Yasutaka Chiba*

Bayesian Inference of Causal Effects for an Ordinal Outcome in Randomized Trials

Abstract: In randomized trials in which two treatment arms are compared with a binary outcome, the causal effect can be identified by assuming that the two treatment arms are exchangeable. In trials with an ordinal outcome, which is categorized as more than two, the causal effect can be identified by assuming that the potential outcomes are independent and that the two treatment arms are exchangeable. In this article, we propose a Bayesian approach to causal inference that does not rely on these two assumptions. To achieve this purpose, we use a randomization-based approach and response type. Then, the likelihood function is derived by physical randomization in which subjects who belong to a response type are randomly assigned to the treatment or control, with no modeling assumption on the outcome. Our approach can derive not only the posterior distribution of the causal effect but also that of the number of subjects in each response type. The proposed approach is illustrated with two examples from randomized clinical trials.

Keywords: Contingency Table, Potential Outcome, Response Type

1 Introduction

The main purpose of randomized trials is to draw inferences regarding causal effects. When two treatment arms are compared with binary outcomes, causal effects can be identified by assuming that the two treatment arms are exchangeable. This assumption means that the risk of the event in the treatment arm would have been the same as the risk of the event in the control arm had subjects in the treatment arm been assigned to the control arm [1, 2], and it is often assumed under random assignment. In trials with an ordinal outcome, which is categorized as more than two, we can identify causal effects by making the assumption that the potential outcomes are independent, in addition to the assumption that the two treatment arms are exchangeable. The assumption of independent potential outcomes means that the potential outcome if a subject was assigned to the treatment arm is independent of the potential outcome if the subject was assigned to the control arm [3]. In general, it is impossible to identify causal effects under the frequentist approach without making these two assumptions or other strict assumptions. Therefore, in this article, we propose a Bayesian approach to causal inference that does not rely on these assumptions, with no modeling assumption. To achieve this purpose, we apply the randomization-based approach, in which the trial subjects are viewed as a finite population of interest and probabilities arise only through the random assignment [4, 5]. Therefore, we do not require the observed data to be a random sample from an infinite population and do not need to apply the large sample approximation. We further apply the response types, which is the pair of potential outcomes for a subject under treatment and control conditions.

For the case of a binary outcome, Ding and Miratrix [6] developed a Bayesian approach, using the randomization-based approach and response types. Their approach requires that researchers fix the number of subjects who belong to one of four response types. In this article, we discuss not only a binary outcome but also an ordinal outcome with more than two categories. Furthermore, we do not fix the number of subjects who belong to one of the response types. Therefore, the approach proposed here can be regarded as an extension of Ding and Miratrix [6].

---
*Corresponding author: Yasutaka Chiba, Kinki University Hospital, Osaka, Japan, e-mail: chibay@med.kindai.ac.jp

In Section 2, we introduce notation and definitions. The Bayesian approach to causal inference is described in Section 3. In Section 4, we illustrate our approach using data from two randomized clinical trials. We conclude the article with a discussion in Section 5.

2 Notation and definitions

Throughout this article, we denote X for the assigned treatment; $X = 1$ if the subject was assigned to the treatment arm and $X = 0$ if the subject was assigned to the control arm. Y denotes the ordinal outcome with J categories labeled $0, \ldots, J - 1$, where 0 and $J - 1$ represent the worst and best categories, respectively. Furthermore, we denote the potential outcome for a subject with $X = x$ as $Y(x)$ [7, 8], which is the outcome that would occur if the subject were assigned to a specific type of treatment. It is not possible to know the values of both $Y(1)$ and $Y(0)$. If the subject is assigned to the treatment arm ($X = 1$), then we observe $Y(1)$ but not $Y(0)$. Conversely, if the subject is assigned to the control arm ($X = 0$), then $Y(1)$ is not observed and $Y(0)$ is observed; i. e., $Y = Y(1)X + Y(0)(1 - X)$.

We define the causal parameters for an ordinal outcome in terms of the response type, which is a pair of potential outcomes for a subject under treatment and control conditions; i. e., $\{Y(1), Y(0)\}$. Thus, $\Pr\{Y(1) = k, Y(0) = l\}$ ($k, l = 0, \ldots, J - 1$) denotes the proportion of subjects whose potential outcome is k under the treatment condition and l under the control condition.

Recently, some new causal parameters were developed (e. g., [9–11]). Volfovsky et al. [9] studied the following conditional median as a causal measure for an ordinal outcome:

$$M_l = \text{median}\{Y(1) \mid Y(0) = l\},$$

which yields J conditional medians. There may be some cases in which $M_l > l$ and $M_{l'} < l'$ ($l \neq l'$). In such cases, it is difficult to determine whether the treatment is superior to the control. Although this causal measure may be valuable as a local causal measure, as Lu et al. [10] noted, it is not a direct measure of the treatment effect.

Lu et al. [10] proposed the use of the following two causal parameters:

$$\tau = \Pr\{Y(1) \geq Y(0)\} \quad \text{and} \quad \eta = \Pr\{Y(1) > Y(0)\}.$$

The use of τ may be misleading because $\tau = 1$ (and $\eta = 0$) under the sharp causal null hypothesis [12] that $Y(1) = Y(0)$ for all subjects. Nevertheless, $1 - \tau = \Pr\{Y(1) < Y(0)\} = 0$ under the sharp causal null hypothesis. Then, $1 - \tau$ indicates the probability that the control is beneficial over the treatment, whereas η indicates the probability that the treatment is beneficial over the control. Owing to the symmetry of the treatment and control labels, Lu et al. [10] considered that τ and η are equally useful. Consequently, they suggested using both τ and η in practice. Surely, if $\eta = 1 - \tau = 0.5$, it is not concluded that a treatment is more beneficial than the control. Nevertheless, if $\eta = 0.2$ and $1 - \tau = 0$, it may be concluded that the treatment is more beneficial than the control.

Chiba [11] proposed the following causal parameter:

$$\theta_x = \Pr\{Y(x) \geq Y(1 - x)\} - \Pr\{Y(x) = Y(1 - x) = 0\}$$

for $x = 0, 1$, which is similar to the causal parameter proposed by Lu et al. [10], but slightly different. The causal parameter θ_x can be interpreted as the proportion of subjects for whom the treatment would not be more harmful than the control for $X = 1$, and, similarly, the proportion of subjects for whom the control would not be more harmful than the treatment for $X = 0$. For the case of a binary outcome ($J = 2$), which is the simplest ordinal outcome, θ_x is equivalent to the well-defined causal risk $\Pr\{Y(X) = 1\}$, while τ and η are not. Chiba [11] proposed comparing θ_1 and θ_0 as the causal measure, such as in the case of a binary outcome.

Again, Lu et al. [10] suggested using the two symmetric parameters to evaluate the causal effect of treatment. However, to determine whether the treatment is beneficial over the control, it is more efficient to use

one causal measure rather than using two parameters. Therefore, we consider the relative treatment effect that can be expressed as follows on the difference scale:

$$\Pr\{Y(1) > Y(0)\} - \Pr\{Y(1) < Y(0)\},\qquad(1)$$

which is equal to $\theta_1 - \theta_0$ and $\eta - (1 - \tau)$. This can be interpreted as a causal quantity to indicate how much larger the proportion of subjects for whom the treatment would be more beneficial (or not more harmful) than the control is than the proportion of subjects for whom the control would be more beneficial (or not more harmful) than the treatment. Obviously, (1) > 0 if the treatment is superior to the control, and (1) < 0 if the treatment is inferior to the control. Under the sharp causal null hypothesis, (1) = 0. By using the relative treatment effect of (1), researchers can clearly consider whether the treatment is beneficial in comparison with the control. We note that (1) can also be expressed as

$$\Pr\{Y(1) > Y(0)\} - \Pr\{Y(1) < Y(0)\} = \sum_{k=1}^{J-1}\sum_{l=0}^{k-1}[\Pr\{Y(1) = k, Y(0) = l\} - \Pr\{Y(1) = l, Y(0) = k\}].$$

For a binary outcome ($J = 2$), (1) degenerates into $\Pr\{Y(1) = 1\} - \Pr\{Y(0) = 1\}$. Therefore, under the exchangeability assumption that $\Pr\{Y(x) = j\} = \Pr\{Y(x) = j \mid X = 1\} = \Pr\{Y(x) = j \mid X = 0\}$ for $x = 0, 1$ [1, 2], (1) can be identified as $\Pr(Y = 1 \mid X = 1) - \Pr(Y = 1 \mid X = 0)$. However, for $J \geq 3$, the exchangeability assumption is not sufficient for (1) to be identified since $\Pr\{Y(1) = k, Y(0) = l\}$ cannot be identified only under this assumption. In general, to identify (1), we further must make the assumption of independent potential outcomes that $Y(1)$ is independent of $Y(0)$ [3], or other strict assumptions. Under the assumptions of exchangeability and independent potential outcomes, (1) can be identified as

$$\sum_{k=1}^{J-1}\sum_{l=0}^{k-1}\{\Pr(Y = k \mid X = 1)\Pr(Y = l \mid X = 0) - \Pr(Y = l \mid X = 1)\Pr(Y = k \mid X = 0)\}.\qquad(2)$$

This is equivalent to the relative treatment effect version of stochastic superiority [13–15].

In this article, we consider the sample causal effect version of (1). Let n denote the total number of subjects in a sample and n_{kl} denote the unobserved number of subjects with response type $\{Y(1), Y(0)\} = (k, l)$ ($k, l = 0, \ldots, J - 1$), where $\sum_{k=0}^{J-1}\sum_{l=0}^{J-1} n_{kl} = n$. The sample causal effect corresponding to (1) can then be expressed as

$$\frac{1}{n}\sum_{k=1}^{J-1}\sum_{l=0}^{k-1}(n_{kl} - n_{lk}).\qquad(3)$$

In Section 3, we present a model-free Bayesian approach to the causal inference of (3) that does not rely on the assumptions of exchangeability or independent potential outcomes. In randomized trials, the exchangeability assumption is a standard assumption, and it is often taken for granted. Nevertheless, as Hernán and Robins [2] noted, we are generally unable to determine whether $\Pr\{Y(x) = j \mid X = 1\} = \Pr\{Y(x) = j \mid X = 0\}$ holds in a sample.

Here, we assume that $n_{kl,x}$ of n_{kl} subjects are randomly assigned to the arm with $X = x$. We then construct a $2 \times J$ contingency table for the unobserved number $n_{kl,x}$, as shown in Table 1. We also construct a $2 \times J$ contingency table for the observed number, as shown in Table 2. In this $2 \times J$ contingency table for the sample, m_{xj} is the observed number of subjects in the category with $(X, Y) = (x, j)$. Finally, we define $N \equiv (n_{00}, \ldots, n_{kl}, \ldots, n_{(J-1)(J-1)})$ and $N_x \equiv (n_{00,x}, \ldots, n_{kl,x}, \ldots, n_{(J-1)(J-1),x})$. In the following section, a combination of $(n_{00}, \ldots, n_{kl}, \ldots, n_{(J-1)(J-1)})$ is noted as "an N."

Table 1: $2 \times J$ contingency table constructed from the unobserved number $n_{kl,x}$.

| Treatment | Outcome | | | | | Total |
	$Y = 0$		$Y = j$		$Y = J - 1$	
$X = 1$	$\sum_l n_{0l,1}$	...	$\sum_l n_{jl,1}$	...	$\sum_l n_{(J-1)l,1}$	$\sum_k \sum_l n_{kl,1}$
$X = 0$	$\sum_k n_{k0,0}$	...	$\sum_k n_{kj,0}$	...	$\sum_k n_{k(J-1),0}$	$\sum_k \sum_l n_{kl,0}$

Table 2: $2 \times J$ contingency table constructed from the observed number m_{xj}.

Treatment	Outcome						Total
	$Y = 0$		$Y = j$		$Y = J - 1$		
$X = 1$	m_{10}	...	m_{1j}	...	$m_{1(J-1)}$		$\sum_j m_{1j}$
$X = 0$	m_{00}	...	m_{0j}	...	$m_{0(J-1)}$		$\sum_j m_{0j}$

3 Bayesian inference of causal effects

In Section 3.1, we present the region of N, in which the likelihood function is nonzero. In Section 3.2, we present a Bayesian approach to make inferences about (3), using the region given in Section 3.1.

3.1 Region of N

The $2 \times J$ contingency table for the unobserved number $n_{kl,x}$ (Table 1) corresponds to the table for the observed number m_{xj} (Table 2). Therefore, as shown in Chiba [11], n_{kl} must satisfy the following inequality:

$$\sum_{a=1}^{j'} \sum_{b=1}^{j^*-j'} n_{k_a l_b} \le \sum_{a=1}^{j'} m_{1k_a} + \sum_{b=1}^{j^*-j'} m_{0l_b}, \tag{4}$$

where $2 \le j^* \le 2J - 1$ and $1 \le j' \le \min\{j^* - 1, J\}$. k_a is an arbitrary j in the j' categories with $X = 1$, and the tuple $(k_1, \ldots, k_{j'})$ takes on values in $\{0, \ldots, J - 1\}$ with $k_1 \ne \cdots \ne k_{j'}$. Similarly, l_b is an arbitrary j in the $(j^* - j')$ categories with $X = 0$, and the tuple $(l_1, \ldots, l_{j^*-j'})$ takes on values in $\{0, \ldots, J - 1\}$ with $l_1 \ne \cdots \ne l_{j^*-j'}$. Inequality (4) has to hold for all possible choices of $(k_1, \ldots, k_{j'})$ and $(l_1, \ldots, l_{j^*-j'})$. The right hand side of (4) is the sum of m_{xj} for j^* categories, with j' categories for $X = 1$ and $(j^* - j')$ categories for $X = 0$. The left hand side is the sum of $n_{k_a l_b} = n_{k_a l_b,1} + n_{k_a l_b,0}$ made in these j^* categories. For example, for two categories ($j^* = 2$) with one category for $X = 1$ ($j' = 1$) and one category for $X = 0$ ($j^* - j' = 1$),

$$m_{1k_1} + m_{0l_1} = \sum_{b=1}^{J} n_{k_1 l_b,1} + \sum_{a=1}^{J} n_{k_a l_1,0} = n_{k_1 l_1} + \sum_{b=2}^{J} n_{k_1 l_b,1} + \sum_{a=2}^{J} n_{k_a l_1,0} \ge n_{k_1 l_1},$$

where $0 \le k_1, l_1 \le J - 1$. This inequality implies

$$\binom{J}{1} \times \binom{J}{1} = J^2$$

inequalities with $n_{00} \le m_{10} + m_{00}, \ldots, n_{kl} \le m_{1k} + m_{0l}, \ldots, n_{(J-1)(J-1)} \le m_{1(J-1)} + m_{0(J-1)}$. Similarly, for three categories ($j^* = 3$) with two categories for $X = 1$ ($j' = 2$) and one category for $X = 0$ ($j^* - j' = 1$),

$$m_{1k_1} + m_{1k_2} + m_{0l_1} = \sum_{b=1}^{J} n_{k_1 l_b,1} + \sum_{b=1}^{J} n_{k_2 l_b,1} + \sum_{a=1}^{J} n_{k_a l_1,0} \ge n_{k_1 l_1} + n_{k_2 l_1},$$

which implies

$$\binom{J}{2} \times \binom{J}{1} = \frac{J^2(J-1)}{2}$$

inequalities. As the above inequalities are derived, inequalities that n_{kl} must satisfy are derived for all j^* satisfying $2 \le j^* \le 2J - 1$ and all j' satisfying $1 \le j' \le \min\{j^* - 1, J\}$. Equation (4) expresses the inequalities by one formula.

In the case of a binary outcome ($J = 2$), (4) has to hold for $(j^*, j') = (2, 1), (3, 1)$, and $(3, 2)$, because $2 \le j^* \le 3$ and $1 \le j' \le \min\{j^* - 1, 2\}$. Then, (4) implies the following three inequalities:

- $n_{k_1 l_1} \leq m_{1 k_1} + m_{0 l_1}$ for $(j^*, j') = (2, 1)$,
- $n_{k_1 l_1} + n_{k_1 l_2} \leq m_{1 k_1} + m_{0 l_1} + m_{0 l_2}$ $(l_1 \neq l_2)$ for $(j^*, j') = (3, 1)$,
- $n_{k_1 l_1} + n_{k_2 l_1} \leq m_{1 k_1} + m_{1 k_2} + m_{0 l_1}$ $(k_1 \neq k_2)$ for $(j^*, j') = (3, 2)$,

where $k_1, k_2, l_1, l_2 = 0, 1$; thus,

$$
\begin{cases} n_{00} \leq m_{10} + m_{00} \\ n_{01} \leq m_{10} + m_{01} \\ n_{10} \leq m_{11} + m_{00} \\ n_{11} \leq m_{11} + m_{01} \end{cases} \text{ and } \begin{cases} n_{00} + n_{01} \leq m_{10} + m_{00} + m_{01} \\ n_{10} + n_{11} \leq m_{11} + m_{00} + m_{01} \\ n_{00} + n_{10} \leq m_{10} + m_{11} + m_{00} \\ n_{01} + n_{11} \leq m_{10} + m_{11} + m_{01} \end{cases} . \tag{5}
$$

Consequently, $(n_{00}, n_{01}, n_{10}, n_{11})$ must exist in the region satisfying these eight inequalities and $n_{00} + n_{01} + n_{10} + n_{11} = n$. Using this region of $(n_{00}, n_{01}, n_{10}, n_{11})$, we can derive

$$
-\frac{m_{10} + m_{01}}{n} \leq \frac{n_{10} - n_{01}}{n} \leq \frac{m_{11} + m_{00}}{n},
$$

which are the sharp nonparametric bounds of the causal risk difference, stated in previous papers [16, 17]. In general, for a $2 \times J$ contingency table, the region of N is determined as $N \in \mathcal{F}$ with

$$
\mathcal{F} = \left\{ N : \sum_{k=0}^{J-1} \sum_{l=0}^{J-1} n_{kl} = n \text{ and } (4) \right\} .
$$

Even in the case of $J \geq 3$, we can derive the sharp nonparametric bounds of (3) by examining all Ns in this region numerically [11].

3.2 Proposed Bayesian approach

We assume that the number of subjects assigned to each arm, $\sum_{k=0}^{J-1} \sum_{l=0}^{J-1} n_{kl,x}$, is fixed to the actual assigned number $m_{x+} = \sum_{j=0}^{J-1} m_{xj}$. Then, for an N in the region $N \in \mathcal{F}$, the probability that $n_{kl,1}$ of the n_{kl} subjects are randomly assigned to the treatment arm $(X = 1)$ can be expressed as

$$
\prod_{k=0}^{J-1} \prod_{l=0}^{J-1} \binom{n_{kl}}{n_{kl,1}} \bigg/ \binom{n}{m_{1+}} .
$$

This can be regarded as a natural extension of the hypergeometric distribution to the response type version. Unfortunately, as we cannot know the value of $n_{kl,1}$, even if N is fixed to a set of the values of $(n_{00}, \ldots, n_{kl}, \ldots, n_{(J-1)(J-1)})$ in the region $N \in \mathcal{F}$, we cannot calculate this probability from the observed data. However, if we limit N_1 to the region $N_1 \in \mathcal{F}_1$ with

$$
\mathcal{F}_1 = \left\{ N_1 : \sum_{l=0}^{J-1} n_{kl,1} = m_{1k} \quad \text{and} \quad \sum_{k=0}^{J-1} n_{kl,1} = \sum_{k=0}^{J-1} n_{kl} - m_{0l} \right\}, \tag{6}
$$

where the set $\mathcal{F}_1$ is conditional on N, then we can make Bayesian inferences about (3). This region is derived because each category in Table 2 corresponds to that in Table 1. The latter equation is derived from $m_{0l} = \sum_{k=0}^{J-1} n_{kl,0} = \sum_{k=0}^{J-1} (n_{kl} - n_{kl,1})$.

Using this region of N_1, we can express the likelihood function for N_1, $f(N_1 \mid N)$, as

$$
f(N_1 \mid N) = \sum_{N_1 \in \mathcal{F}_1} \prod_{k=0}^{J-1} \prod_{l=0}^{J-1} \binom{n_{kl}}{n_{kl,1}} \bigg/ \binom{n}{m_{1+}} \tag{7}
$$

in the region $N \in \mathcal{F}$, and $f(N_1 \mid N) = 0$ outside $N \in \mathcal{F}$. After the prior probability $\Pr(N)$ is determined, the posterior probability $\Pr(N \mid N_1)$ is calculated from

$$\Pr(N \mid N_1) = \frac{f(N_1 \mid N)\Pr(N)}{\sum_{N \in \mathcal{F}} f(N_1 \mid N)\Pr(N)}.$$

The posterior distribution of (3) can be derived by summing the posterior probability $\Pr(N \mid N_1)$ for all Ns that equal a value in (3). For example, the posterior probability of (3) = 0 is derived by summing $\Pr(N \mid N_1)$ for all N with a combination of n_{kl} and n_{lk} equaling $\sum_{k=1}^{J-1} \sum_{l=0}^{k-1} (n_{kl} - n_{lk})/n_{0+} = 0$. Similarly, we can derive the posterior distribution of n_{kl}, which is the number of subjects who belong to the response type with $\{Y(1),\ Y(0)\} = (k,\ l)$. It is important to note that the likelihood is completely derived from the physical randomization with no modeling assumption on the outcome, and we do not require the assumptions of exchangeability and independent potential outcomes.

To complement the above explanation of the proposed approach, let us consider the simple hypothetical example of a 2×2 contingency table with $(m_{00},\ m_{01}, m_{10}, m_{11}) = (2,\ 2, 1, 3)$. For this, we have 60 combinations of $(n_{00},\ n_{01}, n_{10}, n_{11})$ that satisfy $n_{00} + n_{01} + n_{10} + n_{11} = 8$ and (4) (i. e., eight inequalities in (5)). For example, one of the 60 combinations is $(n_{00},\ n_{01}, n_{10}, n_{11}) = (2,\ 0, 2, 4)$, which yields the causal risk difference of $(n_{10} - n_{01})/n = (2 - 0)/8 = 0.25$. In other words, the 60 combinations of $(n_{00},\ n_{01}, n_{10}, n_{11})$ including $(2, 0, 2, 4)$ are in the region $N \equiv (n_{00},\ n_{01}, n_{10}, n_{11}) \in \mathcal{F}$, and the other combinations are outside the region $N \in \mathcal{F}$. For each combination in the region $N \in \mathcal{F}$, we search combinations of $(n_{00,1},\ n_{01,1}, n_{10,1}, n_{11,1})$ satisfying the two equations in (6). For $(n_{00},\ n_{01}, n_{10}, n_{11}) = (2,\ 0, 2, 4)$, we have one combination of $(n_{00,1},\ n_{01,1}, n_{10,1}, n_{11,1}) = (1,\ 0, 1, 2)$. This implies that only $(n_{00,1},\ n_{01,1}, n_{10,1}, n_{11,1}) = (1,\ 0, 1, 2)$ is in the region $N_1 \equiv (n_{00,1},\ n_{01,1}, n_{10,1}, n_{11,1}) \in \mathcal{F}_1$ for $(n_{00},\ n_{01}, n_{10}, n_{11}) = (2, 0, 2, 4)$. Then, we calculate the likelihood $f(N_1 \mid N)$ for $(n_{00},\ n_{01}, n_{10}, n_{11}) = (2, 0, 2, 4)$ and $(n_{00,1},\ n_{01,1}, n_{10,1}, n_{11,1}) = (1, 0, 1, 2)$ from (7). Similarly, the likelihood is calculated for all combinations of $(n_{00},\ n_{01}, n_{10}, n_{11})$, where the likelihood is zero outside $N \in \mathcal{F}$. After the prior probability $\Pr(N)$ is determined, the posterior probability $\Pr(N \mid N_1)$ is calculated for 60 combinations in the region $N \in \mathcal{F}$. When we assume the non-informative prior distribution that the probability for a combination of $(n_{00},\ n_{01}, n_{10}, n_{11})$ is equal to that for the other combination, the posterior probability for $(n_{00},\ n_{01}, n_{10}, n_{11}) = (2, 0, 2, 4)$ is calculated as 0.035. The posterior probability for the causal risk difference $(n_{10} - n_{01})/n = 0.25$ is calculated by summing the posterior probabilities for all combinations of $(n_{00},\ n_{01}, n_{10}, n_{11})$ with $(n_{10} - n_{01})/n = 0.25$. After calculating the posterior probabilities for the other values of $(n_{10} - n_{01})/n$, we obtain the posterior distribution of $(n_{10} - n_{01})/n$.

Finally, we note that Chiba [11] discussed the setting of Bernoulli trials in the context of the frequentist approach. In their setting, the number of subjects who were assigned to an arm was not fixed. Instead, the number depended on the ratio of random assignment. In the context of our Bayesian approach, when the assignment ratio is $1 : r$, the likelihood function for N_1, $g(N_1 \mid N)$, can be expressed as

$$g(N_1 \mid N) = \sum_{N_1 \in \mathcal{F}_1} \prod_{k=0}^{J-1} \prod_{l=0}^{J-1} \binom{n_{kl}}{n_{kl,1}} \left(\frac{1}{1+r}\right)^{n_{kl,1}} \left(\frac{r}{1+r}\right)^{n_{kl}-n_{kl,1}}$$

in the region $N \in \mathcal{F}$, and $g(N_1 \mid N) = 0$ outside $N \in \mathcal{F}$. Although the likelihood function $g(N_1 \mid N)$ is not equal to $f(N_1 \mid N)$, it is simple to verify that both functions yield the same posterior probability.

4 Illustration

We will now illustrate our proposed Bayesian approach using data from two randomized clinical trials. We analyze a trial with a binary outcome in Section 4.1 and a trial with an ordinal outcome with three categories in Section 4.2.

4.1 Example 1: Trial with a binary outcome

Harms et al. [18] reported the results of a randomized clinical trial of preventive antibacterial therapy for patients who had suffered an acute ischemic stroke. The purpose of this trial was to evaluate the effectiveness of moxifloxacin for preventing post-stroke infections. Seventy-nine patients were randomly assigned to either the moxifloxacin arm ($X = 1$) or the placebo arm ($X = 0$), with an assignment ratio of $1 : 1$. The primary endpoint was infection within 11 days. The results of this trial are summarized in Table 3.
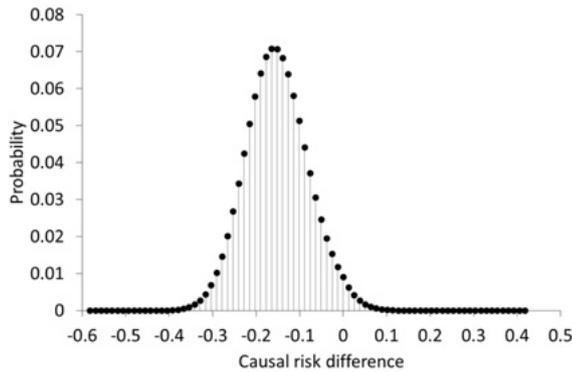

Figure 1: Posterior distribution of the causal risk difference (3), derived from data in Table 3.

Figure 1 shows the posterior distribution of (3) in the region $N \in \mathcal{F}$, derived by applying the Bayesian approach to the data in Table 3 with a non-informative prior distribution, so that the probability for each N was equal. In Table 4, we show specific estimated statistical measures, such as the maximum a posteriori estimate (MAP), 95 % credible interval (CI), expected a posteriori estimate (EAP), and 95 % highest density region (HDR). As a reference, in this table, we also show the crude risk difference under the exchangeability assumption, i. e., the estimate of $\Pr(Y = 1 \mid X = 1) - \Pr(Y = 1 \mid X = 0)$, which is equal to (2) for $J = 2$, and the exact 95 % confidence interval [19]. The MAP and EAP estimates were more conservative than the crude risk difference, and the widths of the CI and HDR were narrower than the width of the exact confidence interval.

We also show the posterior distributions of n_{kl} (k, $l = 0, 1$) in the appendix.

Table 3: Results from the randomized clinical trial of preventive antibacterial therapy: infection within 11 days.

Arm	Infection status		Total
	No infection ($Y = 0$)	Infection ($Y = 1$)	
Moxifloxacin ($X = 1$)	33	6	39
Placebo ($X = 0$)	27	13	40

Table 4: Estimates calculated from the data in Table 3.

Measure	Estimate
Maximum a posteriori estimator	$-13/79 = -0.165$
95 % credible interval	$(-23/79, 0/79) = (-0.291, 0.000)$
Expected a posteriori estimator	-0.152
95 % highest density region	$(-23/79, -1/79) = (-0.291, -0.013)$
Crude risk difference (2)[a]	$6/39 - 13/40 = -0.171$
Exact 95 % confidence interval[a]	$(-27/79, 2/79) = (-0.342, 0.025)$

[a] Crude risk difference (2) and the 95 % confidence interval were calculated under the exchangeability assumption.

4.2 Example 2: Trial with an ordinal outcome with three categories

Fox et al. [20] reported the results of a randomized clinical trial of preventive antiemetic therapy for patients with germ cell tumors or small-cell lung cancer. The purpose of this trial was to evaluate the effectiveness of combining ondansetron (OND) with dexamethasone and chlorpromazine (ODC) for preventing emetic episodes in patients receiving cisplatin. Forty-four patients were randomly assigned to either the ODC ($X = 1$) or OND alone ($X = 0$) arms, with a 1 : 1 assignment ratio. The responses were classified into three categories: complete response (if no emesis occurred during the study period, the response was classified as complete; $Y = 2$); major response (if one to two emetic episodes occurred during the study period, the response was classified as major; $Y = 1$); and minor or no response (if at least three emetic episodes occurred during the study period, the response was classified as minor; $Y = 0$). The results are summarized in Table 5.
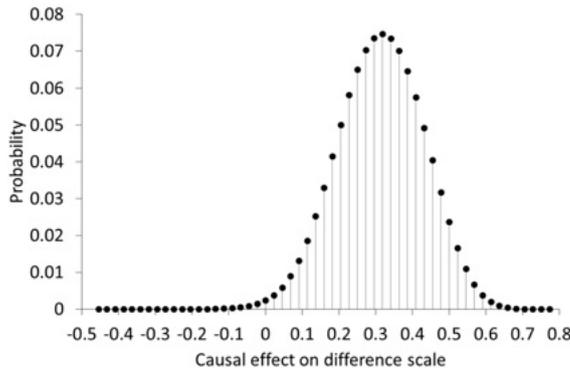

Figure 2: Posterior distribution of the causal effect on the difference scale (3), derived from data in Table 5.

Figure 2 shows the posterior distribution of (3) in the region $N \in \mathcal{F}$, derived by applying the Bayesian approach to the data in Table 5 with a non-informative prior distribution, so that the probability for each N was equal. In Table 6, we show the MAP estimate, 95 % CI, EAP estimate, and 95 % HDR. As a reference, in this

Table 5: Results from the randomized clinical trial of preventive antiemetic therapy: antiemetic response throughout the study period.

Arm	Level of response			Total
	Minor or no ($Y = 0$)	Major ($Y = 1$)	Complete ($Y = 2$)	
ODC[a]($X = 1$)	3	7	12	22
OND[b]($X = 0$)	12	3	7	22

[a]Combination of ondansetron, dexamethasone, and chlorpromazine.
[b]Ondansetron alone.

Table 6: Estimates calculated from the data in Table 5.

Measure	Estimate
Maximum a posteriori estimator	$14/44 = 0.318$
95 % credible interval	$(4/44, 23/44) = (0.091, 0.523)$
Expected a posteriori estimator	0.313
95 % highest density region	$(4/44, 23/44) = (0.091, 0.523)$
Crude difference (2)[a]	0.382
Exact 95 % confidence interval[a]	$(-5/44, 34/44) = (-0.114, 0.773)$

[a]Crude difference (2) and the 95 % confidence interval were calculated under the assumption of exchangeability and independent potential outcomes.

table, we also show the estimate of (2) for $J = 3$ under the assumptions of exchangeability and independent potential outcomes, and the exact 95 % confidence interval [11]. As with the results from Example 1, the MAP and EAP estimates were more conservative than the crude difference (2), and the widths of the CI and HDR were narrower than the width of the exact confidence interval. The differences between the widths were more notable than in Example 1.

We also show the posterior distributions of n_{kl} (k, $l = 0, 1, 2$) in the appendix.

5 Discussion

We have developed a new Bayesian method for making causal inferences from the results of randomized trials with ordinal outcomes, including binary outcomes. The advantage of this approach is that we do not have to place the assumptions of exchangeability, which is often assumed under random assignment, or independent potential outcomes, which is often assumed to identify causal effects for $J \geq 3$. We also do not require any modeling assumptions. This advantage is realized in comparison with approaches used to make the inference of the other causal measures introduced in Section 2. Volfovsky et al. [9] required a full parametric model and fixed the correlation between $Y(1)$ and $Y(0)$ to make a Bayesian inference of the conditional median $M_l = \text{median}\{Y(1) \mid Y(0) = l\}$. Lu et al. [10] required the exchangeability assumption to derive the closed forms of sharp bounds for $\tau = \Pr\{Y(1) \geq Y(0)\}$ and $\eta = \Pr\{Y(1) > Y(0)\}$. In comparison, our approach does not require these constraints to make a Bayesian inference of (3).

Our approach also has the advantage that it can make an inference about the number of subjects in each response type. Such an inference is potentially useful for making the detailed consideration of the characteristics of the treatment in the target population. For example, in the population for Example 1 in Section 4, we can infer that roughly 60 % of subjects might not be infected regardless of whether they received moxifloxacin or the placebo (see Appendix).

One disadvantage of our proposed approach is that the computational effort increases dramatically with the number of categories for the outcome. In Example 1 in Section 4 with a 2×2 contingency table for which the sample size was 79, there were 23,798 Ns in the region $N \in \mathcal{F}$, and we required less than one second to derive the posterior distribution shown in Figure 1. However, in Example 2 with a 2×3 contingency table, although the sample size was 44, there were 104 million Ns in the region $N \in \mathcal{F}$, and we required 100 minutes to derive Figure 2. The computational effort increases dramatically with the sample size and the number of categories. Although our approach is feasible for a small sample, it may not be feasible for a large sample, especially for an ordinal outcome with more than two categories. Further studies are required to develop a calculation method with less computational effort, for example, by proposing an efficient algorithm and by generating an approximation formula. An immediate approach would sample the Ns rather than enumerating all Ns, although this could still be a difficult process as there may be no simple uniform sampler.

Appendix

In this appendix, we present the posterior distributions of components of N derived from the data in Tables 3 and 5. Figure A.1 shows the posterior distributions of n_{kl} (k, $l = 0, 1$) derived from the data in Table 3. The EAP estimates were

$$\begin{pmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{pmatrix} = \begin{pmatrix} 46.3 & 19.0 \\ 7.0 & 6.8 \end{pmatrix};$$
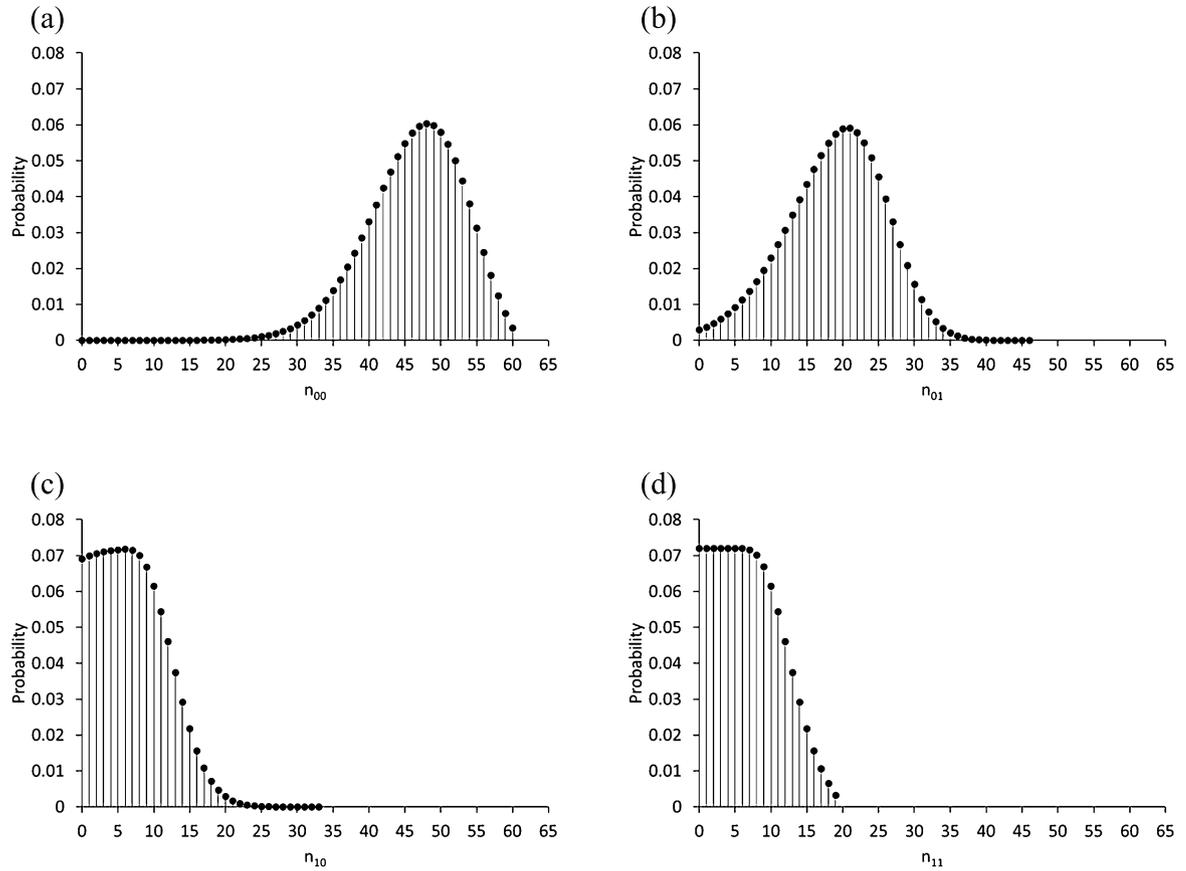
Figure A.1: Posterior distributions of n_{kl} (k, $l = 0$, 1) derived from the data in Table 3; (a) n_{00}, (b) n_{01}, (c) n_{10}, and (d) n_{11}.

the percentages for all 79 subjects were

$$\begin{pmatrix} 58.6\,\% & 24.0\,\% \\ 8.8\,\% & 8.6\,\% \end{pmatrix}.$$

Figure A.2 show the posterior distributions of n_{kl} (k, $l = 0$, 1, 2) derived from the data in Table 5. The posterior distributions of n_{21}, n_{11}, and n_{22} were the same as those of n_{00}, n_{02}, and n_{10}, respectively. This is because, in Table 5, the numbers in the categories with $(X, Y) = (1, 0)$, $(1, 1)$, and $(1, 2)$ are the same as those with $(X, Y) = (0, 1)$, $(0, 2)$, and $(0, 0)$, respectively. Therefore, in Figure A.2, we show the posterior distribution of n_{21}, n_{11}, and n_{22} in the same figure as the posterior distribution of n_{00}, n_{02}, and n_{10}, respectively. The EAP estimates were

$$\begin{pmatrix} n_{00} & n_{01} & n_{02} \\ n_{10} & n_{11} & n_{12} \\ n_{20} & n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} 3.2 & 1.6 & 2.5 \\ 6.8 & 2.5 & 4.8 \\ 12.6 & 3.2 & 6.8 \end{pmatrix};$$

the percentages for all 44 subjects were

$$\begin{pmatrix} 7.3\,\% & 3.5\,\% & 5.7\,\% \\ 15.5\,\% & 5.7\,\% & 10.8\,\% \\ 28.7\,\% & 7.3\,\% & 15.5\,\% \end{pmatrix}.$$
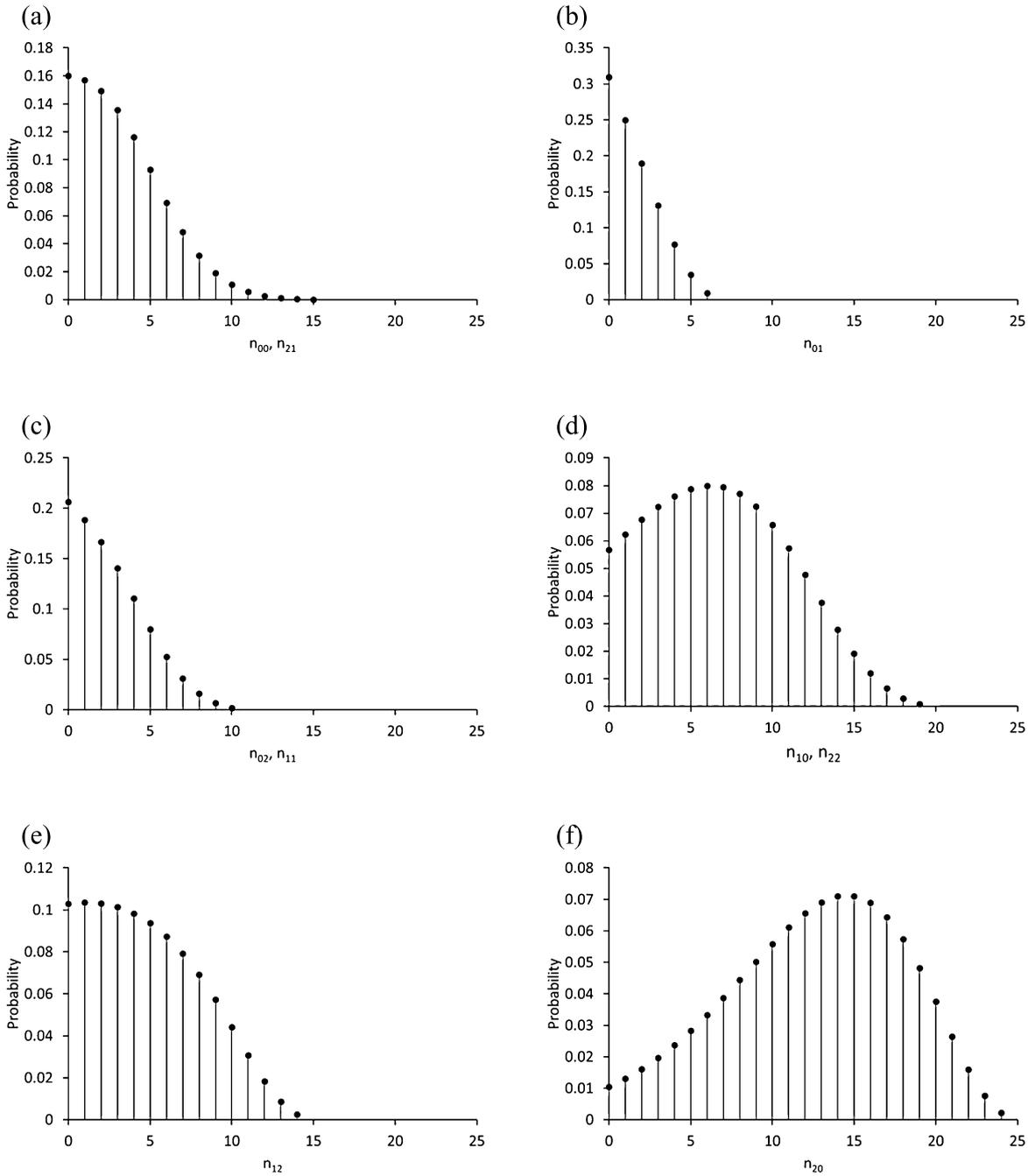
Figure A.2: Posterior distributions of n_{kl} (k, $l = 0, 1, 2$) derived from the data in Table 5; (a) n_{00} and n_{21}, (b) n_{01}, (c) n_{02} and n_{11}, (d) n_{10} and n_{22}, (e) n_{12}, and (f) n_{20}.

References

1. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Epidemiology. 1986;15:413–9.
2. Hernán MA, Robins JM. Causal inference. Boca Raton: Chapman and Hall/CRC; 2018.
3. Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons amongst survivors in randomized trials. Biometrics. 2005;61:305–10.
4. Rosenbaum P. Observational studies. New York: Springer; 2002.

5. Rigdon J, Hudgens M. Randomization inference for treatment effects on a binary outcome. Stat Med. 2015;34:924–35.
6. Ding P, Miratrix LW. Model-free causal inference of binary experimental data. Available at https://arxiv.org/abs/1705.08526.
7. Rubin DB. Bayesian inference for causal effects: the role of randomization. Ann Stat. 1978;6:34–58.
8. Rubin DB. Formal models of statistical inference for causal effects. J Stat Plan Inference. 1990;25:279–92.
9. Volfovsky A, Airoldi EM, Rubin DB. Causal inference for ordinal outcomes. Available at https://arxiv.org/abs/1501.01234.
10. Lu J, Ding P, Dasgupta T. Treatment effects on ordinal outcomes: causal estimands and sharp bounds. Available at https://arxiv.org/abs/1507.01542.
11. Chiba Y. Sharp nonparametric bounds and randomization inference for treatment effects on an ordinal outcome. Stat Med. 2017;36:3966–75.
12. Greenland S. On the logical justification of conditional tests for two-by-two contingency tables. Am Stat. 1992;45:248–51.
13. Klotz JH. The Wilcoxon ties, and the computer. J Am Stat Assoc. 1966;61:772–87.
14. Vargha A, Delaney HD. The Kruskal-Wallis test and stochastic homogeneity. J Educ Behav Stat. 1998;59:137–42.
15. Agresti A. Analysis of ordinal categorical data. 2nd ed. New Jersey: John Wiley and Sons; 2010.
16. Manski CF. Nonparametric bounds on treatment effects. Am Econ Rev. 1990;80:319–23.
17. Pearl J. Causal inference from indirect experiments. Artif Intell Med. 1995;7:561–82.
18. Harms H, Prass K, Meisel C, et al.. Preventive antibacterial therapy in acute ischemic stroke: a randomized controlled trial. PLoS ONE. 2008;3:e2158.
19. Chiba Y. Exact tests for the weak causal null hypothesis on a binary outcome in randomized trials. J Biometr Biostat. 2015;6:244.
20. Fox SM, Einhorn LH, Cox E, Powell N, Abdy A. Ondansetron versus ondansetron, dexamethasone, and chlorpromazine in the prevention of nausea and vomiting associated with multiple-day cisplatin chemotherapy. J Clin Oncol. 1993;11:2391–5.