Opinion   Open Access

Causal Inference in Randomized Trials: A Shift from the Sharp Causal Null Hypothesis to the Weak Causal Null Hypothesis

Yasutaka Chiba[*]

Clinical Research Center, Kinki University Hospital, Japan

[*]**Corresponding author:** Yasutaka Chiba, Clinical Research Center, Kinki University Hospital, Japan, Tel: +81-72-366-0221; Fax: +81-72-368-1193; E-mail: chibay@med.kindai.ac.jp

Introduction

In randomized trials, statistical inference of the average causal effect (ACE) of a treatment in comparison with a control treatment is desired with a focus on a particular outcome. Often, in addition to estimation of the ACE, the confidence interval (CI) is calculated, and a hypothesis test is performed. Nevertheless, many tests employed in randomized trials that are currently underway do not allow any statistical inference to be made about the ACE unless strict assumptions are satisfied. For example, the permutation test, which corresponds to Fisher's exact test in the case of a binary outcome, is a hypothesis test for the sharp causal null hypothesis (i.e., the causal effect of treatment is the same for all subjects), but not for the weak causal null hypothesis (i.e., the causal risks are equal in the two groups). In this article, I discuss causal inference in the context of randomized trials with a binary outcome. First, I state that a hypothesis test for the sharp causal null hypothesis is generally different from that for the weak causal null hypothesis, as I showed in a recent publication [1]. Next, I demonstrate that previously proposed CIs linking to exact tests are not informative in terms of the ACE; I use hypothetical data to make this point. Finally, I discuss the future prospects for causal inference in randomized trials.

Sharp and Weak Causal Null Hypothesis

For demonstration purposes, I use hypothetical data with a small sample in Table 1, where X denotes a treatment, and Y the outcome. Let $Y(x)$ denote the potential outcome for each subject under $X = x$; let n_{st} denote the number of subjects with $(Y(1), Y(0)) = (s, t)$, where $s, t = 0, 1$; and n is the total number of subjects. Then, all subjects belong to either $(Y(1), Y(0)) = (1, 1), (1, 0), (0, 1),$ or $(0, 0)$; and $\Sigma_{s,t} n_{st} = n$. In randomized trials, it is generally sought to derive inferences about the ACE and thus to compare $\Pr(Y(1) = 1)$ and $\Pr(Y(0) = 1)$. The null hypothesis of interest is thus the following weak causal null hypothesis:

H_0: $\Pr(Y(1) = 1) = \Pr(Y(0) = 1)$.

	Outcome		
Group	Y = 1	Y = 0	Total
X = 1	1	4	5
X = 0	3	2	5

Table 1: Hypothetical data with a small sample size.

Using n_{st}, $\Pr(Y(1) = 1)$ and $\Pr(Y(0) = 1)$ can be expressed as $(n_{11} + n_{10})/n$ and $(n_{11} + n_{01})/n$, respectively, and then the weak causal null hypothesis corresponds to

H_0: $n_{10} = n_{01}$.

Although the null hypothesis of interest is the weak causal null hypothesis, the hypothesis tests that are commonly used do not explore this null hypothesis. For instance, Fisher's exact test, which is the exact test most commonly used to evaluate two-by-two contingency tables, is a hypothesis test for the following sharp causal null hypothesis [1,2]:

H_0: $Y(1) = Y(0)$ for all subjects.

Under this null hypothesis, the combination of $(Y(1), Y(0))$ is limited to $(Y(1), Y(0)) = (1, 1)$ or $(0, 0)$. Because subjects with $(Y(1), Y(0)) = (1, 0)$ or $(0, 1)$ do not exist, the sharp causal null hypothesis corresponds to

H_0: $n_{10} = n_{01} = 0$.

Clearly, the sharp causal null hypothesis is a special case of the weak causal null hypothesis, and the proposition "the weak causal null hypothesis holds if the sharp causal null hypothesis holds" is true. However, the inverse "the weak causal null hypothesis does not hold if the sharp causal null hypothesis does not hold" is not true. This can be illustrated using the hypothetical data in Table 1 as follows. Assume that the 10 subjects in Table 1 comprise $(n_{11}, n_{10}, n_{01}, n_{00}) = (4, 1, 1, 4)$. Then, Table 1 is obtained as a result that $(1, 0, 1, 3)$ of $(n_{11}, n_{10}, n_{01}, n_{00}) = (4, 1, 1, 4)$ is randomly assigned to the group $X = 1$. Because $n_{10} = n_{01} = 1$, the sharp causal null hypothesis does not hold, but the weak causal null hypothesis does hold. This shows that the sharp causal null hypothesis can be rejected even when $\Pr(Y(1) = 1) - \Pr(Y(0) = 1) = 0$. In other words, rejection of the sharp causal null hypothesis does not mean that $\Pr(Y(1) = 1) - \Pr(Y(0) = 1) \neq 0$.

Nevertheless, the inverse is true under the following monotonicity assumption [3]:

Monotonicity assumption: $Y(1) \leq Y(0)$ for all subjects.

This assumption implies that there is no subject with $(Y(1), Y(0)) = (1, 0)$; i.e., $n_{10} = 0$. Then, $n_{01} > n_{10} = 0$ corresponds to the situation that the sharp causal null hypothesis does not hold, and if this is the case, the weak causal null hypothesis also does not hold. Consequently, the sharp causal null hypothesis is equivalent to the weak causal null hypothesis under the monotonicity assumption.

Confidence Interval

Next, I discuss previously proposed CIs linking to exact tests. The limits of the CI for the ACE cannot be outside the nonparametric bounds [4], which are given by

$\Pr(Y = 1, X = x) \leq \Pr(Y(x) = 1) \leq 1 - \Pr(Y = 0, X = x)$.

This is because the nonparametric bounds are the range within which the ACE must exist. For the data in Table 1, the nonparametric bounds for the causal odds ratio (OR) are $0.028 \leq$ causal OR ≤ 3.500.

However, the usual and matching exact CIs based on the hypergeometric distribution [5], which links to Fisher's exact test, yield 95% CIs of (0.003, 4.586) and (0.005, 3.172), respectively. Note that the conditional maximum likelihood estimator yields an OR of 0.203. Both lower limits are smaller than the lower bound of 0.028. This shows that these 95% CIs include values that the causal OR cannot take. Consequently, these exact CIs for the OR are not the exact CIs for the causal OR.

On the risk difference (RD) scale, Santner-Snell CI [6], which is an exact CI linking to Barnard's exact test, yields a 95% CI of (-0.867, 0.305), while the bounds for the causal RD are $-0.700 \le$ causal RD ≤ 0.300 [7]. This also shows that the Santner-Snell CI is not the exact CI for the ACE.

Discussion

Finally, I discuss future prospects for causal inference in randomized trials. As mentioned above, inference about the ACE is generally of most interest. Nevertheless, in general, the existing hypothesis tests and the CIs linking to them are not appropriate for evaluating the ACE. Fisher's exact test, which is the exact test commonly used to evaluate two-by-two contingency tables, is a hypothesis test for the sharp causal null hypothesis, but unfortunately, in general, rejection of the sharp causal null hypothesis does not mean that $\Pr(Y(1) = 1) - \Pr(Y(0) = 1) \ne 0$.

To increase the quality of statistical inferences about ACE in randomized trials, hypothesis tests for the weak causal null hypothesis and CIs linking to them require further attention. Although a few methods have recently been developed on binary outcomes [1,8-10], no well-established method exists yet. Also, new methods of sample size calculation are required, and it is necessary to create an algorithm permitting efficient use of such newly developed methods.

Furthermore, according to trial design, the methods need improvement. Such methods must be applicable to not only superiority but also non-inferiority trials; the methods will differ in the type of randomization applied. If simple (or equally complete) randomization is employed, the method must not require that the number of subjects assigned to each group is fixed. For randomized trials with restrictions, a stratified analysis is better than a crude analysis.

	Outcome		
Group	Y = 1	Y = 0	Total
X = 1	1	69	70
X = 0	8	62	70

Table 2: Hypothetical data used in Chiba [1].

It is also important, in some settings, to consider whether an assumption made actually holds. For instance, the monotonicity assumption will be reasonable in many vaccine trials, in which there is no subject who would become infected if a vaccine was received, but would not become infected if a vaccine was not received. For the hypothetical data in Table 2, Chiba's conditional exact test [1] yields an RD of -0.100 (95% CI: -0.200, -0.014; two-sided p-value = 0.034) under the monotonicity assumption, but an RD of -0.100 (95% CI: -0.207, 0.007; two-sided p-value = 0.074) without the assumption. The latter result may give a somewhat more negative impression than the former. Thus, the conclusions drawn from the randomized trial may be in error, because the monotonicity was (or was not) assumed by mistake. When considering assumptions, cooperation between clinicians and biostatisticians is essential.

The hypothesis tests and CIs in randomized trials are not a new problem (indeed, the problem arose several decades ago). However, the ACE-related concerns described above are indeed new. Further work is needed.

Acknowledgements

References

1. Chiba Y (2015) Exact tests for the weak causal null hypothesis on a binary outcome in randomized trials. J Biomet Biostat 6: 244.

2. Greenland S (1992) On the logical justification of conditional tests for two-by-two contingency tables. Am Stat 45: 248-251.

3. Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables (with discussion). J Am Stat Assoc 91: 444-455.

4. Pearl J (1995) Causal inference from indirect experiments. Artif Intell Med 7: 561-582.

5. Fay MP (2010) Two-sided exact tests and matching confidence intervals for discrete data. R Journal 2: 53-58.

6. Santner T, Snell M (1980) Small-sample confidence intervals for p1 – p2 and p1 / p2 in 2 × 2 contingency tables. J Am Stat Assoc 75: 386-394.

7. Chiba Y (2015) A note on exact tests and confidence intervals for two-by-two contingency tables in randomized trials. Int J Clin Res Trials 1: 102.

8. Rigdon J, Hudgens MG (2015) Randomization inference for treatment effects on a binary outcome. Stat Med 34: 924-935.

9. Li X, Ding P (2016) Exact confidence intervals for the average causal effect on a binary outcome. Stat Med 35: 957-960.

10. Chiba Y (2016) A note on exact confidence interval for causal effects on a binary outcome in randomized trials. Stat Med 35: 1739-1741.