\right] \tag{36}$$

where Δ_0 is the null difference. Thus, we have that;

$$\hat{\alpha} = P\left[\frac{\hat{\Delta}_0 - \Delta_0}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha}\right]$$

$$\hat{\alpha} = P\left[\frac{\hat{\Delta}_0}{SE(\hat{\Delta}_0)} < -Z_{1-\alpha} + \frac{\Delta_0}{SE(\hat{\Delta}_0)}\right]$$

$$\rightarrow \quad \hat{\alpha} = P\left[Z < -Z_{1-\alpha} + \frac{\Delta_0}{SE(\hat{\Delta}_0)}\right]$$

$$\therefore \quad \hat{\alpha} = \Phi\left[-Z_{1-\alpha} + \frac{\Delta_0}{SE(\hat{\Delta}_0)}\right] \tag{37}$$

Thus, the power function and the estimator of the significance level α for the proposed Bayesian test statistic considered in this paper are presented in Table 1.

Table 1: *The Power function (Z_B-Power) and estimator of the Nominal Type I error rate α (Z_B-Alpha (α))of the proposed Bayesian test for two independent Binomial Proportions.*

Description	Estimator
Z_B-Alpha (α)	$\Phi\left[-Z_{1-\alpha}+\Delta_0 \div \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)\times\sigma_{01}^2}{\pi_{01}(1-\pi_{01})+\sigma_{01}^2 n_1}+\frac{\hat{\pi}_2(1-\hat{\pi}_2)\times\sigma_{02}^2}{\pi_{02}(1-\pi_{02})+\sigma_{02}^2 n_2}}\right]$
Z_B-Power	$\Phi\left[-Z_{1-\alpha}+\Delta_1 \div \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)\times\sigma_{01}^2}{\pi_{01}(1-\pi_{01})+\sigma_{01}^2 n_1}+\frac{\hat{\pi}_2(1-\hat{\pi}_2)\times\sigma_{02}^2}{\pi_{02}(1-\pi_{02})+\sigma_{02}^2 n_2}}\right]$

3. Simulation Study

3.1 Simulation Scheme

The data utilized for this work were simulated using the R statistical package (R Core Team [18]). Two independent binomial random variables X_1 and X_2 were generated at different sample sizes $n = 20, 40, 100$ using varying probabilities of success π_1 and π_2 with varying effect size Δ_0 where $\pi_1 = \pi_2 + \Delta_0$. The Power (%) and the nominal size α (%) values of the proposed test statistic were computed using 10,000 iterations of the basic experiment. The implementation of the Bayesian methodology is in two phases:

1. Obtaining the Bayesian estimate of $\hat{\pi}_1$ and $\hat{\pi}_2$ by updating the data using conjugate beta-prior.
2. Estimating the corresponding Power (%) and the nominal size α (%) values of the proposed Bayesian test statistic.

Recall that the non-inferiority hypothesis test is given as H_0: $\pi_1 - \pi_2 = \Delta_0$ vs. H_a: $\pi_1 - \pi_2 < \Delta_0$.

The section that follows presents the results for different values of π_1 and π_2 at various sample sizes.

4. Results

The results of the proposed non-inferiority test between two binomial proportions are presented in phases based on the different parameters combinations as provided under the simulation scheme.

Table 2 presents the estimated power of the proposed Bayesian test first at the effect size $\Delta_0 = 0.05$ but under two different sizes of proportion pairs ($\pi_1 = 0.2, \pi_2 = 0.25$) and ($\pi_1 = 0.6$, $\pi_2 = 0.65$) with the associated Bayesian test statistics Z_{B1} and Z_{B2} respectively at varying sample sizes. Table 2 equally presents the results of the estimated powers of the proposed test at the effect size $\Delta_0 = 0.1$ and under two different sizes of proportion pairs ($\pi_1 = 0.2, \pi_2 = 0.3$) and ($\pi_1 = 0.6$, $\pi_2 = 0.7$) with the associated Bayesian test statistics Z_{B3} and Z_{B4} respectively also at varying sample sizes. Expectedly, it can be observed that the power of the proposed Bayesian test method increases as the sample size increases.

Table 2: *The Power (in %) of the proposed Binomial Non-inferiority Test (ZB) at various sample sizes (n), with the following test parameters;* $\pi_1 = 0.2, \pi_2 = 0.25$, *effect size* $\Delta_0 = 0.05$ *for test* Z_{B1}; $\pi_1 = 0.6, \pi_2 = 0.65$ *effect size* $\Delta_0 = 0.05$ *for test* Z_{B2}; $\pi_1 = 0.2, \pi_2 = 0.3$, *effect size* $\Delta_0 = 0.1$ *for test* Z_{B3}; $\pi_1 = 0.6, \pi_2 = 0.7$ *effect size* $\Delta_0 = 0.1$ *for test* Z_{B4}.

Power of Bayesian Non-Inferiority Tests				
Sample Size (n)	Effect Size $\Delta_0 = 0.05$		Effect Size $\Delta_0 = 0.1$	
	Z_{B1} $(\pi_1, \pi_2) = (0.2,\ 0.25)$	Z_{B2} $(\pi_1, \pi_2) = (0.6,\ 0.65)$	Z_{B3} $(\pi_1, \pi_2) = (0.2,\ 0.3)$	Z_{B4} $(\pi_1, \pi_2) = (0.6,\ 0.7)$
20	21.91	16.35	70.42	65.69
40	34.92	28.16	85.01	79.77
60	44.29	37.23	91.78	89.25
80	50.86	43.48	96.16	93.90
100	56.80	49.50	97.96	96.51
120	64.02	55.29	99.03	98.17
140	68.96	59.67	99.62	98.92
160	73.81	65.24	99.77	99.28
180	78.11	69.20	99.88	99.67
200	81.25	71.57	99.96	99.86

It can also be observed from the results in Table 2 that, the power of the test increases remarkably as the effect size Δ_0 increases. All these results are presented in the line graphs in Fig 1 (left) in which the powers of the test were plotted against the various sample sizes under the two effect sizes chosen.

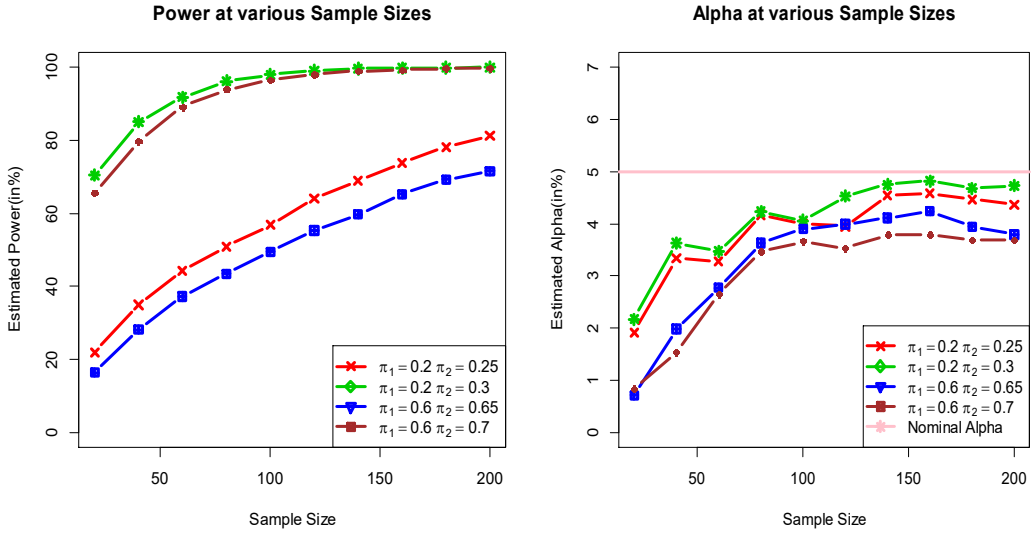

Figure 1: *The graphs of the estimated Powers (left) and nominal Type I error rate α (right) of the proposed Bayesian Non-inferiority test for two Independent Population Proportions at different sample sizes and varying effect sizes.*

Regarding the ability of the proposed test at returning the 5% nominal Type I error rate α set for it, it can be observed from Table 3 that the test returns nominal alpha values that are quite close to the 5% alpha level set for it at all the effect sizes most especially at higher sample sizes. However, the test under-estimated this nominal 5% α level of the test as all the estimated nominal values are below the 5% line as shown in Fig. 1 (right).

Table 3: *The estimated nominal Type I error rate α (in %) of the Proposed Binomial Non-inferiority Test (ZB) at various sample sizes (n), with the following test parameters; $\pi_1 = 0.2, \pi_2 = 0.25$, effect size $\Delta_0 = 0.05$ for test Z_{B1}; $\pi_1 = 0.6, \pi_2 = 0.65$ effect size $\Delta_0 = 0.05$ for test Z_{B2}; $\pi_1 = 0.2, \pi_2 = 0.3$, effect size $\Delta_0 = 0.1$ for test Z_{B3}; $\pi_1 = 0.6$, $\pi_2 = 0.7$ effect size $\Delta_0 = 0.1$ for test Z_{B4}.*

Estimated Nominal Type I Error Rate $\boldsymbol{\alpha}$ (in %) of Bayesian Non-Inferiority Tests				
Sample Size (n)	Effect Size $\Delta_0 = 0.05$		Effect Size $\Delta_0 = 0.1$	
	Z_{B1} $(\pi_1, \pi_2) = (0.2,\ 0.25)$	Z_{B2} $(\pi_1, \pi_2) = (0.6,\ 0.65)$	Z_{B3} $(\pi_1, \pi_2) = (0.2,\ 0.3)$	Z_{B4} $(\pi_1, \pi_2) = (0.6,\ 0.7)$
20	1.91	0.71	2.17	0.83
40	3.33	1.98	3.63	1.54
60	3.28	2.77	3.47	2.65
80	4.17	3.63	4.23	3.46
100	4.00	3.90	4.06	3.66
120	3.95	3.99	4.53	3.53
140	4.54	4.11	4.75	3.78
160	4.58	4.24	4.82	3.79
180	4.47	3.94	4.68	3.68
200	4.36	3.79	4.73	3.69

5. Discussion of Results

In this paper, the Bayesian test statistic for testing the difference of two binomial proportions under the non-inferiority condition was presented. The estimators of the power and significance level for the proposed Bayesian test were summarized in Table 1. Table 2 presents the empirical percentage power of the proposed test at varying sample sizes under the four scenarios considered while Table 3 presents the nominal Type I error rate of the test at various sample sizes under different effect sizes.

The Bayesian test statistic of non-inferiority Z_{B1} examines a test condition with a low effect size ($\Delta_0 = 0.05$) and relatively small sizes of the proportion pair ($\pi_1 = 0.2, \pi_2 = 0.25$). Under this scenario, the proposed Bayesian test method requires an average sample size of about 150 observations to achieve an approximate 70% power as can be observed in Table 2 and Fig 1 (left).

Unlike the test statistic Z_{B1}, the Bayesian test statistic Z_{B2} examines a test condition with a low effect size ($\Delta_0 = 0.05$) but relatively large sizes of the proportion pair ($\pi_1 = 0.6, \pi_2 = 0.65$) in the hypothesis set to be tested. Here, the Bayesian test Z_{B2} requires more sample units in the neighborhood of 180 samples before it could achieve about 70% power. It can therefore be concluded that the size of the binomial proportion pair (π_1, π_2) influences the size of the power of the test. Thus, the smaller the sizes of the two population proportions π_1 and π_2 being tested, the higher the power of the test.

The Bayesian test statistic Z_{B3} unlike Z_{B1}, examines a test condition with a relatively large effect size ($\Delta_0 = 0.1$) and smaller sizes of the proportion pair ($\pi_1 = 0.2, \pi_2 = 0.3$). Here, the proposed Bayesian test method achieved a reasonable power of about 70% even at a very small sample size as low as 20 samples as can be observed from the results in Table 1 and clearly shown by the power plot in Fig 1 (left).

Finally, the Bayesian test statistic Z_{B4} in Table 2 tested a pair of relatively large binomial proportions ($\pi_1 = 0.6, \pi_2 = 0.7$) with the same effect size ($\Delta_0 = 0.1$) as used in Z_{B3} test. Given these parameters settings, the Bayesian test Z_{B4} requires more samples to attain the same feat of about

70% power achieved by Z_{B3} at a relatively smaller sample size simply due to an increase in the sizes of the two proportions π_1 and π_2. In all cases considered, the power of the proposed Bayesian test increases as the sample size increases.

To further examine the goodness of the proposed test method, its ability to retain its size α of 5% set for it was evaluated. Table 3 presents the empirical percentage significance levels α returned by the test at varying sample sizes under the four parameters combinations considered. It could be observed from Table 3 that the empirical percentage significance levels α provided by the test are closer to the 5% nominal level set for it, especially at sample sizes 100 and above. Finally, all these results in Table 3 showed that the ability of the proposed Bayesian test method to commit the Type I error is lower than the 5% nominal level set for it. At all the sample sizes considered, the test under-estimates the 5% nominal level set for it.

6. Conclusion

In this work, an efficient Bayesian test method for testing non-inferiority between two independent binomial proportions is proposed. The goodness of the proposed test method was assessed based on the power and the empirical Type I error rates provided by the test across the various sample sizes considered. Results in Table 2 clearly showed that the proposed test is quite efficient at detecting the significance of the non-inferiority parameter value when such is not negligible in all sample sizes.

Without loss of generality, the various results from power analysis and analysis of nominal Type I error rates reported by the test are a clear indication that the proposed Bayesian test method is quite efficient and good for testing and establishing non-inferiority between two binomial population proportions. Besides the high power reported by the proposed test, the empirical levels of significance $\hat{\alpha}$ estimated by the test method that was closer to the 5% nominal level set for the test in all the cases considered also confirmed the goodness of the proposed Bayesian test method. It is therefore recommended that the proposed Bayesian test method be employed whenever it is desirable to establish the existence of non-inferiority or otherwise between a pair of treatments in which the preference of users is of the essence.

References

[1] Yahya, W. B., Olaniran, O. R., Garba, M. K., Banjoko, A. W., Dauda. K. A., and Olorede, K. O. (2016). A Test Procedure for Ordered Hypothesis of Population Proportions against a Control. *Türkiye Klinikleri Journal of Biostatistics*, 8(1):1-12. doi:10.5336/biostatic.2016-50196.

[2] Chen, D. G. and Peace, K. E. (2013). *Clinical Trial Data Analysis using R*. Chapman and Hall/CRC Press, UK. https://doi.org/10.1177/0962280211425588

[3] Ng T-H. (2014). *Noninferiority Testing in Clinical Trials: Issues and Challenges*. 1st Edition, Chapman and Hall/CRC, UK.

[4] Norleans, M. X. (2019). *Statistical Methods for Clinical Trials*. 1st Edition. Chapman and Hall/CRC Press, UK.

[5] Lesaffre, E. (2008). Superiority, Equivalence and Non-Inferiority Trials. *Bulletin of the NYU Hospital for Joint Diseases*, **66**(2):150-154.

[6] Yahya, W. B. and Jolayemi, E. T. (2008). A note on the Test of Hypothesis with Ordered Alternatives. *Scientific Annals of "Alexandru Ioan Cuza "*, 54(1), 197-208.

[7] Kawasaki, Y., Zhang, F. and Miyaoka, E. (2010). Comparisons of Test Statistics for Non-inferiority Test for the Difference between Two Independent Binominal Proportions. *American Journal of Biostatistics*, **1**(1): 23-31.

[8] European Medicines Agency (EMA) (2005). Pre-authorization Evaluation of Medicines for Human Use. Canary Wharf, London.Doc. Ref. EMEA/CPMP/EWP/2158/99.

[9] Temple, R. J. (2022). FDA Guidance on Non-Inferiority Trials General Issues, 2022. URL: www.fda.gov (Accessed on 30th March 2022).

[10] D'Agostino, R. B., Massaro, J. M. and Sullivan, L. M. (2003). Non-Inferiority trials. Design concepts and issues – the encounters of academic consultants in statistics. *Statistics in Medicine*, **22**(2): 169 – 186. DOI:10.1002/sim.1425

[11] Pocock, J. S. (2003). The pros and cons of non-inferiority trials. *Blackwell Publishing Fundamental & Clinical Pharmacology*, **17**:483 – 490.

[12] Kawasaki, Y. and Miyaoka, E. (2013). A Bayesian non-inferiority test for two independent binomial proportions. *Pharmaceutical Statistics*, 12(4): 201-206. https://doi.org/10.1002/pst.1571

[13] Garrett, A. (2003). Therapeutic equivalence; Fallacies and falsification. *Statistics in Medicine*, **22**: 741 – 762.

[14] Leung, J. T., Barnes, S. L., Lo, S. T. and Leung, D. Y. (2020). Non-inferiority trials in cardiology: what clinicians need to know. *Heart*, 106**:** 99–104. doi:10.1136/heartjnl-2019-315772

[15] Ward, M. D. and Ahlquist, J. S. (2018). *Maximum Likelihood for Social Science: Strategies for Analysis*. Cambridge University Press. P. 36.

[16] Di-Caterina, C., Kosmidis, I. (2018). Location-adjusted Wald statistics for scalar parameters. *Statistics (Methodology)*, 1-21. https://arxiv.org/pdf/1710.11217.pdf

[17] Munzel, U.and Hsuschke, D. (2003). A nonparametric test for proving non-inferiority in clinical trials with ordered categorical data. *Pharma. Stat*., **2**:31 – 37. DOI: 10.1002/pst.17.

[18] R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.