

# Sample Size Calculation of Exact Tests for the Weak Causal Null Hypothesis in Randomized Trials with a Binary Outcome

Yasutaka Chiba

Clinical Research Center, Kinki University Hospital, Osaka, Japan

Email: [chibay@med.kindai.ac.jp](mailto:chibay@med.kindai.ac.jp)

**How to cite this paper:** Chiba, Y. (2016) Sample Size Calculation of Exact Tests for the Weak Causal Null Hypothesis in Randomized Trials with a Binary Outcome. *Open Journal of Statistics*, 6, 766-776. <http://dx.doi.org/10.4236/ojs.2016.65063>

**Received:** August 10, 2016

**Accepted:** October 5, 2016

**Published:** October 8, 2016

Copyright © 2016 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The main purpose in many randomized trials is to make an inference about the average causal effect of a treatment. Therefore, on a binary outcome, the null hypothesis for the hypothesis test should be that the causal risks are equal in the two groups. This null hypothesis is referred to as the weak causal null hypothesis. Nevertheless, at present, hypothesis tests applied in actual randomized trials are not for this null hypothesis; Fisher's exact test is a test for the sharp causal null hypothesis that the causal effect of treatment is the same for all subjects. In general, the rejection of the sharp causal null hypothesis does not mean that the weak causal null hypothesis is rejected. Recently, Chiba developed new exact tests for the weak causal null hypothesis: a conditional exact test, which requires that a marginal total is fixed, and an unconditional exact test, which does not require that a marginal total is fixed and depends rather on the ratio of random assignment. To apply these exact tests in actual randomized trials, it is inevitable that the sample size calculation must be performed during the study design. In this paper, we present a sample size calculation procedure for these exact tests. Given the sample size, the procedure can derive the exact test power, because it examines all the patterns that can be obtained as observed data under the alternative hypothesis without large sample theories and any assumptions.

## Keywords

Causal Inference, Conditional and Unconditional Exact Test, Potential Outcome, Two-by-Two Contingency Table

## 1. Introduction

In superiority randomized trials in which subjects are assigned to one of two treatment

groups and the outcome is binary, data can be summarized in a two-by-two contingency table. Investigators are often interested in testing the equality of the causal risks of the two groups, using a hypothesis test. A popular method for the hypothesis test is Fisher's exact test [1] [2]. However, the null hypothesis of this test is that the causal effect of treatment is the same for all subjects. This null hypothesis is referred to as the sharp causal null hypothesis [3]-[5]; rejection of this null hypothesis does not mean that the causal risks of the two groups are different, *i.e.*, the causal risk difference is not zero. Therefore, it is inevitable to examine the null hypothesis that the causal risks are equal in the two groups, which is referred to as the weak causal null hypothesis [3]-[5], to make an inference about the average causal effect of treatment. Nevertheless, few hypothesis tests for the weak causal null hypothesis have been discussed.

Recently, two exact tests for the weak causal null hypothesis were developed [6]; one is a conditional exact test, which requires that a marginal total is fixed, and the other is an unconditional exact test, which does not require that a marginal total is fixed and depends rather on the ratio of random assignment. Under simple (or equally complete) randomization, the unconditional exact test, rather than the conditional exact test, may be applied, because the number of subjects assigned to each group is not fixed under simple randomization. Conversely, under randomization with any restriction, the conditional exact test, rather than the unconditional exact test, may be applied. These exact tests have the advantages that they are not based on large sample theories and do not require any assumptions, and they can be extended to non-inferiority trials and to the construction of a confidence interval (CI) in a straightforward manner. Therefore, these exact tests can be applied as a unified approach.

To conduct statistical hypothesis testing in an actual randomized trial, the sample size need in the trial must be calculated during the study design. Moher *et al.* [7] wrote the following about the necessity of sample size calculation as follows: "For scientific and ethical reasons, the sample size for a trial needs to be planned carefully, with a balance between medical and statistical considerations. Ideally, a study should be large enough to have a high probability (power) of detecting as statistically significant a clinically important difference of a given size if such a difference exists." A randomized trial with a smaller sample size than the sample size needed may cause type I or II error and produce a scientifically unreliable result. Conversely, an excessive sample size may cause ethical problems, because researchers have to evaluate more subjects. Although some sample size calculation methods have been developed [8] [9] and compared [10]-[12], none have been applied to the weak causal null hypothesis. Therefore, in this paper, we present a procedure for calculating the sample size for the conditional and unconditional exact tests introduced by Chiba [6].

The paper is organized as follows. In Section 2, we describe the notation used throughout this paper. In Section 3, we review the unconditional and conditional exact tests. In Section 4, we present a procedure of the sample size calculation for these exact tests. The procedure is examined through a numerical example in Section 5. Finally, we discuss the results in Section 6 and state the conclusion in Section 7.

## 2. Notation

Throughout this paper, we denote  $X$  as the assigned treatment;  $X = 1$  if a subject was assigned to the treatment group, and  $X = 0$  if assigned to the control group.  $Y$  denotes the binary outcome;  $Y = 1$  if the event occurred, and  $Y = 0$  if it did not. The results are summarized in **Table 1**, where  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $n$  are the numbers of subjects.

For each subject, it is also possible to consider the potential outcomes [13] [14] that correspond to the subject's outcome had he/she been in the other trial group.  $Y_i(x)$  denotes the potential outcome for  $i$ th subject  $i$  ( $i = 1, \dots, n$ ) under  $X = x$ . Using the potential outcomes, we consider the following four types of subject to define the four principal strata:

- i) Type 11: individuals who would experience the event regardless of the assigned treatment group; *i.e.*,  $(Y_i(1), Y_i(0)) = (1, 1)$ .
- ii) Type 10: individuals who would experience the event if assigned to the treatment group but would not experience the event if assigned to the control group; *i.e.*,  $(Y_i(1), Y_i(0)) = (1, 0)$ .
- iii) Type 01: individuals who would not experience the event if assigned to the treatment group but would experience the event if assigned to the control group; *i.e.*,  $(Y_i(1), Y_i(0)) = (0, 1)$ .
- iv) Type 00: Individuals who would not experience the event regardless of the assigned treatment group; *i.e.*,  $(Y_i(1), Y_i(0)) = (0, 0)$ .

Note that all subjects belong to one of these four types.

Let  $n_{st}$  denote the number of subjects with  $(Y_i(1), Y_i(0)) = (s, t)$ , where  $s, t = 0, 1$ . The causal risk if all subjects were assigned to the treatment group ( $X = 1$ ) can be expressed as

$$\frac{1}{n} \sum_{i=1}^n Y_i(1) = \frac{n_{11} + n_{10}}{n}. \quad (1)$$

This is because only subjects with type 11 or type 10 would experience the event. Likewise, the causal risk if all subjects were assigned to the control group ( $X = 0$ ) can be expressed as

$$\frac{1}{n} \sum_{i=1}^n Y_i(0) = \frac{n_{01} + n_{00}}{n}, \quad (2)$$

because only subjects with type 11 or type 01 would experience the event. Therefore, the sample average treatment effect (the difference between (1) and (2)) can be expressed

**Table 1.** Two-by-two contingency table obtained from a randomized trial, where  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $n$  indicate the number of subjects.

Group	Event		Total
	Occurred ( $Y = 1$ )	Not occurred ( $Y = 0$ )	
Treatment ( $X = 1$ )	$a$	$b$	$a + b$
Control ( $X = 0$ )	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$n$

as

$$\frac{1}{n} \sum_{i=1}^n \{Y_i(1) - Y_i(0)\} = \frac{n_{10} - n_{01}}{n}, \quad (3)$$

and thus the null hypothesis can be expressed as

$$H_0: n_{10} = n_{01}, \quad (4)$$

which corresponds to the weak causal null hypothesis of

$$H_0: \frac{1}{n} \sum_{i=1}^n Y_i(1) = \frac{1}{n} \sum_{i=1}^n Y_i(0). \quad (5)$$

This null hypothesis will be the main interest in many clinical trials.

Here, we consider the null hypothesis of

$$H_0: n_{10} = n_{01} = 0, \quad (6)$$

which is a special case for the weak causal null hypothesis (4). The null hypothesis (6) implies that the combination of  $(Y_i(1), Y_i(0))$  is limited to  $(Y_i(1), Y_i(0)) = (1, 1)$  or  $(0, 0)$ , and thus subjects with  $(Y_i(1), Y_i(0)) = (1, 0)$  or  $(0, 1)$  do not exist. Therefore, this null hypothesis corresponds to the following sharp causal null hypothesis:

$$H_0: Y_i(1) = Y_i(0) \text{ for all subjects}, \quad (7)$$

which is also a special case for the weak causal null hypothesis (5). It is obvious that the weak causal null hypothesis holds whenever the sharp causal null hypothesis holds. However, in general, the rejection of the sharp causal null hypothesis does not mean that the weak causal null hypothesis is rejected (*i.e.*,  $\sum_i \{Y_i(1) - Y_i(0)\} / n \neq 0$ ) [6].

### 3. Exact Tests for the Weak Causal Null Hypothesis

#### 3.1. Unconditional and Conditional Exact Tests

When the random assignment is conducted by the ratio of 1: $r$ , we assume that subjects are assigned as in **Table 2** under the weak causal null hypothesis; *i.e.*, of the  $n_{st}$  subjects,  $n_{st,1}$  subjects are assigned to the treatment group ( $X = 1$ ) with the probability of  $1/(1+r)$ , and  $n_{st,0}$  subjects are assigned to the control group ( $X = 0$ ) with the probability of  $r/(1+r)$ . Then, as each subject is independently assigned, the probability that  $n_{st,1}$  of  $n_{st}$  subjects are assigned to the treatment group can be expressed as follows:

$$\prod_{s=0}^1 \prod_{t=0}^1 \binom{n_{st}}{n_{st,1}} \left( \frac{1}{1+r} \right)^{n_{st,1}} \left( \frac{r}{1+r} \right)^{n_{st,0}}, \quad (8)$$

where  $\binom{j}{k} = {}_jC_k = \frac{j!}{k!(j-k)!}$ , and the following set of conditions is required:

**Set of conditions 1:**

$$n_{10} = n_{01}, \quad \sum_{s=0}^1 \sum_{t=0}^1 n_{st} = n, \quad \begin{cases} n_{11} \leq a+c \\ n_{10} \leq a+d \\ n_{01} \leq b+c \\ n_{00} \leq b+d \end{cases} \text{ and } \begin{cases} n_{11} + n_{10} \leq n-b \\ n_{11} + n_{01} \leq n-d \\ n_{00} + n_{10} \leq n-c \\ n_{00} + n_{01} \leq n-a \end{cases}. \quad (9)$$

**Table 2.** Two-by-two contingency table with the numbers for the four types of subjects defining the four principal strata.

Group	Event		
	Occurred ( $Y=1$ )	Not occurred ( $Y=0$ )	Total
Treatment ( $X=1$ )	$n_{11,1} + n_{10,1}$	$n_{00,1} + n_{01,1}$	$n_{11,1} + n_{10,1} + n_{01,1} + n_{00,1}$
Control ( $X=0$ )	$n_{11,0} + n_{01,0}$	$n_{00,0} + n_{10,0}$	$n_{11,0} + n_{10,0} + n_{01,0} + n_{00,0}$
Total	$n_{11} + n_{10,1} + n_{01,0}$	$n_{00} + n_{01,1} + n_{10,0}$	$n$

The first condition is the null hypothesis (4), and the second is the total number of subjects. The last two conditions are derived on the basis of **Table 1** and **Table 2**; e.g.,  $n_{11} \leq a + c$  is derived from  $n_{11} + n_{10,1} + n_{01,0} = a + c$ , and  $n_{11} + n_{10} \leq a + c + d$  is derived from  $n_{11} + n_{10} + n_{01,0} + n_{00,1} = a + c + d$ .

The risk difference estimated from the observed data,  $RD_O$ , is

$$RD_O = \frac{a}{a+b} - \frac{c}{c+d} \quad (10)$$

from **Table 1**, and the risk difference under the null hypothesis,  $RD_N$ , is

$$RD_N = \frac{n_{11,1} + n_{10,1}}{\sum_s \sum_t n_{st,1}} - \frac{n_{11,0} + n_{01,0}}{\sum_s \sum_t n_{st,0}} \quad (11)$$

from **Table 2**. In this paper, we consider only the case of  $RD_O \leq 0$ , but the following methods can easily be applied to the case of  $RD_O \geq 0$ . For  $RD_O \leq 0$ , the unconditional exact test yields the one-sided p-value,  $p$ , using the following formula:

$$p = \sup \left\{ p^U : (n_{11}, n_{10}, n_{01}, n_{00}) \text{ satisfying set of conditions 1(9)} \right\} \quad (12)$$

with

$$p^U = \sum_{n_{11,1}=0}^{n_{11}} \sum_{n_{10,1}=0}^{n_{10}} \sum_{n_{01,1}=0}^{n_{01}} \sum_{n_{00,1}=0}^{n_{00}} I(z) \prod_{s=0}^1 \prod_{t=0}^1 \binom{n_{st}}{n_{st,1}} \left( \frac{1}{1+r} \right)^{n_{st,1}} \left( \frac{r}{1+r} \right)^{n_{st,0}}, \quad (13)$$

where  $I(z) = 1$  if  $z \leq 0$  and  $I(z) = 0$  if  $z > 0$  with  $z = RD_N - RD_O$  (the difference between (11) and (10)). This is the unconditional exact test introduced by Chiba [6].

For the conditional exact test, the numbers of subjects assigned to the two groups are fixed. Thus, instead of the probability (8), the following probability is used:

$$\prod_{s=0}^1 \prod_{t=0}^1 \binom{n_{st}}{n_{st,1}} \bigg/ \binom{n}{a+b}, \quad (14)$$

where the following conditions are required:

**Set of conditions 2:** Set of conditions 1 (9) plus  $\sum_{s=0}^1 \sum_{t=0}^1 n_{st,1} = a + b$ . (15)

Consequently, the conditional exact test yields the one-sided p-value,  $p$ , using the following formula:

$$p = \sup \left\{ p^C : (n_{11}, n_{10}, n_{01}, n_{00}) \text{ satisfying set of conditions 2(15)} \right\} \quad (16)$$

with

$$p^C = \sum_{n_{11,1}=0}^{n_{11}} \sum_{n_{10,1}=0}^{n_{10}} \sum_{n_{01,1}=0}^{n_{01}} \sum_{n_{00,1}=0}^{n_{00}} I(z) \prod_{s=0}^1 \prod_{t=0}^1 \binom{n_{st}}{n_{st,1}} / \binom{n}{a+b}. \quad (17)$$

We note that, under the following monotonicity assumption [15] [16]:

$$\textbf{Assumption 1 (monotonicity): } Y_i(1) \leq Y_i(0) \text{ for all subjects,} \quad (18)$$

the weak causal null hypothesis (4) is equivalent to the sharp causal null hypothesis (6). This is because there is no subject with type 10, *i.e.*,  $n_{10} = 0$ , under this assumption. We further note that the conditional exact test degenerates to Fisher's exact test under the monotonicity assumption (18) [6].

In this paper, we define a two-sided p-value as twice the one-sided p-value.

### 3.2. Extension to Non-Inferiority Trials

Hypothesis tests of non-inferiority focus on the null hypothesis of  $H_0$ :

$\sum_i \{Y_i(1) - Y_i(0)\} / n = \delta$  rather than  $H_0: \sum_i \{Y_i(1) - Y_i(0)\} / n = 0$ , where  $\delta (> 0)$  is a small quantity specified in advance. Therefore, from the formula (3), the null hypothesis for non-inferiority can be expressed as  $n_{10} - n_{01} = \delta n$ . To take the case in which  $\delta n$  is not an integer value into account, we set the null hypothesis to a maximum integer value satisfying  $n_{10} - n_{01} \leq \delta n$ . Consequently, for non-inferiority trials, the one-sided p-value can be calculated by substituting  $n_{10} = n_{01}$  in the set of conditions 1 (9) by  $n_{10} - n_{01} = m$ , where  $m$  is a maximum integer value satisfying  $m \leq \delta n$ .

We note that we can also yield the  $100\alpha$  (%) CI, which is a CI corresponding to a significance level of  $\alpha$  (two-sided), by finding the range in which the null value of  $n_{10} - n_{01}$  is not rejected at a significance level of  $\alpha/2$  based on the two separate one-sided hypothesis tests. Chiba [17] demonstrated that such an exact CI was narrower than that using the other approaches [18] [19] to derive as an exact CI for a data set.

### 4. Sample Size Calculation

In the situation in which a randomized clinical trial with the assignment ratio 1: $r$  is planned, we set the sample size in the treatment group to  $N_1 = N/(1+r)$  and that in the control group to  $N_0 = rN/(1+r)$ , where the total number is  $N = N_1 + N_0$ . Furthermore, we set the response probabilities under the alternative hypothesis as follows:  $P_1$  if all subjects are assigned to the treatment group, and  $P_0$  if all subjects are assigned to the control group.

First, we derive the power function for a given sample size  $N$  for the unconditional exact test. When the one-sided p-value is set to  $\alpha/2$ , the power function can be derived by the following procedure:

- 1) Derive combinations of  $(n_{11}, n_{10}, n_{01}, n_{00})$  under the alternative hypothesis, which satisfy  $n_{10} - n_{01} = M_A$ , where  $M_A$  is a maximum integer value satisfying  $M_A \leq N \times (P_1 - P_0)$ ,  $NP_1 - 1 < n_{11} + n_{10} < NP_1 + 1$ , and  $NP_0 - 1 < n_{11} + n_{01} < NP_0 + 1$ .
- 2) For each combination of  $(n_{11}, n_{10}, n_{01}, n_{00})$  in Step 1, derive all combinations of  $(a, b, c, d)$ , which can be obtained as observed data under the combination of  $(n_{11}, n_{10}, n_{01}, n_{00})$ , from  $a = n_{11,1} + n_{10,1}$ ,  $b = n_{01,1} + n_{00,1}$ ,  $c = n_{11,0} + n_{01,0}$ ,  $d = n_{10,0} + n_{00,0}$  (see **Table 1**

- and **Table 2**), and calculate the probability (8).
- 3) For each combination of  $(n_{11}, n_{10}, n_{01}, n_{00})$ ,  $(a, b, c, d)$  and the probability (8) in Step 2, using the other combination  $(n_{11}^*, n_{10}^*, n_{01}^*, n_{00}^*)$  corresponding to  $(a, b, c, d)$ , derive the one-sided p-value from the unconditional exact test for the null hypothesis of  $H_0: n_{10}^* - n_{01}^* = M_N$ , where  $M_N$  is a maximum integer value satisfying  $M_N \leq \delta N$  ( $\delta = 0$  for the superiority trial and  $\delta > 0$  for the non-inferiority trial).
  - 4) Derive the conditional power given  $(n_{11}, n_{10}, n_{01}, n_{00}), p^*$ , by summing the probability (8) in Step 2 for cases in which the one-sided p-value in Step 3 is smaller than  $\alpha/2$ .
  - 5) Derive the power from  $\inf\{p^*: (n_{11}, n_{10}, n_{01}, n_{00})\}$ .

In Step 5, we take the infimum. This is because we cannot know which combination of  $(n_{11}, n_{10}, n_{01}, n_{00})$  is the most plausible from the assumed true values of  $P_1$  and  $P_0$ . Nevertheless, if investigators have plausible information about  $(n_{11}, n_{10}, n_{01}, n_{00})$ , such as the monotonicity assumption (18), the power can be calculated using the information. The hypothesis test can also be performed using the information. The required sample size can be determined by the smallest integer value of  $N$ , such that the power derived in Step 5 is larger than or equal to the power given in advance.

For the conditional exact test, the power function is obtained by adding the condition of  $\sum_s \sum_t n_{st,1} = N_1$  and applying the probability (14) instead of the probability (8) in Step 2 and by using the conditional exact test in Step 3.

The procedure to calculate the power presented here examines all of the patterns that can be obtained as observed data under the alternative hypothesis by applying an exact test without large sample theories and any assumptions. Therefore, the calculated power is exact. However, the procedure requires large computer memory in addition to significant computing time, especially for the unconditional exact test. Unfortunately, it is very difficult to perform the procedure without any assumptions in actual clinical trials. In the next section, we will illustrate the procedure under the monotonicity assumption (18) using an example.

## 5. Numerical Example

For the illustration, we have used the data from a superiority randomized clinical trial to evaluate the effects of subcutaneous drainage during digestive surgery [20]. In this trial, patients who underwent an elective primary resection of colorectal cancer were randomized into either a group that would receive subcutaneous passive drainage (PD) or a group with no drainage (ND). The randomization was performed by the minimization method [21], and the assignment ratio was 1:1. The endpoint was the incidence of superficial surgical site infections (SSI), and the result is summarized in **Table 3**. The risk difference is  $-0.066$ . Under the monotonicity assumption (18), the conditional exact test yields the one-sided p-value of 0.031 and 95% CI of  $(-32/246, 0/246) = (-0.130, 0.000)$ . The unconditional exact test yields the one-sided p-value of 0.018 and 95% CI of  $(-33/246, -1/246) = (-0.134, -0.004)$ .

For the sample size calculation, it was assumed that the true SSI incidence proportion would be 0.10 in the ND group and 0.02 in the PD group with a significance level of 0.05 (two-sided) and a power of 0.80 [20]. Under this setting and the monotonicity

assumption (18), we examined the sample size calculation presented in Section 4 for both conditional and unconditional exact tests, and the arc sine approximation with continuity-correction [8] as a reference. These three power functions are displayed in **Figure 1**. The required sample size per group with a power  $\geq 0.80$  is 132 for the unconditional exact test, 144 for the conditional exact test, and 149 for the arc sine approximation with continuity-correction.

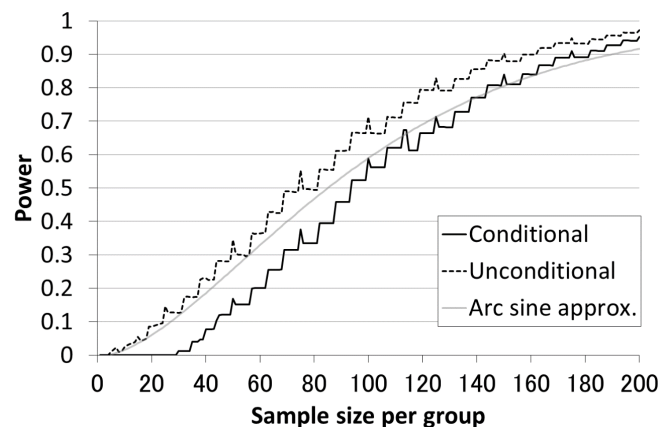
**Figure 1** showed that the power for the unconditional exact test was the highest. The power for the conditional exact test was close to that for the arc sine approximation with continuity-correction for moderate to high powers, especially near a power of 0.80. Note that without the monotonicity assumption (18), the powers for the two exact tests would be lower than those in **Figure 1**. This is because we need to consider the other combinations of  $(n_{11}, n_{10}, n_{01}, n_{00})$  with  $n_{10} \neq 0$ , and these combinations may derive the lower power.

## 6. Discussion

In this paper, we proposed a sample size calculation method for the exact tests introduced by Chiba [6], which are tests for the weak causal null hypothesis. The method can derive the exact power, because it examines all of the patterns that can be obtained as observed data under the alternative hypothesis by applying an exact test without large sample theories and any assumptions. However, unfortunately, it is very difficult

**Table 3.** Results from a superiority randomized clinical trial to evaluate the effects of subcutaneous drainage during digestive surgery.

Group	Superficial surgical site infections (SSI)		
	Yes	No	Total
Passive drainage (PD)	4	120	124
No drainage (ND)	12	110	122



**Figure 1.** Power functions under  $P_1 = 0.02$ ,  $P_0 = 0.10$ ,  $\alpha/2 = 0.025$ , and  $\delta = 0$  for the assignment ratio of 1:1: the black solid line indicates the conditional exact test, the black broken line indicates the unconditional exact test, and the gray solid line indicates the arc sine approximation with continuity-correction.



to perform the presented methods in actual clinical trials without any assumptions such as the monotonicity assumption (18), due to limitations in computing power. Further work is needed to create an efficient algorithm and to develop an approximation method.

At present, for small to moderate sample sizes, randomization with any restriction is recommended rather than simple randomization to balance some background factors between two groups. This is natural if the same hypothesis testing method is applied under either randomization method. However, if a different hypothesis testing method corresponding to the randomization method is applied, the recommendation may be changed. In Section 5, the illustration showed that the power for the unconditional exact test was higher than that for the conditional exact test. This result could be predicted, because the p-value was smaller for the unconditional exact test compared with the conditional exact test. In general, the p-value will be smaller for the unconditional exact test than for the conditional exact test, because the conditional exact test is more discrete compared with the unconditional exact test by conditioning on  $\sum_s \sum_t n_{st,1} = a + b$ . In other words, in general, the power will be higher for the unconditional exact test compared with the conditional exact test. Consequently, if the unconditional exact test is applied under simple randomization and the conditional exact test is applied under randomization with any restriction, the test power will be higher under simple randomization than under randomization with any restriction. Although lack of balance of some background factors between the two groups is a problem for small to moderate sample sizes under simple randomization, this problem may be removed by stratified (adjusted) analysis, in which the covariates to be included in the analysis will be pre-specified in the protocol [22]. Such an analysis can increase the efficiency and power of a study without introducing a risk of bias [23]-[25].

The unconditional exact test also has an advantage in that, for the sample size calculation, it takes into account cases in which the actual ratio of the numbers assigned to the two groups is not just 1:*r*, whereas the conditional exact test assumes that the ratio is just 1:*r*.

## 7. Conclusion

Whenever we conduct a statistical hypothesis test for the weak causal null hypothesis, which is the main interest in many clinical trials, we need to apply the corresponding sample size calculation method. Of the hypothesis tests, the unconditional test may have greater test power compared with the conditional test. The unconditional test and corresponding sample size calculation method should be discussed further.

## Acknowledgements

The author thanks the reviewers for helpful comments. The author also thanks Dr. Masataka Taguri for introducing the paper [20]. This work was supported partially by Grant-in-Aid for Scientific Research (No. 15K00057) from Japan Society for the Promotion of Science.

## References

- [1] Fisher, R.A. (1926) The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture of Great Britain*, **33**, 503-513.
- [2] Fisher, R.A. (1966) The Design of Experiments. 8th Edition, Oliver and Boyd, Edinburgh.
- [3] Copas, J.B. (1973) Randomization Models for the Matched and Unmatched 2x2 Tables. *Biometrika*, **60**, 467-476. <http://dx.doi.org/10.2307/2334995>
- [4] Robins, J.M. (1988) Confidence Intervals for Causal Parameters. *Statistics in Medicine*, **7**, 773-785. <http://dx.doi.org/10.1002/sim.4780070707>
- [5] Greenland, S. (1992) On the Logical Justification of Conditional Tests for Two-by-Two Contingency Tables. *American Statistician*, **45**, 248-251.
- [6] Chiba, Y. (2015) Exact Tests for the Weak Causal Null Hypothesis on a Binary Outcome in Randomized Trials. *Journal of Biometrics and Biostatistics*, **6**, 244. <http://dx.doi.org/10.4172/2155-6180.1000244>
- [7] Moher, D., Hopewell, S., Schlz, K.F., Montori, V., Gøtzsche, P., Devereaux, P.J., Elbourne, D., Egger, M. and Altman, D.G. (2010) CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials. *Journal of Clinical Epidemiology*, **63**, e1-e37. <http://dx.doi.org/10.1016/j.jclinepi.2010.03.004>
- [8] Walters, D.E. (1979) In Defense of the Arc Sine Approximation. *The Statistician*, **28**, 219-222. <http://dx.doi.org/10.2307/2987871>
- [9] Fleiss, J.L., Tytun, A. and Ury, H.K. (1980) A Simple Approximation for Calculating Sample Size for Comparing Independent Proportions. *Biometrics*, **36**, 343-346. <http://dx.doi.org/10.2307/2529990>
- [10] Ury, H.K. (1981) Continuity-Corrected Approximations to Sample Size or Power When Comparing Two Proportions: Chi-Squared or Arc Sine? *The Statistician*, **30**, 199-203. <http://dx.doi.org/10.2307/2988050>
- [11] Dobson, A.J. and Gebski, V.J. (1986) Sample Size for Comparing Two Independent Proportions Using the Continuity-Corrected Arc Sine Transformation. *The Statistician*, **35**, 51-53. <http://dx.doi.org/10.2307/2988298>
- [12] Vorburger, M. and Munoz, B. (2009) Approximations to Power When Comparing Two Small Independent Proportions. *Journal of Modern Applied Statistical Methods*, **8**, 17.
- [13] Rubin, D.B. (1978) Bayesian Inference for Causal Effects: The Role of Randomization. *Annals of Statistics*, **6**, 34-58. <http://dx.doi.org/10.1214/aos/1176344064>
- [14] Rubin, D.B. (1990) Formal Models of Statistical Inference for Causal Effects. *Journal of Statistical Planning and Inference*, **25**, 279-292. [http://dx.doi.org/10.1016/0378-3758\(90\)90077-8](http://dx.doi.org/10.1016/0378-3758(90)90077-8)
- [15] Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996) Identification of Causal Effects Using Instrumental Variables (with Discussion). *Journal of the American Statistical Association*, **91**, 444-472. <http://dx.doi.org/10.1080/01621459.1996.10476902>
- [16] Manski, C.F. (1997) Monotone Treatment Response. *Econometrica*, **65**, 1311-1334. <http://dx.doi.org/10.2307/2171738>
- [17] Chiba, Y. (2016) A Note on Exact Confidence Interval for Causal Effects on a Binary Outcome in Randomized Trials. *Statistics in Medicine*, **35**, 1739-1741. <http://dx.doi.org/10.1002/sim.6826>
- [18] Rigdon, J. and Hudgens, M. (2015) Randomization Inference for Treatment Effects on a Binary Outcome. *Statistics in Medicine*, **34**, 924-935. <http://dx.doi.org/10.1002/sim.6384>

- [19] Li, X. and Ding, P. (2016) Exact Confidence Intervals for the Average Causal Effect on a Binary Outcome. *Statistics in Medicine*, **35**, 957-960. <http://dx.doi.org/10.1002/sim.6764>
- [20] Numata, M., Godai, T., Shirai, J., Watanabe, K., Inagaki, D., Hasegawa, S., Sato, T., Oshima, T., Fujii, S., Kunisaki, C., Yukawa, N., Rino, Y., Taguri, M., Morita, S. and Masuda, M. (2014) A Prospective Randomized Controlled Trial of Subcutaneous Passive Drainage for the Prevention of Superficial Surgical Site Infections in Open and Laparoscopic Colorectal Surgery. *International Journal of Colorectal Disease*, **29**, 353-358. <http://dx.doi.org/10.1007/s00384-013-1810-x>
- [21] Pocock, S.J. and Simon, R. (1975) Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial. *Biometrics*, **31**, 103-115. <http://dx.doi.org/10.2307/2529712>
- [22] European Medicines Agency (2015) Guideline on Adjustment for Baseline Covariates in Clinical Trials. London.
- [23] Cochran, W.G. (1957) Analysis of Covariance: Its Nature and Uses. *Biometrics*, **13**, 261-281. <http://dx.doi.org/10.2307/2527916>
- [24] Cox, D.R. and McCullagh, P. (1982) Some Aspects of Analysis of Covariance. *Biometrics*, **38**, 541-561. <http://dx.doi.org/10.2307/2530040>
- [25] Tsiatis, A., Davidian, M., Zhang, M. and Lu, X. (2008) Covariate Adjustment for Two-Sample Treatment Comparisons in Randomized Clinical Trials: A Principled yet Flexible Approach. *Statistics in Medicine*, **27**, 4658-4677. <http://dx.doi.org/10.1002/sim.3113>

Scientific Research Publishing

---

**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [ojs@scirp.org](mailto:ojs@scirp.org)

