

Research Article

Thomas Leavitt*

Randomization-based, Bayesian inference of causal effects

<https://doi.org/10.1515/jci-2022-0025>

received April 10, 2022; accepted March 21, 2023

Abstract: Bayesian causal inference in randomized experiments usually imposes model-based structure on potential outcomes. Yet causal inferences from randomized experiments are especially credible because they depend on a known assignment process, *not* a probability model of potential outcomes. In this article, I derive a randomization-based procedure for Bayesian inference of causal effects in a finite population setting. I formally show that this procedure satisfies Bayesian analogues of unbiasedness and consistency under weak conditions on a prior distribution. Unlike existing model-based methods of Bayesian causal inference, my procedure supposes neither probability models that generate potential outcomes nor independent and identically distributed random sampling. Unlike existing randomization-based methods of Bayesian causal inference, my procedure does not suppose that potential outcomes are discrete and bounded. Consequently, researchers can reap the benefits of Bayesian inference without sacrificing the properties that make inferences from randomized experiments especially credible in the first place.

Keywords: design-based inference, potential outcomes, finite population inference, limited information likelihood

MSC 2020: 62-02, 62B15, 62C10, 62D99

1 Introduction

Causal inferences from randomized experiments are especially credible – and indeed are indispensable to the “credibility revolution” [1] – since their validity depends primarily on the integrity of the data collection and the adherence of statistical analyses to fundamental features of the design [2]. Importantly, causal inferences from randomized experiments do *not* depend on probability models for the response variable or on an assumed sampling process from an often vaguely defined superpopulation [3]. Instead, randomized experiments draw on the randomization process itself as the “reasoned basis” for inference [4, p. 14].

Scholars also acknowledge the benefits that Bayesian inference could have for methods associated with the credibility revolution. One such benefit is the interpretation and presentation of statistical analyses [5–7]. Another is the apparatus that Bayesian inference provides for formally quantifying how much one learns from a new randomized experiment – the value of which Deaton and Cartwright [8, p. 3] underscore in their emphasis on “understanding how the results from randomized controlled trials relate to the knowledge that you already possess about the world.”

The contribution of this article is to develop Bayesian inference that is justified by the experimental design. The development of such inference has been difficult because, absent a probability model of potential outcomes, a likelihood function based solely on the assignment mechanism will be unidentified for causal effects of interest – i.e., flat over different hypothetical values of causal effects. Randomization

* **Corresponding author: Thomas Leavitt**, Marxe School of Public and International Affairs, Baruch College, City University of New York (CUNY), 135 East 22nd Street, New York, NY 10011, USA, e-mail: thomas.leavitt@baruch.cuny.edu

serves as the basis for Bayesian causal inference only for specific cases in which binary, ordinal, or otherwise discrete and bounded potential outcomes provide additional structure [9–12]. Otherwise, Bayesian causal inference of randomized experiments typically embeds potential outcomes in probability models or assumes random sampling from an infinite superpopulation [13–18].

To circumvent this problem, I use a normal “working model” to derive a likelihood for all data types that conditions *not* on the full realized data, but a suitably defined function (test-statistic) of them. While the likelihood is derived from a normal density of the test-statistic, this normal density is used only to construct the Bayesian procedure and is not assumed to be true when deriving the procedure’s randomization-based properties. I then prove, first, a randomization-based, Bayesian analog of unbiasedness wherein a flat prior on causal effects implies that the posterior’s most plausible effect is equal, in expectation (over random assignments), to the true effect. I then prove a randomization-based, Bayesian analog of consistency wherein, if the true effect is in the prior’s support, then, as the number of experimental subjects increases indefinitely, the resulting sequence of posterior distributions concentrates around the true effect. Both results hold in a strictly randomization-based, finite population setting, thereby opening up new possibilities for Bayesian inference of causal effects.

Section 2 provides the formal setup for the article and Section 3 describes extant approaches to Bayesian inference. Section 4 introduces the likelihood for randomization-based, Bayesian inference. Section 5 derives randomization-based, Bayesian unbiasedness and consistency for the average effect via two different procedures. Each procedure postulates a hypothetical average effect, but the first uses a conservative plug-in variance estimator, while the second uses a sharp causal effect consistent with the hypothetical average effect to directly calculate the variance of the difference-in-means. Section 6 concludes by pointing to open questions about the performance of this article’s Bayesian procedure relative to existing randomization- and model-based alternatives.

2 Formal setup

Consider a finite study population that consists of $N \geq 4$ units, and let the index $i = 1, \dots, N$ run over these N units. The indicator variable $z_i = 1$ or $z_i = 0$ denotes whether individual unit i is assigned to treatment ($z_i = 1$) or control ($z_i = 0$). The vector $\mathbf{z} = [z_1 \ z_2 \ \dots \ z_N]^\top$, where the superscript \top denotes matrix transposition, is the collection of N individual treatment indicator variables. The set of treatment assignment vectors is $\{0, 1\}^N$, which has cardinality $|\{0, 1\}^N| = 2^N$, where $|\cdot|$ denotes the cardinality of (i.e., number of elements in) a set.

Adopting the terminology of Freedman [19] and later Gerber and Green [20], a potential outcome schedule is defined as a vector-valued function, $\mathbf{y} : \{0, 1\}^N \mapsto \mathbb{R}^N$, which maps the set of assignments to an N -dimensional vector of real numbers. The vectors of potential outcomes, denoted by $\mathbf{y}(\mathbf{z})$ for $\mathbf{z} \in \{0, 1\}^N$, are the elements in the range of the potential outcomes schedule. The individual potential outcomes for unit i are the i th entries of each of the N -dimensional vectors of potential outcomes, denoted by $y_i(\mathbf{z})$ for $\mathbf{z} \in \{0, 1\}^N$.

With 2^N assignments, there are in principle 2^N potential outcomes for each individual unit. The stable unit treatment value assumption (SUTVA) [21–23] states that (1) units in the experiment would respond only to the condition to which each unit could be individually assigned and (2) the treatment condition would actually be the same treatment for all units and the control condition would actually be the same control for all units. In other words, SUTVA means that each unit has at most two distinct potential outcomes.

Assumption 1. (SUTVA) For all $i = 1, \dots, N$ units, $y_i(\mathbf{z})$ takes on a fixed value, $y_i(1)$, for all $\mathbf{z} : z_i = 1$ and another fixed value, $y_i(0)$, for all $\mathbf{z} : z_i = 0$.

SUTVA implies that a potential outcome for unit i can be expressed as $y_i(\mathbf{z})$, which is either $y_i(1)$ or $y_i(0)$ depending on whether \mathbf{z} is with $z_i = 1$ or $z_i = 0$. Under SUTVA, the vector of potential outcomes when all units are assigned to treatment is equal to the collection of all units’ treated potential outcomes,

$\mathbf{y}(\mathbf{1}) = [y_1(1) \ y_2(1) \ \dots \ y_N(1)]^\top$, and the vector of potential outcomes when all units are assigned to control is equal to the collection of all units' control potential outcomes, $\mathbf{y}(\mathbf{0}) = [y_1(0) \ y_2(0) \ \dots \ y_N(0)]^\top$.

The assignment mechanism selects a single $\mathbf{z} \in \{0, 1\}^N$ with probability $p(\mathbf{z})$. Hence, the treatment assignment vector is a random quantity, \mathbf{Z} , which takes on the value $\mathbf{z} \in \{0, 1\}^N$ with probability $\Pr(\mathbf{Z} = \mathbf{z})$. Likewise, the vector of potential outcomes is a random quantity, $\mathbf{y}(\mathbf{Z})$, in which its randomness is inherited from only \mathbf{Z} . Going forward, I assume complete random assignment (CRA), whereby, of the $N \geq 4$ units in the finite study population, $n_1 \geq 2$ are assigned to treatment and the remaining $n_0 := N - n_1 \geq 2$ are assigned to control.

Assumption 2. (CRA) The set of allowable assignments is $\Omega := \{\mathbf{z} : p(\mathbf{z}) > 0\} = \{\mathbf{z} : \sum_{i=1}^N z_i = n_1\}$ with $n_1 \geq 2$, $n_0 \geq 2$ and $p(\mathbf{z}) = 1/|\Omega|$ for all $\mathbf{z} \in \Omega$.

Under CRA, the number of treated units, n_1 , is fixed by design. However, under N independent Bernoulli trials, which can yield 2^N possible assignments, n_1 can be fixed by conditioning on its observed value. The randomization distribution conditional on the realized n_1 yields the same randomization distribution one would obtain if n_1 had been fixed ex ante by design [24, p. 289–290]. Hence, this general setup pertains to both simple and complete randomized assignments, although the argument by which one can regard n_1 as fixed is slightly different for each assignment mechanism.

To define the randomized experiment's inferential target, first denote the individual treatment effect on the additive scale by $\tau_i := y_i(1) - y_i(0)$ and the collection of the N individual treatment effects by $\boldsymbol{\tau} := [\tau_1 \ \tau_2 \ \dots \ \tau_N]^\top = \mathbf{y}(\mathbf{1}) - \mathbf{y}(\mathbf{0})$. For the purposes of this article, the primary causal target, usually associated with Neyman [25], is the average effect, denoted by $\tau := N^{-1} \sum_{i=1}^N \tau_i = \bar{y}_1 - \bar{y}_0$, where \bar{y}_z is the mean of either treated ($z = 1$) or control ($z = 0$) potential outcomes. This causal target differs from an alternative target, usually associated with Fisher [4], which in principle could be the full N -dimensional vector of individual causal effects, $\boldsymbol{\tau}$. A key difference between the average effect and the entire N -dimensional vector of individual effects is that knowledge of the latter, along with observed outcomes, implies all values of unobserved potential outcomes. Hence, $\boldsymbol{\tau}$ is a *sharp* causal effect. The average effect, τ , does not (together with observed outcomes) imply values of all unobserved potential outcomes; hence, τ is a *weak* causal effect.

The average effect, τ , is a single, unknown value. However, there are many possible values to which this unknown τ could be equal. I therefore denote a hypothetical value of the true but unknown average effect by τ_h and the set of these hypothetical values by $\mathcal{T}_h \subseteq \mathbb{R}$. I partition the set \mathcal{T}_h as $\mathcal{T}_h^* := \{\tau_h : |\tau - \tau_h| \leq \varepsilon\}$, $\mathcal{T}_h^- := \{\tau_h : \tau - \tau_h > \varepsilon\}$ and $\mathcal{T}_h^+ := \{\tau_h : \tau_h - \tau > \varepsilon\}$, where ε is an arbitrarily small constant greater than 0. The set \mathcal{T}_h^* consists of the hypothetical average effects within a distance of ε from the true, unknown effect. By contrast, \mathcal{T}_h^- and \mathcal{T}_h^+ are the hypothetical average effects that are smaller and larger than the true effect by a distance of ε , respectively. Subjective uncertainty about \mathcal{T}_h is represented by a prior density function, $r : \mathbb{R} \mapsto [0, \infty)$, where $r(\tau_h)$ for all $\tau_h \in \mathcal{T}_h$ represents the subjective belief that τ_h is equal to the true, unknown τ .

3 Extant approaches to Bayesian causal inference

Bayesian inference typically proceeds by defining a likelihood that conditions on the full realized data, which are usually conceived as independent and identically distributed (i.i.d.) draws from a probability distribution of potential outcomes. However, in the randomization-based, finite population setting described in Section 2, potential outcomes are fixed quantities and randomness stems solely from the probability distribution on the set of assignments. In this setting, an identified likelihood that conditions on the full realized data is difficult to construct.

To unpack this difficulty, let $\boldsymbol{\tau}_h$ denote a vector of the hypothetical individual effects for all N experimental units, and then denote the outcomes adjusted (or centered) by this vector of hypothetical individual

effects by $\mathbf{y}(\mathbf{Z}) - \boldsymbol{\tau}_h \odot \mathbf{Z}$, where \odot is the Hadamard (element-wise) product of two matrices with the same dimensions. Because $\boldsymbol{\tau}_h$ is a sharp hypothetical effect, we can derive an analog to what Aronow and Miller [26, p. 93] refer to as a finite population probability mass function (PMF):

$$f(\mathbf{t}) = \sum_{\mathbf{z} \in \Omega} \mathbb{I}\{\mathbf{y}(\mathbf{z}) - \boldsymbol{\tau}_h \odot \mathbf{z} = \mathbf{t}\} \Pr(\mathbf{Z} = \mathbf{z}), \quad (1)$$

where $\mathbf{t} \in \mathbb{R}^N$ and $\mathbb{I}\{\cdot\}$ is the indicator function. Equation (1) is the true, unknown PMF of outcomes adjusted by $\boldsymbol{\tau}_h$.

In practice, a researcher observes only one assignment and then constructs a PMF implied by the supposition that $\boldsymbol{\tau}_h$ is true. The supposition that $\boldsymbol{\tau}_h = \boldsymbol{\tau}$ implies that the vector of adjusted outcomes is fixed over all $\mathbf{z} \in \Omega$. I therefore write this PMF, given a single realization of data, $\{\mathbf{z}, \mathbf{y}(\mathbf{z})\}$, as follows:

$$\hat{f}(\mathbf{t}) = \sum_{\mathbf{w} \in \Omega} \mathbb{I}\{\mathbf{y}(\mathbf{z}) - \boldsymbol{\tau}_h \odot \mathbf{z} = \mathbf{t}\} \Pr(\mathbf{W} = \mathbf{w}), \quad (2)$$

where \mathbf{w} is a relabeling of the treatment allocations after observing data, $\{\mathbf{z}, \mathbf{y}(\mathbf{z})\}$, and the “hat” operator denotes functions of observed quantities – functions whose stochastic properties will often be the object of analysis herein. For any $\boldsymbol{\tau}_h$, the vector of adjusted outcomes in equation (2), $\mathbf{y}(\mathbf{z}) - \boldsymbol{\tau}_h \odot \mathbf{z}$, would be fixed at its observed value over all assignments if $\boldsymbol{\tau}_h$ were true. Hence, a likelihood derived from equation (2) will be flat (unidentified) over different values of $\boldsymbol{\tau}_h$.

The PMFs in equations (1) and (2) use outcomes adjusted by $\boldsymbol{\tau}_h$, as in Rosenbaum [27,28]. However, the same identification problem holds for analogous PMFs of treated and control potential outcomes that use the “science table” implied by $\boldsymbol{\tau}_h$, as in Rubin [29]. The supposition that $\boldsymbol{\tau}_h$ is true implies that treated potential outcomes are $\mathbf{z} \odot \mathbf{y}(\mathbf{z}) + (\mathbf{1} - \mathbf{z}) \odot (\mathbf{y}(\mathbf{z}) + \boldsymbol{\tau}_h)$ and control potential outcomes are $\mathbf{z} \odot (\mathbf{y}(\mathbf{z}) - \boldsymbol{\tau}_h) + (\mathbf{1} - \mathbf{z}) \odot \mathbf{y}(\mathbf{z})$. Given some realization of data, treated potential outcomes implied by any $\boldsymbol{\tau}_h$ are always equal to the treated units’ observed outcomes and control potential outcomes implied by any $\boldsymbol{\tau}_h$ are always equal to the control units’ observed outcomes. Hence, all $\boldsymbol{\tau}_h$ are equally consistent with the observed data.

This identification problem can be resolved when outcomes are discrete and bounded. In particular, Copas [30] shows that, under reasonable assumptions, it is possible to derive an identified, randomization-based likelihood that conditions on the full data when outcomes are binary. Ding and Miratrix [9] then show how scholars can conduct model-free, Bayesian inference of causal effects via such a randomization-based likelihood [see also 11,12]. Chiba [10] extends this logic for the case of binary outcomes to that of ordinal outcomes. Yet the lack of such outcomes in many applications often makes this randomization-based likelihood untenable.

Given the difficulties of such model-free Bayesian inference, an alternative is to derive a likelihood from a probability model of the joint distribution of potential outcomes, which is the standard approach to Bayesian causal inference [13,18]. In this approach, as Imbens and Rubin [15, p. 141] state, “potential outcomes themselves are also viewed as random variables, even in the finite sample.” With stochastic potential outcomes, the essential role of randomization is that it implies that one can “ignore the assignment mechanism when making causal inferences” [31, p. 233], hence the term “ignorability” [13,31].

Under random assignment, we can ignore the assignment process and instead consider only (1) the prior distribution of the potential outcome model’s parameters and (2) the likelihood of the potential outcomes conditional on the model parameters governing the joint distribution of potential outcomes. With this likelihood, we can derive a posterior distribution of the average effect, defined as a function of the potential outcome model’s parameters: Upon updating on the parameters governing potential outcomes’ marginal distributions [32], we can (1) draw from the posterior distribution of the model parameters, (2) input each draw as the parameters of the model of potential outcomes, (3) draw each missing potential outcome from the model of potential outcomes, and (4) directly calculate a function of the two vectors of (partially observed and partially imputed) potential outcomes, namely, the average effect. Repeating this procedure many times yields a simulation-based approximation to the average effect’s posterior distribution.

A concern of this methodology is that inference now depends on a stochastic model of potential outcomes, not only a known assignment process. Thus, one of the central appeals of experiments – their “reasoned basis” for inference – may no longer hold. Imbens and Rubin [15, p. 142] emphasize that “[o]ne of the practical issues in the model-based approach is the choice of a credible model for imputing the missing potential outcomes” and that “fundamentally the resulting inference may be more sensitive to the modeling assumptions.” Thus, the ability to conduct Bayesian inference from an experiment may come at the expense of inference that is justified by a known assignment process.

For this reason, understanding the randomization-based properties of model-based, Bayesian inference has been of longstanding interest [33–35]. To this end, both Dasgupta et al. [34] and Ding and Dasgupta [35] derive the posterior mean and variance of the average effect under specific models of potential outcomes and prior distributions of the models’ parameters, and then show how the posterior means and variances differ from the difference-in-means and Neyman’s conservative variance estimator, respectively. For example, Ding and Dasgupta [35] show that, under a standard modeling assumption for binary potential outcomes and a Beta prior, the average effect’s posterior variance is less than Neyman’s conservative variance estimator. Hence, resulting credible intervals may not have correct Frequentist coverage when individual causal effects are homogeneous.

Much progress can be made from a case-by-case analysis of the model-based approach’s randomization-based properties under particular prior distributions and outcome models for specific data types. The following section takes a different tack by proposing a single likelihood that works for different data types. This likelihood enables practitioners to define priors directly on the average effect, *not* parameters of a stochastic model, and yields unbiasedness and consistency under a mild condition on the average effect’s prior distribution.

4 A normal likelihood for the difference-in-means

In this section, I propose a normal likelihood of causal effects that conditions *not* on the full realized data, but rather on a test-statistic of them. A suitable test-statistic for inference of the average effect under SUTVA and CRA is the difference-in-means:

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) = \left(\frac{1}{n_1} \right) \mathbf{Z}^\top \mathbf{y}(\mathbf{Z}) - \left(\frac{1}{n_0} \right) (\mathbf{1} - \mathbf{Z})^\top \mathbf{y}(\mathbf{Z}), \quad (3)$$

where, under CRA, n_1 and n_0 are fixed over all $\mathbf{z} \in \Omega$. Under SUTVA and CRA, the variance of the difference-in-means, as Neyman [25] shows, is

$$\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} - \frac{S_{10}^2}{N}, \quad (4)$$

where

$$\begin{aligned} S_1^2 &= \frac{1}{(N-1)} \sum_{i=1}^N (y_i(1) - \bar{y}_1)^2 \\ S_0^2 &= \frac{1}{(N-1)} \sum_{i=1}^N (y_i(0) - \bar{y}_0)^2 \\ S_{10}^2 &= \frac{1}{(N-1)} \sum_{i=1}^N (\tau_i - \tau)^2. \end{aligned}$$

Both the expectation, $E[\cdot]$, and variance, $\text{Var}[\cdot]$, are taken over the set of assignments, as are all other expectations and variances going forward.

The finite population central limit theorem (CLT) [36] implies that the standardized difference-in-means converges in distribution to standard normal, i.e.,

$$G(t) = \sum_{z \in \Omega} \mathbb{1} \left\{ \frac{\hat{\tau}(z, y(z)) - E[\hat{\tau}(Z, y(Z))]}{\sqrt{\text{Var}[\hat{\tau}(Z, y(Z))}}} \leq t \right\} \Pr(Z = z) \xrightarrow{d} \Phi, \quad (5)$$

where $t \in \mathbb{R}$, Φ is the standard normal cumulative distribution function, and \xrightarrow{d} denotes convergence in distribution. Associated theory, namely, Berry-Esseen bounds [37,38], has been established for finite population sampling without replacement in Bikelis [39] and Höglund [40] and, more recently, for randomization-based causal inference in Wang and Li [41] and Shi and Ding [42]. Such bounds imply that, so long as an experiment is of at least moderate size and units' potential outcomes are not too skewed or characterized by extreme outliers, the distribution of the standardized difference-in-means will be well approximated by the standard normal distribution.

For inference via null hypothesis significance testing, the convergence of the standardized difference-in-means in equation (5) and associated theory justify the use of normal approximation-based p -values. For Bayesian inference, by contrast, a similar justification for a normal approximation-based likelihood is *not* available. In short, the standardized difference-in-means' convergence in distribution to standard normal does not imply the convergence of the standardized difference-in-means' PMF to a standard normal density. Boos [43] and Sweeting [44] derive conditions under which convergence in distribution implies convergence in density. However, these conditions – namely, that of an equicontinuous sequence of densities – are difficult to reconcile with a randomization-based framework in which the difference-in-means is discrete.

Despite the inapplicability of the finite population CLT, I draw on a standard normal density for the likelihood of the standardized difference-in-means. Use of this normal likelihood is to be interpreted as a “working model” that is used to construct a Bayesian inferential procedure, but is not actually assumed to be true. The use of such “working models” is common in randomization-based causal inference, such as in the literature on ordinary least squares (OLS) regression adjustment wherein Lin [45, p. 314] writes, “[o]ne does not need to believe in the classical linear model to tolerate or even advocate OLS adjustment.” Likewise, one does not need to believe that the standardized difference-in-means' PMF is well approximated by a standard normal density to justify the use of a standard normal likelihood.

Use of a standard normal likelihood proceeds by, first, centering the difference-in-means by a hypothesis about its expectation, which, under SUTVA and CRA, is a hypothesis about the unknown average effect. In contrast to its expectation, the difference-in-means' variance is a nuisance parameter that is not of primary interest. The standard Bayesian approach to eliminating a nuisance parameter is by marginalizing over its distribution [46,47]. However, applied researchers are unlikely to have well-motivated prior beliefs about the variance of the difference-in-means, which makes this approach unattractive. Instead, to construct the test-statistic, I propose two plug-in approaches to eliminating the variance nuisance parameter, which use both a weak and sharp causal hypothesis, respectively. Under either approach, the likelihood of a hypothetical average effect is the standard normal density evaluated at the difference-in-means centered by this hypothetical effect and then divided by the square root of a variance plug-in that the researcher acts *as if* it were the true variance.

By using such a likelihood that conditions on a test statistic, not the full data themselves, this procedure is a “limited information” approach [48–50]. That is, the likelihood draws information from only the test-statistic, which, if a full-data likelihood from a probability model of potential outcomes were available, could result in a loss of information. Notwithstanding this hypothetical loss of information, the principal justification for a “limited information” likelihood is its asymptotic properties under minimal modeling assumptions [48,49,51]. For this reason, the use of a “limited information” Bayesian approach is not without precedent in randomization-based causal inference [52].

A crucial practical benefit of this “limited information” approach with a normal “working model” is its ease for practitioners. Under SUTVA and CRA, the possible values of the causal target, the average effect, are identical to the possible values of the normal model's location parameter; hence, practitioners can define priors directly on possible values of the average effect. Existing model-based procedures, by contrast, suppose that researchers define priors (in principle, researchers' subjective credences) over parameters of statistical models that are extrinsic to the randomization-based, finite population setting. Defining priors on

such quantities that are not grounded in potential outcomes or (to the extent possible) mapping from priors on functions of potential outcomes to parameters of a statistical model is a difficult exercise.

In what follows, I show that inference via a “limited information,” normal likelihood results in both Bayesian unbiasedness and consistency *regardless of how well the standard normal density approximates the PMF of the standardized difference-in-means*. The statistical properties – and, hence, credibility – of the procedures are based on randomization alone. While it is possible that different models under the standard Bayesian approach may achieve the same properties, the benefit of the forthcoming analysis is that it establishes the credibility of a single Bayesian procedure that researchers can use out of the box for different data types.

5 Bayesian unbiasedness and consistency

The argument to follow embeds a finite experimental population in an imaginary sequence of finite populations of increasing sizes. Bayesian unbiasedness is a property that holds for a fixed experimental population, i.e., for any finite experimental population in the sequence of populations. In contrast to Bayesian unbiasedness, Bayesian consistency is a limiting property over this infinite sequence of finite populations of increasing sizes.

A common asymptotic regime in randomization-based causal inference [as in, e.g., 53,54] is given by Brewer [55]. In this conception of asymptotic growth, each finite population in the infinite sequence of finite populations is a concatenation of a specified number of copies of the original finite population. That is, the original population of N units is copied $m - 1$ times, where $m = 1, 2, \dots$, and exactly $n_1 m$ out of the Nm total units are assigned to treatment. In this asymptotic regime, all relevant quantities, namely, n_1 / N , as well as the means, variances, and covariance of potential outcomes, are fixed constants over all elements in the sequence of finite populations.

This asymptotic regime implicitly embeds several regularity conditions that are standard in the literature [see, e.g., 45,56,57, among others]. In contrast to the asymptotic regime in Brewer [55], I assume only these regularity conditions whereby there is no other information whatsoever between any two populations in the sequence of finite populations. For the value of such an asymptotic regime, see the brief but insightful discussion in Sävje et al. [58, Section 5]; Delevoye and Sävje [59] also discuss in passing the value of this asymptotic regime.

The mild regularity conditions on the infinite sequence of finite populations are in Assumption 3. In writing these conditions and in laying out the general asymptotic argument moving forward, one ought to index potential outcomes and other quantities in the infinite sequence of finite populations by $N \in \mathbb{N}_{\geq 4}$, where $\mathbb{N}_{\geq 4}$ is the set of natural numbers greater than or equal to 4. However, for cleaner notation and in accordance with standard practice, I leave this indexing implicit. In addition, for some quantities, e.g., τ , I do not use notation to distinguish between its value for a specific $N \in \mathbb{N}_{\geq 4}$ and its limiting value. References to such quantities' limits should be clear from context.

Assumption 3. (Finite population asymptotic regularity conditions)

Condition 3.1: As $N \rightarrow \infty$, the proportion of treated units, n_1 / N , tends to a positive value strictly greater than 0 and less than 1, i.e., $\lim_{N \rightarrow \infty} n_1 / N = v$, where $v \in (0, 1)$.

Condition 3.2: The population means, variances, and covariance of treated and control potential outcomes are Cesàro summable, i.e., tend to finite limits as $N \rightarrow \infty$, in which the limiting variances, $S_{0,\infty}^2$ and $S_{1,\infty}^2$, are both greater than 0 and the limiting covariance is greater than $(S_{0,\infty}^2 S_{1,\infty}^2)^{-1/2}$.

Condition 3.3: Potential outcomes have bounded fourth moments, i.e., for all $N \in \mathbb{N}_{\geq 4}$,

$$\frac{1}{N} \sum_{i=1}^N (y_i(z) - \bar{y}_z)^4 < L < \infty \quad \text{for } z = 0 \quad \text{and} \quad z = 1,$$

where L is a positive, real number.

The importance of Conditions 3.1 and 3.2 is to ensure that the relevant limits exist over the sequence of finite populations of increasing sizes. The condition on the limiting covariance in 3.2 ensures that potential outcomes are not perfectly negatively correlated and, consequently, that the limiting variance of the difference-in-means is not 0 when $v = (1/2)$. Regularity condition 3.3 ensures the strong law of large numbers of mean and variance estimators in a finite population setting [60, Lemma A3].

5.1 Weak causal effects

The first approach to Bayesian inference of the average effect centers the difference-in-means by a weak causal hypothesis and then plugs in a variance estimator for the difference-in-means' variance. Neyman's conservative variance estimator [25] is a natural plug-in for the variance of the difference-in-means. This variance estimator is

$$\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_0^2}{n_0}, \quad (6)$$

where

$$\begin{aligned} \hat{S}_1^2 &= \frac{1}{(n_1 - 1)} \sum_{i=1}^N Z_i \left(y_i(1) - \frac{1}{n_1} \sum_{i=1}^N Z_i y_i(1) \right)^2 \quad \text{and} \\ \hat{S}_0^2 &= \frac{1}{(n_0 - 1)} \sum_{i=1}^N (1 - Z_i) \left(y_i(0) - \frac{1}{n_0} \sum_{i=1}^N (1 - Z_i) y_i(0) \right)^2. \end{aligned}$$

When causal effects are homogeneous for all units, the conservative variance estimator in equation (6) is equal, in expectation, to $\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]$. Otherwise, the expectation of Neyman's conservative variance estimator is greater than $\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]$. The use of improved, conservative estimators is also possible, e.g., from Aronow et al. [61] in completely randomized experiments and Imai [62], Fogarty [63], and Pashley and Miratrix [64] in finely stratified experiments.

We can write the difference-in-means centered by a hypothetical average effect and then divided by the square root of Neyman's conservative variance estimator as follows:

$$\frac{\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h}{\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}}. \quad (7)$$

This test-statistic's exact, unknown PMF based on only the random assignment process is

$$g(t) = \sum_{\mathbf{z} \in \Omega} \mathbb{1} \left\{ \frac{\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h}{\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} = t \right\} \Pr(\mathbf{Z} = \mathbf{z}). \quad (8)$$

If one were to suppose that $\tau_h = \tau$, the exact PMF in equation (8) would remain unknown since the hypothetical effect is weak, i.e., does not imply values for missing potential outcomes.

Upon observing experimental data summarized by the test-statistic in equation (7), let what I term the reference density be

$$\hat{\phi}(t) = \phi(t), \quad (9)$$

where $\phi(\cdot)$ is the standard normal density. This reference density in equation (9), as distinct from the PMF in equation (8), is analogous to the notion of a reference distribution for the calculation of p -values in null

hypothesis significance testing. Just as one calculates a p -value by evaluating the reference distribution at t equal to the test-statistic in (7), evaluating the reference density at the same test-statistic yields a probability density. For the purposes of Bayesian inference, this reference density reflects the researcher's proceeding, upon observing data, as if τ_h were the unknown average effect, the variance of the difference-in-means were $\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]$, and the normal density were the data-generating process of the test-statistic.

Despite the absence of any guarantee that the reference density in equation (9) approximates the exact PMF in equation (8), the use of this density suffices for a Bayesian analog of unbiasedness. That is, with a flat prior, the maximum *a posteriori* probability (MAP), i.e., the posterior mode, is equal, in expectation (over random assignments), to the true effect. This property is formally established in Proposition 1.

Proposition 1. (Bayesian unbiasedness) *Suppose Assumptions 1 and 2, and let the prior density of \mathcal{T}_h be uniform. It follows that the MAP is equal, in expectation, to the average effect, i.e.,*

$$\mathbb{E} \left[\operatorname{argmax}_{\tau_h \in \mathcal{T}_h} \frac{\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h) / \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}) r(\tau_h)}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h) / \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}) r(\tau_h) d\tau_h} \right] = \tau. \quad (10)$$

The proof of this proposition as well as all other formal results are in Appendix A. The general logic of Proposition 1 is that, given some realization of data, the value of τ_h that maximizes the standard normal likelihood is whichever value of τ_h is equal to the difference-in-means. With a uniform prior, the MAP is equal to the value of τ_h that maximizes the likelihood; hence, the MAP is the difference-in-means. Since the expectation of the difference-in-means is equal to τ , the expected MAP is also equal to τ . This property of Bayesian unbiasedness, despite the use of a standard normal likelihood for inference, holds regardless of the accuracy of the normal approximation and for randomized experiments with as few as $N = 2$ units with one unit assigned to treatment and the other to control.

Having established a Bayesian analog of unbiasedness, I now turn to Bayesian consistency, which states that, as the size of the experiment increases indefinitely, the posterior distribution concentrates around the true effect with probability 1. Formally, we can state this theorem as follows:

Theorem 1. (Bayesian consistency) *Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. If $\tau_h : \tau_h = \tau$ is in the support of the prior distribution, then, as the size of the experiment increases indefinitely, the posterior probability of \mathcal{T}_h^* converges a.s. to 1:*

$$\frac{\int_{\tau_h \in \mathcal{T}_h^*} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h) / \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}) r(\tau_h) d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h) / \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}) r(\tau_h) d\tau_h} \xrightarrow{\text{a.s.}} 1. \quad (11)$$

Here, I present a proof sketch in five primary steps. (1) The first step establishes that the variance estimator converges a.s. to a fixed constant that is greater than or equal to the limit of the (suitably scaled) test-statistic's true variance. (2) The second step builds on the a.s. convergence result in step (1) to show that, for all $\tau_h : |\tau - \tau_h| > \varepsilon$, the standardized test-statistic in equation (7) diverges a.s. to ∞ , which implies that, for all $\tau_h : |\tau - \tau_h| > \varepsilon$, the standard normal density of the test-statistic in equation (7) converges a.s. to 0. (3) A subsequent application of the CMT implies that, for all $\tau_h : |\tau_h - \tau| > \varepsilon$, the likelihood multiplied by the prior, $r(\tau_h)$, which is fixed over all $N \in \mathbb{N}_{\geq 4}$, converges a.s. to 0. (4) Step 4 begins by noting, first, that the integrals over \mathcal{T}^- and \mathcal{T}^+ of the limiting likelihoods multiplied by the priors are equal to 0. The dominated convergence theorem (DCT) implies that the integrals over \mathcal{T}^- and \mathcal{T}^+ are continuous functions and, hence, the CMT implies that the limits of the integrals are equal to the integrals of the limits. Hence, the limiting integrals over \mathcal{T}^- and \mathcal{T}^+ of the likelihoods multiplied by the priors are both equal to 0. (5) This final step shows that, whenever the true effect is in the prior's support, the denominator of equation (11) is bounded away from 0, which, by the CMT and law of total probability, then implies that the posterior probability of \mathcal{T}^* converges a.s. to 1.

The use of a normal density and randomization play crucial roles for Bayesian unbiasedness and consistency. For Bayesian unbiasedness, the normal density implies that the value of τ_h that maximizes the likelihood is equal to the difference-in-means. Hence, with a uniform prior, the MAP is equal, in expectation, to $E[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]$. Randomization then implies that this expected difference-in-means is equal to τ , thereby establishing that the expected MAP is also equal to τ .

For Bayesian consistency, the normal density's monotonically decreasing tails imply that the density of the test-statistic in equation (7), centered by a value of τ_h not equal to the difference-in-means' expectation, tends to 0. When this same test-statistic is centered by the difference-in-means' true expectation, the test-statistic converges in distribution to standard normal (scaled by the square root of a quantity greater than 0 and less than or equal to 1 since the test-statistic uses a conservative variance estimator of the difference-in-means' true variance). Since the likelihood in equation (9) is standard normal, and the standard normal density cannot be equal to 0 for any finite input, the likelihood evaluated at the test-statistic centered by the difference-in-means' true expectation is bounded away from 0 in the limit. Hence, so long as the difference-in-means' expected value is in the prior's support, the denominator of Bayes' rule is bounded away from 0 in the limit. Randomization then implies that the difference-in-means' expectation is equal to the true causal effect, τ . Without random assignment, the difference-in-means could converge to the true effect plus or minus some bias term. Consequently, the absence of random assignment would imply that the test-statistic in equation (7) centered by the true effect would diverge to $\pm\infty$ and the posterior distribution would concentrate around the τ_h equal to the true τ plus or minus the bias term.

To illustrate Theorem 1, consider the following simulation exercise. The experimental population consists of $N = 20$ units whose true control potential outcomes are randomly drawn from the distribution $\mathcal{N}(50, 100)$, where $\mathcal{N}(\dots, \dots)$ is the Normal distribution. I fix these control potential outcomes at their realized values and then construct the treated potential outcomes for each unit as its control potential outcome plus a constant effect of 10. I suppose CRA of these $N = 20$ units in which $n_1 = 10$ and $n_0 = 10$. For expository purposes, I define a prior that assigns equal probability mass of 0.5 to only two weak causal hypotheses, $\tau_h = 5$ and $\tau_h = \tau = 10$, where, for any ε greater than 0 and less than 5, $\tau_h = 5$ belongs to the set $\mathcal{T}_h^- := \{\tau_h : \tau - \tau_h > \varepsilon\}$ and $\tau_h = \tau = 10$ belongs to the set $\mathcal{T}_h^* := \{\tau_h : |\tau - \tau_h| \leq \varepsilon\}$.

Drawing on the asymptotic regime of Brewer [55], which satisfies the regularity conditions in Assumption 3, I let the sequence of finite populations increase by copying the initial finite population of $N = 20$ units an increasing number of times. That is, I let $m = 1, 2, \dots$, and then construct each finite experimental population of increasing sizes by (1) copying the original population of N units $m - 1$ times; (2) within each of the m populations, following the original CRA process with $n_1 = 10$ and $n_0 = 10$; and (3) collecting the m populations into a single population with mN total units, mn_1 treated units and mn_0 control units. For each element in the sequence of finite populations, indexed by m , I randomly draw 1,000 assignments from the set Ω_m , and, for each draw, calculate two versions of the test-statistic in equation (7) – one of which is centered by the false $\tau_h = 5$ and the other by the true $\tau_h = 10$. To generate a posterior probability for the false $\tau_h = 5$ and the true $\tau_h = 10$ under each draw from Ω_m , I input each test-statistic to the standard normal density, multiply each density by its respective prior, and then divide by the total probability of the evidence.

Figure 1 plots both the respective likelihoods and posterior probabilities for $\tau_h = 5$ and $\tau_h = 10$ over all 1,000 draws of assignments from Ω_m . The first panel is the original population in which $m = 1$ (i.e., $N = 20$). The remaining three panels show the likelihoods and posterior probabilities for each causal hypothesis over all 1,000 draws from Ω_m with $m = 5$ (i.e., $N = 100$), $m = 10$ (i.e., $N = 200$), and $m = 50$ (i.e., $N = 1,000$). The y-axis of Figure 1 is the simulation-based approximation to the randomization probability based on the probability distribution over each Ω_m . By contrast, the x-axis of Figure 1 refers to the standard normal density in the first row and the posterior probability in the second row. Figure 1 therefore shows the distributions of the likelihoods and posterior probabilities for each causal hypothesis over repeated draws from Ω_m .

Since $\tau_h = 5$ is too small relative to the true effect of $\tau = 10$, the randomization distribution of the test-statistic in equation (7) diverges in probability to $-\infty$. Thus, as Figure 1 shows, when $\tau_h : |\tau - \tau_h| > \varepsilon$, the probability density that the standard normal distribution assigns to the standardized test-statistic in

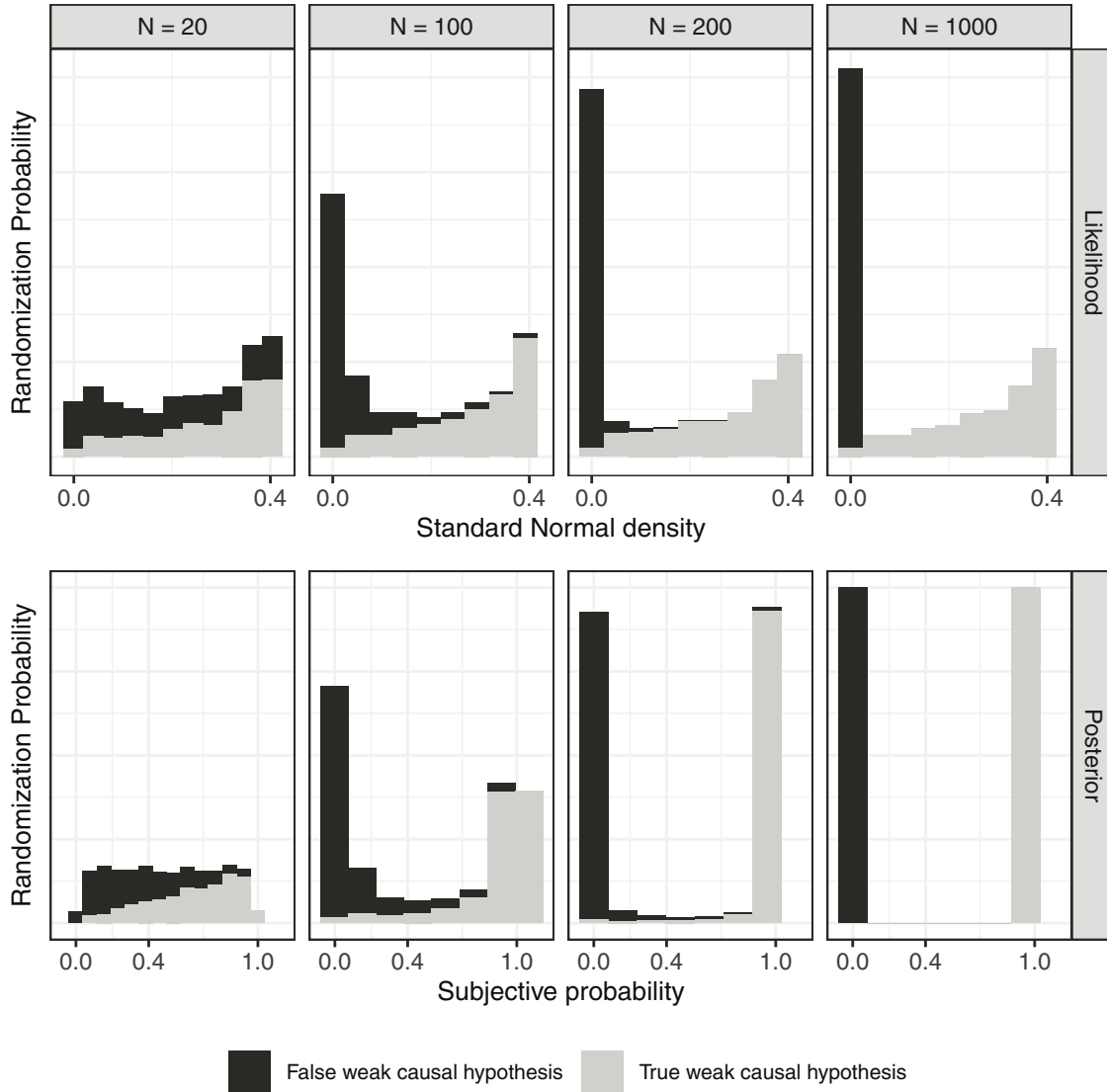


Figure 1: Distributions of likelihoods and posterior probabilities over repeated randomizations.

equation (7) tends to 0 as $N \rightarrow \infty$. For $\tau_h : \tau_h = \tau$, the probability density that the standard normal distribution assigns to the standardized test-statistic takes on values strictly greater than 0 and less than or equal to $1/\sqrt{2\pi} \approx 0.4$, which is the maximum probability density of the standard normal distribution. Since the likelihood of the false weak causal hypothesis ($\tau_h = 5$) tends to 0 asymptotically, so does the product of its prior and likelihood. Normalizing by the total probability of the evidence thereby implies that the posterior distribution concentrates increasingly around the true weak causal hypothesis ($\tau_h = 10$) as $N \rightarrow \infty$.

5.2 Sharp causal effects

In contrast to constructing a reference density for a weak causal hypothesis, an alternative is to construct a reference PMF via a sharp effect consistent with a hypothesis about only the average effect. The imputation of a sharp causal effect consistent with a hypothetical weak effect enables exact null hypothesis significance tests of an average effect. However, Bayesian inference via an exact likelihood constructed via a sharp effect does not necessarily satisfy the properties of Bayesian unbiasedness and consistency. As I now

demonstrate, an exact likelihood takes us a long way in proving a version of Theorem 1, but does not take us quite far enough. In what follows, I explain why and show that this problem can be alleviated by using a standard normal likelihood with a variance parameter that is eliminated based on a direct calculation of observed and imputed potential outcomes.

When outcomes are binary, the imputation of a sharp effect consistent with a hypothetical weak effect often proceeds by imputing potential outcomes under the worst-case (i.e., highest variance) allocation of potential outcomes consistent with a hypothetical average effect [65–68]. However, this approach is tractable only for binary outcomes. For the difficulty of this approach with ordinal outcomes, see Lu et al. [69]. When outcomes are not binary, implementing this approach often proceeds under a specific model of effects [27, Chapter 5] – usually that of a constant effect in which $\tau_i = \tau_h$ for all $i = 1, \dots, N$ units. For example, Middleton and Aronow [53, p. 50–53] propose eliminating the variance nuisance parameter by, first, imputing missing potential outcomes via a homogeneous individual effect for all units that is consistent with a given weak null hypothesis and then directly calculating the implied variance of the difference-in-means. Samii and Aronow [70] show that this variance estimator is equivalent (up to a scaling factor) to the widely used homoscedastic variance estimator of OLS.

The test-statistic follows this approach of using a model of a homogeneous individual effect consistent with a hypothetical average effect. Analogous to the test-statistic in equation (7), which centers the difference-in-means by a hypothetical average effect, the random difference-in-means calculated on potential outcomes in which treated potential outcomes are adjusted by a hypothetical effect, τ_h , is as follows:

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) = \left(\frac{1}{n_1} \right) \mathbf{Z}^\top (\mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) - \left(\frac{1}{n_0} \right) (\mathbf{1} - \mathbf{Z})^\top (\mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}), \quad (12)$$

where, under CRA, n_1 , and n_0 are fixed over all $\mathbf{z} \in \Omega$. As in equation (7), the expected value of $\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$ is equal to 0 when $\tau_h = \tau$ regardless of whether individual effects are homogeneous.

The unknown PMF of the difference-in-means with treated potential outcomes adjusted by a sharp hypothetical effect of τ_h is given as follows:

$$k(t) = \sum_{\mathbf{z} \in \Omega} \mathbb{I}\{\hat{\tau}(\mathbf{z}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}) = t\} \Pr(\mathbf{Z} = \mathbf{z}). \quad (13)$$

The exact PMF in equation (13) is unknown since the researcher can observe potential outcomes under only one assignment. Instead, under whichever assignment is realized, the researcher constructs a PMF implied by the supposition that τ_h is true. Since the observed vector of adjusted outcomes would be fixed over all assignments if it were the case that $\tau_i = \tau_h$ for all $i = 1, \dots, N$, the PMF implied by the supposition that τ_h is the true homogeneous effect, given a single realization of data, is

$$\hat{k}(t) = \sum_{\mathbf{w} \in \Omega} \mathbb{I}\{\hat{\tau}(\mathbf{w}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}) = t\} \Pr(\mathbf{W} = \mathbf{w}). \quad (14)$$

Analogous to a reference distribution for the calculation of p -values, $\hat{k}(t)$ is a reference PMF in which a researcher proceeds as if τ_h were the true homogeneous effect, and, hence, $\mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}$ were the fixed, adjusted response over all assignments. Unlike the unknown PMF in equation (13), this reference PMF, $\hat{k}(t)$, is random in that, whenever τ_h is false, $\hat{k}(t)$ varies with \mathbf{Z} .

To see why Bayesian inference via a likelihood derived from equation (14) does not necessarily satisfy Bayesian unbiasedness and consistency, first consider the centered difference-in-means in equation (12). Lemma 3 in Appendix A shows that this test-statistic converges a.s. to $\tau - \tau_h$. That is,

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) \xrightarrow{\text{a.s.}} \tau - \tau_h.$$

Lemmas 4 and 5 likewise show that

$$\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}) \xrightarrow{\text{a.s.}} 0$$

for all sequences of \mathbf{Z} .

Given these lemmas, Proposition 2 shows that, for all $\tau_h : |\tau - \tau_h| > \varepsilon$, the exact reference PMF in equation (14) evaluated at the centered difference-in-means in equation (12) converges a.s. to 0.

Proposition 2. *Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. It follows that*

$$\hat{k}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})) \xrightarrow{\text{a.s.}} 0 \quad \text{for all } \tau_h : |\tau - \tau_h| > \varepsilon.$$

The proof of Proposition 2 draws upon the exponential tail bound from Bloniarz et al. [71, Lemma S1], [following 72], which has been elaborated upon by Wu and Ding [60, Lemma A2]. This exponential tail bound yields a continuous upper bound for the probability that the absolute value of $\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z})$ is equal to any $t \geq 0$. The availability of this continuous upper bound sidesteps the task of directly showing that the discrete reference PMF evaluated at the test-statistic in equation (12) converges a.s. to 0 for all $\tau_h : |\tau - \tau_h| > \varepsilon$. Instead, the proof uses the CMT to show that the continuous upper bound of the reference PMF converges a.s. to 0 for all $\tau_h : |\tau - \tau_h| > \varepsilon$. Since the reference PMF's upper bound converges a.s. to 0 for all $\tau_h : |\tau - \tau_h| > \varepsilon$, it follows that the same holds for the actual reference PMF.

Proposition 2 shows that the likelihood of any $\tau_h : |\tau - \tau_h| > \varepsilon$ converges a.s. to 0. However, to establish Bayesian consistency, it remains to be shown that the likelihood of the true τ_h , i.e., $\tau_h : \tau_h = \tau$, is bounded away from 0 in the limit. Both the test-statistic, $\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$, and the reference test-statistic, $\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$ converge a.s. to 0. Although both converge a.s. to the same value, the discreteness of the reference PMF means that the CMT does not immediately imply that the reference PMF is bounded away from 0 when evaluated at the limit of the difference-in-means in equation (12). In other words, distributions of both $\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$ and $\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$ will ultimately lie within a narrow interval around 0. However, the theory developed thus far is insufficient to rule out the possibility that the reference distribution of $\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$ contains no values that are exactly equal to $\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})$.

One way to alleviate this issue is by using a standard normal density in place of the reference PMF in equation (14). The causal inference literature often uses normal approximations to sharp null distributions [as in, e.g., 73] via appeals to the finite population CLT and associated theory. However, as in the previous section, the finite population CLT and associated theory do not imply that a normal density approximates the PMF of the difference-in-means under a sharp null hypothesis. Nevertheless, the following Corollaries to Proposition 1 and Theorem 1 show that both Bayesian unbiasedness and consistency hold.

For inference via a sharp effect consistent with a hypothetical weak effect, the density function is standard normal, as in equation (9). However, the standardized test-statistic is expressed as follows:

$$\frac{\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})}{\sqrt{\text{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}}. \quad (15)$$

The numerator of the test-statistic in (15) is equivalent to the numerator of (7); hence, they both converge a.s. to the same limit of $\tau - \tau_h$. (See Lemma 1 in Appendix A.) However, their variances do not necessarily converge a.s. to the same limit.

A hypothesis about the average effect under the (potentially false) model of a homogeneous individual effect is strong enough to imply both the expected value and variance of the difference-in-means. Hence, unlike in (7), the variance in equation (15), given some realization of data, is taken over all $\mathbf{w} \in \Omega$ holding adjusted outcomes fixed at $\mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}$. That is, this variance refers to the difference-in-means' variance under the supposition that τ_h is the true constant effect for all units. This supposition may be false, in which case the variance in the denominator of equation (15) is a random quantity that varies with \mathbf{Z} .

Corollary 1. (Bayesian unbiasedness) *Suppose Assumptions 1 and 2, and let the prior density of τ_h be uniform. It follows that the MAP is equal, in expectation, to the average effect, i.e.,*

$$\mathbb{E} \left[\underset{\tau_h \in \mathcal{T}_h}{\operatorname{argmax}} \frac{\hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\text{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h)}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\text{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h) d\tau_h} \right] = \tau. \quad (16)$$

The logic of Corollary 1 is the same as that of Proposition 1. For any finite plug-in greater than 0 for the variance parameter, the value of τ_h that maximizes the standard normal density will be whichever value is equal to the observed difference-in-means. Hence, the procedure for eliminating the variance nuisance parameter via a sharp effect is irrelevant and the same logic of Proposition 1 holds.

Corollary 2. (Bayesian consistency) *Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. If $\tau_h : \tau_h = \tau$ is in the support of the prior distribution, then, as the size of the experiment increases indefinitely, the posterior probability of \mathcal{T}_h^* converges a.s. to 1:*

$$\frac{\int_{\tau_h \in \mathcal{T}_h^*} \hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\text{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h) d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\text{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h) d\tau_h} \xrightarrow{\text{a.s.}} 1. \quad (17)$$

The proof of Corollary 2 is identical to the proof of Theorem 1 except that the variance in equation (15) does not necessarily converge to the same constant as that of the variance estimator in equation (7).

The proofs of Corollaries 1 and 2 do not assume a constant effect for all units. Imputing missing potential outcomes via a sharp effect consistent with a weak causal hypothesis represents only another way of eliminating the variance nuisance parameter. Bayesian unbiasedness and consistency is with respect to the average effect. For a comparison of performances between these two approaches to eliminating the variance parameter, see Ding [74].

6 Conclusion

This article has established Bayesian analogs of unbiasedness and consistency for the average effect in a randomization-based, finite population setting. An important feature of the randomization-based, Bayesian procedure is that it avoids imposing a model on potential outcomes by instead using a standard normal model for the standardized difference-in-means. Crucially, the randomization-based, Bayesian procedure yields the aforementioned properties of Bayesian unbiasedness and consistency regardless of how well the standard normal density approximates the PMF of the standardized difference-in-means.

The aim of this article has been to develop credible Bayesian inference that is justified by the experimental design. In so doing, this article's procedure improves upon the generality of existing, randomization-based methods of Bayesian inference that suppose potential outcomes are binary, ordinal, or, more generally, discrete and bounded. In addition, this article's procedure avoids assuming probability models of potential outcomes, such as in dominant, model-based methods of Bayesian causal inference. A benefit of this article's method relative to existing randomization- and model-based alternatives is that it produces credible inference under a prior directly on the causal target of interest and a single likelihood function that can be used out of the box for different data types. Nevertheless, open questions for future research include how this article's procedure performs relative to existing randomization-based methods for data that are discrete and bounded and model-based methods when they are evaluated from a randomization-based, finite population perspective.

Acknowledgements: For their valuable feedback and advice, I thank Peng Ding (the associate editor), three anonymous reviewers, Don Green, Macartan Humphreys, Jake Bowers, Naoki Egami, Fredrik Sävje, Winston Lin, Peter Cohen, Ben Hansen, Anna Wilke, Georgiy Syunyaev, Nicole Pashley, Luke Miratrix, Cyrus Samii, P. M. Aronow and José Zubizarreta, as well as audiences at Wissenschaftszentrum Berlin für Sozialforschung (WZB), the NYU Quantitative Methods Seminar and PolMeth XXXIX.

Funding information: The authors state no funding involved.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

- [1] Angrist JD, Pischke JS. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *J Econ Perspectives*. 2010;24(2):3–30.
- [2] Pashley NE, Basse GW, Miratrix LW. Conditional as-if analyses in randomized experiments. *J Causal Inference*. 2021;9(1), 264–84.
- [3] Berk RA, Freedman DA. Statistical assumptions as empirical commitments. In: Blomberg TG, Cohen S, editors. *Punishment and social control: essays in honor of Sheldon L. Messinger*. 2nd ed. New York, NY: Aldine De Gruyter; 2003. p. 235–54.
- [4] Fisher RA. *The design of experiments*. Edinburgh, SCT: Oliver and Boyd; 1935.
- [5] King G, Tomz M, Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. *Amer J Polit Sci*. 2000;44(2):341–55.
- [6] Tomz M, Wittenberg J, King G. Clarify: software for interpreting and presenting statistical results. *J Stat Softw*. 2003;8(1):1–30.
- [7] Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press; 2006.
- [8] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2–21.
- [9] Ding P, Miratrix L. Model-free causal inference of binary experimental data. *Scand J Stat*. 2019;46(1):200–14.
- [10] Chiba Y. Bayesian inference of causal effects for an ordinal outcome in randomized trials. *J Causal Inference*. 2018;6(2):1–12.
- [11] Keele L, Quinn KM. Bayesian sensitivity analysis for causal effects from 2×2 tables in the presence of unmeasured confounding with application to presidential campaign visits. *Ann Appl Stat*. 2017;11(4):1974–97.
- [12] Humphreys M, Jacobs AM. Mixing methods: a Bayesian approach. *Amer Polit Sci Rev*. 2015;109(04):653–73.
- [13] Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat*. 1978;6(1):34–58.
- [14] Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Stat*. 1997;25(1):305–27.
- [15] Imbens GW, Rubin DB. *Causal inference for statistics, social, and biomedical sciences: an introduction*. New York, NY: Cambridge University Press; 2015.
- [16] Zhang JL, Rubin DB, Mealli F. Likelihood-based analysis of causal effects of job-training programs using principal stratification. *J Amer Stat Assoc*. 2009;104(485):166–76.
- [17] Ding P, Li F. Causal inference: a missing data perspective. *Stat Sci*. 2018;33(2):214–37.
- [18] Li F, Ding P, Mealli F. Bayesian Causal Inference: A Critical Review. *Philosophical Transactions of the Royal Society A*. 2023;381(2247):20220153.
- [19] Freedman DA. *Statistical models: theory and practice*. New York, NY: Cambridge University Press; 2009.
- [20] Gerber AS, Green DP. *Field experiments: design, analysis, and interpretation*. New York, NY: W.W. Norton; 2012.
- [21] Cox DR. *Planning of experiments*. New York, NY: Wiley; 1958.
- [22] Rubin DB. Comment on “Randomization analysis of experimental data in the Fisher randomization test” by Basu, D. *J Amer Stat Assoc*. 1980;75(371):591–3.
- [23] Rubin DB. Which ifs have causal answers? (Comment on “statistics and causal inference” by Paul W. Holland). *J Amer Stat Assoc*. 1986;81(396):961–2.
- [24] Rosenbaum PR. *Observation and experiment: an introduction to causal inference*. Cambridge, MA: Harvard University Press; 2017.
- [25] Neyman J. Sur les applications de la théorie des probabilités aux expériences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*. 1923;10:1–51.
- [26] Aronow P, Miller BT. *Foundations of agnostic statistics*. New York, NY: Cambridge University Press; 2019.
- [27] Rosenbaum PR. *Observational studies*. 2nd ed. New York, NY: Springer; 2002.
- [28] Rosenbaum PR. *Design of observational studies*. New York, NY: Springer; 2010.
- [29] Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Amer Stat Assoc*. 2005;100(469):322–31.
- [30] Copas JB. Randomization models for the matched and unmatched 2×2 tables. *Biometrika*. 1973;60(3):467–76.
- [31] Rubin DB. Bayesian inference for causality: the importance of randomization. In: Goldfield ED, editor. *American Statistical Association: 1975 Proceedings of the Social Statistics Section*. Washington, D.C.: American Statistical Association; 1976. p. 233–9.
- [32] Richardson TS, Evans RJ, Robins JM. Transparent parametrizations of models for potential outcomes. In: Bernardo JM, Bayarri MJ, Berger JOA, Dawid P, Heckerman D, Smith AFM, et al., editors. *Bayesian statistics*. vol. 9. New York, NY: Oxford University Press; 2011. p. 569–610.

- [33] Rubin DB. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann Stat.* 1984;12(4):1151–72.
- [34] Dasgupta T, Pillai NS, Rubin DB. Causal inference from 2^K factorial designs by using potential outcomes. *J R Stat Soc Ser B (Stat Methodol)*. 2015;77(4):727–53.
- [35] Ding P, Dasgupta T. A potential tale of two-by-two tables from completely randomized experiments. *J Amer Stat Assoc.* 2016;111(513):157–68.
- [36] Li X, Ding P. General forms of finite population central limit theorems with applications to causal inference. *J Amer Stat Assoc.* 2017;112(520):1759–69.
- [37] Berry AC. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans Amer Math Soc.* 1941;49(1):122–36.
- [38] Esseen CG. On the Liapunoff limit of error in the theory of probability. *Arkiv För Matematik, Astronomi Och Fysik.* 1942;A28:1–19.
- [39] Bikelis A. The estimation of the remainder term in the central limit theorem for samples taken from finite sets. *Studia Scientiarum Mathematicarum Hungarica.* 1969;4:345–54.
- [40] Höglund T. Sampling from a finite population. A remainder term estimate. *Scand J Stat.* 1978;5(1):69–71.
- [41] Wang Y, Li X. Rerandomization with diminishing covariate imbalance and diverging number of covariates. *Ann Stat.* 2022;50(6):3439–65.
- [42] Shi L, Ding P. Berry-Esseen bounds for design-based causal inference with possibly diverging treatment levels and varying group sizes; 2022. Working Paper. <https://arxiv.org/pdf/2209.12345.pdf>.
- [43] Boos DD. A Converse to Scheffé’s theorem. *Ann Stat.* 1985;13(1):423–7.
- [44] Sweeting TJ. On a converse to Scheffé’s theorem. *Ann Stat.* 1986;14(3):1252–6.
- [45] Lin W. Agnostic notes on regression adjustments to experimental data: reexamining Freedman’s critique. *Ann Appl Stat.* 2013;7(1):295–318.
- [46] Berger JO, Liseo B, Wolpert RL. Integrated likelihood methods for eliminating nuisance parameters. *Stat Sci.* 1999;14(1):1–22.
- [47] Liseo B. The elimination of nuisance parameters. In: Dey DK, Rao CR, editors. *Bayesian thinking: modeling and computation*. vol. 25 of *handbook of statistics*. Amsterdam, NL: Elsevier; 2005. p. 193–219.
- [48] Kim JY. Limited information likelihood and Bayesian analysis. *J Econometric.* 2002;107(1–2):175–93.
- [49] Kwan YK. Asymptotic Bayesian analysis based on a limited information estimator. *J Econometric.* 1999;88(1):99–121.
- [50] Boos DD, Monahan JF. Bootstrap methods using prior information. *Biometrika.* 1986;73(1):77–83.
- [51] Greco L, Racugno W, Ventura L. Robust likelihood functions in Bayesian inference. *J Stat Plan Inference.* 2008;138(5):1258–70.
- [52] Li X, Ding P, Lin Q, Yang D, Liu JS. Randomization inference for peer effects. *J Amer Stat Assoc.* 2019;114(528):1651–64.
- [53] Middleton JA, Aronow P. Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Stat Politic Policy.* 2015;6(1–2):39–75.
- [54] Bowers J, Leavitt T. Causality and design-based inference. In: Curini L, Franzese R, editors. *The SAGE handbook of research methods in political science and international relations*. vol. 2. Thousand Oaks, CA: SAGE Publications; 2020. p. 769–804.
- [55] Brewer KRW. A class of Robust sampling designs for large-scale surveys. *J Amer Stat Assoc.* 1979;74(368):911–5.
- [56] Freedman DA. On regression adjustments in experiments with several treatments. *Ann Appl Stat.* 2008;2(1):176–96.
- [57] Cohen PL, Fogarty CB. Gaussian pre pivoting for finite population causal inference. *J R Stat Soc Ser B (Stat Methodol)*. 2022;84(2):295–320.
- [58] Sävje F, Aronow P, Hudgens MG. Average treatment effects in the presence of unknown interference. *Ann Stat.* 2021;49(2):673–701.
- [59] Delevoye A, Sävje F. Consistency of the Horvitz-Thompson estimator under general sampling and experimental designs. *J Stat Plan Inference.* 2020;207:190–7.
- [60] Wu J, Ding P. Randomization tests for weak null hypotheses. *J Amer Stat Assoc.* 2021;116(536):1898–913.
- [61] Aronow P, Green DP, Lee DKK. Sharp bounds on the variance in randomized experiments. *Ann Stat.* 2014;42(3):850–71.
- [62] Imai K. Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Stat Med.* 2008;27(24):4857–73.
- [63] Fogarty CB. On mitigating the analytical limitations of finely stratified experiments. *J R Stat Soc Ser B (Stat Methodol)*. 2018;80(5):1035–56.
- [64] Pashley NE, Miratrix LW. Insights on variance estimation for blocked and matched pairs designs. *J Educat Behav Stat.* 2021;46(3):271–96.
- [65] Fogarty CB, Mikkelsen ME, Gaieski DF, Small DS. Discrete optimization for interpretable study populations and randomization inference in an observational study of severe sepsis mortality. *J Amer Stat Assoc.* 2016;111(514):447–58.
- [66] Fogarty CB, Shi P, Mikkelsen ME, Small DS. Randomization inference and sensitivity analysis for composite null hypotheses with binary outcomes in matched observational studies. *J Amer Stat Assoc.* 2017;112(517):321–31.
- [67] Li X, Ding P. Exact confidence intervals for the average causal effect on a binary outcome. *Stat Med.* 2016;35(13):2296.
- [68] Rigdon J, Hudgens MG. Randomization inference for treatment effects on a binary outcome. *Stat Med.* 2015;34(6):924–35.

- [69] Lu J, Ding P, Dasgupta T. Treatment effects on ordinal outcomes: causal estimands and sharp bounds. *J Educat Behav Stat.* 2018;43(5):540–67.
- [70] Samii C, Aronow P. On equivalencies between design-based and regression-based variance estimators for randomized experiments. *Stat Probabil Lett.* 2012;82(2):365–70.
- [71] Bloniarz A, Liu H, Zhang CH, Sekhon JS, Yu B. Lasso adjustments of treatment effect estimates in randomized experiments. *Proc Nat Acad Sci USA.* 2016;113(27):7383–90.
- [72] Massart P. Rates of convergence in the central limit theorem for empirical processes. In: Fernique X, Heinkel B, Marcus MB, Meyer PA, editors. *Geometrical and statistical aspects of probability in Banach spaces. Lecture Notes in Mathematics.* Berlin, Germany: Springer-Verlag; 1986. p. 73–109.
- [73] Hansen BB, Bowers J. Covariate balance in simple, stratified and clustered comparative studies. *Stat Sci.* 2008;23(2):219–36.
- [74] Ding P. A paradox from randomization-based causal inference. *Stat Sci.* 2017;32(3):331–45.

Appendix

A Formal proofs

A.1 Proof of Proposition 1

Proposition 1. (Bayesian unbiasedness). *Suppose Assumptions 1 and 2, and let the prior density of τ_h be uniform. It follows that the MAP is equal, in expectation, to the average effect, i.e.,*

$$\mathbb{E} \left[\operatorname{argmax}_{\tau_h \in \mathcal{T}_h} \frac{\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))) - \tau_h) / \sqrt{\widehat{\operatorname{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]} r(\tau_h)}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))) - \tau_h) / \sqrt{\widehat{\operatorname{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]} r(\tau_h) d\tau_h} \right] = \tau.$$

Proof. The standard normal density is monotonically decreasing in Euclidean distance from its mean of 0. Hence, conditional on a realization of data, $\{\mathbf{z}, \mathbf{y}(\mathbf{z})\}$, the value of τ_h that maximizes the standard normal likelihood is whichever value is equal to the observed difference-in-means, i.e.,

$$\operatorname{argmax}_{\tau_h \in \mathcal{T}_h} \hat{\phi} \left(\frac{\hat{\tau}(\mathbf{z}, \mathbf{y}(\mathbf{z})) - \tau_h}{\sqrt{\widehat{\operatorname{Var}}[\hat{\tau}(\mathbf{z}, \mathbf{y}(\mathbf{z}))]}} \right) = \hat{\tau}(\mathbf{z}, \mathbf{y}(\mathbf{z})).$$

Since SUTVA in Assumption 1 and CRA in Assumption 2 imply that $\mathbb{E}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \tau$, it follows that the expectation of the maximum likelihood estimator over the set of assignments is unbiased for τ , i.e.,

$$\mathbb{E} \left[\operatorname{argmax}_{\tau_h \in \mathcal{T}_h} \hat{\phi} \left(\frac{\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h}{\sqrt{\widehat{\operatorname{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \right) \right] = \mathbb{E}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \tau. \quad (\text{A1})$$

With a uniform prior, the value of $\tau_h \in \mathcal{T}_h$ that maximizes the posterior distribution is the maximum likelihood estimate. Therefore, equation (A1) implies that the MAP is unbiased for τ . \square

A.2 Lemma 1

Lemma 1. *Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. Then*

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h \xrightarrow{\text{a.s.}} \tau - \tau_h.$$

Proof. Under Assumptions 1–3, Lemma A3 of Wu and Ding [60] implies that

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) \xrightarrow{\text{a.s.}} \tau.$$

The CMT further implies that

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h \xrightarrow{\text{a.s.}} \tau - \tau_h,$$

where τ_h is a fixed constant over all $N \in \mathbb{N}_{\geq 4}$, which concludes the proof. \square

A.3 Lemma 2

Lemma 2. *Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. It follows that*

$$\lim_{N \rightarrow \infty} \text{Var}[\sqrt{N}\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \frac{1}{v}S_{1,\infty}^2 + \frac{1}{(1-v)}S_{0,\infty}^2 - S_{10,\infty}^2 > 0$$

and that

$$\widehat{\text{Var}}[\sqrt{N}\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] \xrightarrow{\text{a.s.}} \frac{1}{v}S_{1,\infty}^2 + \frac{1}{(1-v)}S_{0,\infty}^2 > 0.$$

Proof. First, note that, under SUTVA and CRA in Assumptions 1 and 2,

$$\text{Var}[\sqrt{N}\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \frac{N}{n_1}S_1^2 + \frac{N}{n_0}S_0^2 - S_{10}^2,$$

which, by the asymptotic regularity conditions in Assumption 3 and the CMT, limits to

$$\text{Var}[\sqrt{N}\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = \frac{1}{v}S_{1,\infty}^2 + \frac{1}{(1-v)}S_{0,\infty}^2 - S_{10,\infty}^2.$$

Second, note that, under SUTVA and CRA in Assumptions 1 and 2,

$$\widehat{\text{Var}}[\sqrt{N}\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] = N \left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_0^2}{n_0} \right).$$

Under the asymptotic regularity conditions in Assumption 3, Lemma A3 of Wu and Ding [60] and the CMT imply that

$$\widehat{\text{Var}}[\sqrt{N}\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))] \xrightarrow{\text{a.s.}} \frac{1}{v}S_{1,\infty}^2 + \frac{1}{(1-v)}S_{0,\infty}^2 > 0. \quad (\text{A2})$$

□

A.4 Proof of Theorem 1

Theorem 2. (Bayesian consistency). Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. If $\tau_h : \tau_h = \tau$ is in the support of the prior distribution, then, as the size of the experiment increases indefinitely, the posterior probability of \mathcal{T}_h^* converges a.s. to 1:

$$\frac{\int_{\tau_h \in \mathcal{T}_h^*} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h) / \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}) r(\tau_h) d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h) / \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}) r(\tau_h) d\tau_h} \xrightarrow{\text{a.s.}} 1.$$

Proof. First, rewrite the test-statistic in equation (7) as follows:

$$\frac{\sqrt{N}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)}{\sqrt{N}\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}},$$

which, under SUTVA in Assumption 1, can be equivalently expressed as follows:

$$\frac{\sqrt{N}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau)}{\sqrt{N}\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]} \frac{\sqrt{N\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}}{\sqrt{N}\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]} + \frac{\sqrt{N}(\tau - \tau_h)}{\sqrt{N}\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]} \frac{\sqrt{N\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}}{\sqrt{N}\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}. \quad (\text{A3})$$

The finite population CLT [36] implies that

$$\left(\frac{\sqrt{N}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau)}{\sqrt{N}\sqrt{\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \right) \xrightarrow{d} \Phi.$$

In addition, under CRA in Assumption 2, Lemma 2 and the CMT imply that

$$\left(\frac{\sqrt{N} \sqrt{\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}}{\sqrt{N} \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \right) \xrightarrow{\text{a.s.}} 1/\sqrt{c},$$

where $c \geq 1$.

Therefore, Slutsky's theorem implies that the first term of equation (A3) converges in distribution to standard normal scaled by $1/\sqrt{c}$, i.e.,

$$\left(\frac{\sqrt{N}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau)}{\sqrt{N} \sqrt{\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \right) \left(\frac{\sqrt{N} \sqrt{\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}}{\sqrt{N} \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \right) \xrightarrow{d} \Phi/\sqrt{c}. \quad (\text{A4})$$

For the second term of equation (A3), first note that, under CRA in Assumption 2, Lemma 2 and the CMT likewise imply that

$$\frac{1}{\sqrt{N} \sqrt{\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \frac{\sqrt{N} \sqrt{\text{Var}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}}{\sqrt{N} \sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \xrightarrow{\text{a.s.}} \frac{1}{\sqrt{v \left(\frac{1}{v} S_{1,\infty}^2 + \frac{1}{(1-v)} S_{0,\infty}^2 - S_{10,\infty}^2 \right)}}.$$

Then, since $\sqrt{N}(\tau - \tau_h)$ diverges to either ∞ or $-\infty$ for all $\tau_h : |\tau - \tau_h| > \varepsilon$, the second term of equation (A3) diverges a.s. to $\pm\infty$ for all $\tau_h : |\tau - \tau_h| > \varepsilon$. Therefore, recalling that the first term of equation (A3) converges in distribution to Φ/\sqrt{c} , equation (A3) diverges a.s. to $\pm\infty$ for all $\tau_h : |\tau - \tau_h| > \varepsilon$. Given this a.s. divergence of equation (7) and that the standard normal density is strictly monotonically decreasing in distance from its mean, it follows that, for all $\tau_h : |\tau - \tau_h| > \varepsilon$, the standard normal density evaluated at equation (7) converges a.s. to the standard normal density's lower bound of 0.

It then follows from the CMT that

$$\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h) \xrightarrow{\text{a.s.}} 0 \quad \text{for all } \tau_h : |\tau - \tau_h| > \varepsilon, \quad (\text{A5})$$

where the prior density for all $\tau_h \in \mathcal{T}_h$ is fixed over all $N \in \mathbb{N}_{\geq 4}$.

To show that both

$$\int_{\tau_h \in \mathcal{T}_h^-} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h) d\tau_h \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad (\text{A6})$$

$$\int_{\tau_h \in \mathcal{T}_h^+} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h) d\tau_h \xrightarrow{\text{a.s.}} 0, \quad (\text{A7})$$

it suffices to show that

$$\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)$$

is (1) pointwise convergent to an integrable function and (2) dominated by an integrable function.

Note that (1) is implied by the a.s. convergence of

$$\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)$$

in equation (A5). For (2), note that the prior density dominates

$$\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)$$

for all $N \in \mathbb{N}_{\geq 4}$ since

$$\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})$$

is bounded between 0 and 1.

Taken together, (1) and (2) justify an application of the DCT, which implies that the integrals over \mathcal{T}_h^- and \mathcal{T}_h^+ are continuous functions. The integrals over \mathcal{T}_h^- and \mathcal{T}_h^+ evaluated at the limit of

$$\hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)$$

are both equal to 0. Hence, both (A6) and (A7) then follow from the CMT.

Then, to complete the proof that

$$\frac{\int_{\tau_h \in \mathcal{T}_h^+} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h} \xrightarrow{\text{a.s.}} 1, \quad (\text{A8})$$

note, first, that, as $N \rightarrow \infty$, the denominator of equation (A8) is bounded away from 0: Recall from equation (A4) that, for $\tau_h : \tau_h = \tau$,

$$\frac{(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)}{\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]}} \xrightarrow{d} \Phi/\sqrt{c} \quad (\text{A9})$$

and does not diverge a.s. to $\pm\infty$. By the supposition that $\tau_h : \tau_h = \tau$ is in the prior density's support, i.e., $r(\tau_h) > 0$ for $\tau_h : \tau_h = \tau$, and since the standard normal density is strictly positive for any finite input, it follows from the CMT and DCT that the denominator of equation (A8) is a.s. bounded away from 0 as $N \rightarrow \infty$.

Note, second, that (A6), (A7), and the CMT imply that

$$\frac{\int_{\tau_h \in \mathcal{T}_h^-} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h} \xrightarrow{\text{a.s.}} 0$$

and

$$\frac{\int_{\tau_h \in \mathcal{T}_h^+} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h} \xrightarrow{\text{a.s.}} 0.$$

Therefore, the CMT and the law of total probability imply that

$$\frac{\int_{\tau_h \in \mathcal{T}_h^+} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}((\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h)/\sqrt{\widehat{\text{Var}}[\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}))]})r(\tau_h)d\tau_h} \xrightarrow{\text{a.s.}} 1,$$

which completes the proof. \square

A.5 Lemma 3

Lemma 3. Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. Then

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h \xrightarrow{\text{a.s.}} \tau - \tau_h.$$

Proof. Under SUTVA and CRA, note that

$$\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h = \frac{1}{n_1} \sum_{i=1}^N Z_i [y_i(1) - \tau_h] - \frac{1}{n_0} \sum_{i=1}^N (1 - Z_i) y_i(0) = \hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z})) - \tau_h.$$

The rest of the proof then follows from Lemma 1. \square

A.6 Lemma 4

Lemma 4. Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. It follows that the adjusted potential outcomes, $y(\mathbf{Z}) - \tau_h \mathbf{Z}$, satisfy

$$\lim_{N \rightarrow \infty} \max_z \max_i \left[y_i(z) - \tau_h z_i - \left(\sum_{i=1}^N (y_i(z) - z_i \tau_h) / N \right) \right]^2 / N = 0$$

for all sequences of \mathbf{Z} .

Proof. This proof follows the general logic of Lemma A4 from Wu and Ding [60]. Because the means of potential outcomes limit to finite values under Condition 3.2 of Assumption 3 and τ_h is a fixed constant over $N \in \mathbb{N}_{\geq 4}$, there exists a $B < \infty$ whereby, under SUTVA in Assumption 1,

$$\max_z \{ \max_i |\bar{y}_z|, \max_i |(n_1/N)\tau_h - z_i \tau_h| \} \leq B \quad \text{for all } N \in \mathbb{N}_{\geq 4}. \quad (\text{A10})$$

Also let $A = \max_z \max_i \{y_i(z) - \bar{y}_z\}^2$, which implies that

$$\max_z \max_i \{y_i(z) - \bar{y}_z\} = [\max_z \max_i \{y_i(z) - \bar{y}_z\}^2]^{1/2} \leq A^{1/2}. \quad (\text{A11})$$

In addition, note that

$$\max_i |y_i(z)| \leq \max_z \max_i |y_i(z)| \leq \max_z \max_i |y_i(z) - \bar{y}_z| + \max_z |\bar{y}_z| \leq A^{1/2} + B.$$

Since $|\bar{y}_z| \leq \max_i |y_i(z)|$ for $z = 0$ and $z = 1$, it follows that $\max_z |\bar{y}_z| \leq A^{1/2} + B$ and

$$\max_z \max_i |y_i(z) - \bar{y}_z| \leq \max_z \max_i |y_i(z)| + \max_z |\bar{y}_z| \leq 2(A^{1/2} + B), \quad (\text{A12})$$

where the last inequality follows from (A10) and (A11), and the fact that $|\bar{y}_z| \leq \max_i |y_i(z)|$ for $z = 0$ and $z = 1$.

The inequality in (A12) then implies that

$$(\max_z \max_i y_i(z) - \bar{y}_z)^2 \leq [2(A^{1/2} + B)]^2 \leq 8(A + B^2), \quad (\text{A13})$$

where the last inequality follows from the property that $(a + b)^2 \leq 2(a^2 + b^2)$.

Note that, under CRA in Assumption 2, $y_i(z) - z_i \tau_h - \sum_{i=1}^N [y_i(z) - z_i \tau_h] / N$ can be re-expressed as follows:

$$y_i(z) - \bar{y}_z - z_i \tau_h + (n_1/N)\tau_h.$$

Then, drawing again on the property that $(a + b)^2 \leq 2(a^2 + b^2)$, note that

$$\begin{aligned} \max_z \max_i (y_i(z) - \bar{y}_z - z_i \tau_h + (n_1/N)\tau_h)^2 &\leq 2[\max_z \max_i (y_i(z) - \bar{y}_z)^2 + \max_i ((n_1/N)\tau_h - z_i \tau_h)^2] \\ &= 2 \max_z \max_i (y_i(z) - \bar{y}_z)^2 + 2 \max_i ((n_1/N)\tau_h - z_i \tau_h)^2. \end{aligned}$$

The inequalities in (A10) and (A13) then imply that

$$\max_z \max_i (y_i(z) - \bar{y}_z - z_i \tau_h + (n_1/N)\tau_h)^2 \leq 2[8(A + B^2)] + 2B^2 \leq 16(A + B^2) + 2B^2.$$

Since $\lim_{N \rightarrow \infty} A/N = 0$ by Condition 3.3 of Assumption 3 and $B < \infty$, it follows that

$$\lim_{N \rightarrow \infty} [16(A + B^2) + 2B^2] / N = 0.$$

Finally, since an upper bound of $\max_z \max_i (y_i(z) - \bar{y}_z - z_i \tau_h + (n_1/N)\tau_h)^2 / N$ limits to 0 as $N \rightarrow \infty$, so does $\max_z \max_i (y_i(z) - \bar{y}_z - z_i \tau_h + (n_1/N)\tau_h)^2 / N$ itself, thereby completing the proof. \square

A.7 Lemma 5

Lemma 5. Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. Then

$$\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}) \xrightarrow{\text{a.s.}} 0$$

for all sequences of \mathbf{Z} .

Proof. Under SUTVA and CRA in Assumptions 1 and 2, Lemma A3 of Wu and Ding [60] implies that Lemma 4 suffices for

$$\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}) \xrightarrow{\text{a.s.}} 0,$$

which completes the proof. \square

A.8 Proposition 2

Proposition 2. Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. It follows that

$$\hat{k}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})) \xrightarrow{\text{a.s.}} 0 \quad \text{for all } \tau_h : |\tau - \tau_h| > \varepsilon.$$

Proof. First, note that $E[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z})] = 0$ because the adjusted outcomes $\mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z}$ are fixed over all $\mathbf{w} \in \Omega$. In addition, Lemma 4 implies that, under SUTVA and CRA in Assumptions 1 and 2,

$$\lim_{N \rightarrow \infty} \max_z \max_i \left(y_i(z) - \tau_h z_i - (1/N) \sum_{i=1}^N (y_i(z) - \tau_h z_i) \right)^2 / N = 0$$

for all sequences of \mathbf{Z} . Hence, it follows that

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N \left(y_i(\mathbf{z}) - \tau_h z_i - (1/N) \sum_{i=1}^N y_i(\mathbf{z}) - \tau_h z_i \right)^2 / N = 0,$$

which implies that we can pick a $V < \infty$ that is an upper bound of $\text{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z})]$ for all sequences of \mathbf{Z} .

Next, note that the concentration inequality from Bloniarz et al. [71, Lemma S1] and Wu and Ding [60, Lemma A2] implies that, for all $t \geq 0$ and $N \in \mathbb{N}_{\geq 4}$,

$$\Pr(|\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z})| \geq t) \leq 2 \exp \left\{ -\frac{v^2 N t^2}{4CV} \right\}, \quad (\text{A14})$$

where $C = (71/70)^2$. Since $\Pr(|\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z})| \geq t) \geq \Pr(|\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{z}) - \tau_h \mathbf{z})| = t)$, it follows that $2 \exp \left\{ -\frac{v^2 N t^2}{4CV} \right\}$ in equation (A14) is also an upper bound of the reference PMF in equation (14) for all sequences of \mathbf{Z} .

Let $t = \tau - \tau_h$ and then note that

$$\begin{cases} \lim_{N \rightarrow \infty} 2 \exp \left\{ -\frac{v^2 N (\tau - \tau_h)^2}{4CV} \right\} = 0 & \text{if } \tau \neq \tau_h \\ \lim_{N \rightarrow \infty} 2 \exp \left\{ -\frac{v^2 N (\tau - \tau_h)^2}{4CV} \right\} = 2 & \text{if } \tau = \tau_h. \end{cases}$$

Hence, it follows from Lemma 3 and the CMT that the continuous upper bound in equation (A14) evaluated at t equal to the test-statistic in equation 12 with $\tau_h : |\tau - \tau_h| > \varepsilon$ converges a.s. to 0, which implies that

$$\hat{k}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})) \xrightarrow{\text{a.s.}} 0 \quad \text{for all } \tau_h : |\tau - \tau_h| > \varepsilon,$$

thereby completing the proof. \square

A.9 Proof of Corollary 1

Corollary 1. (Bayesian unbiasedness). *Suppose Assumptions 1 and 2, and let the prior density of \mathcal{T}_h be uniform. It follows that the MAP is equal, in expectation, to the average effect, i.e.,*

$$\mathbb{E} \left[\operatorname{argmax}_{\tau_h \in \mathcal{T}_h} \frac{\hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\operatorname{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h)}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\operatorname{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h) d\tau_h} \right] = \tau.$$

Proof. The proof is identical to that of Proposition 1. \square

A.10 Proof of Corollary 2

Corollary 2. (Bayesian consistency). *Suppose Assumptions 1 and 2 for each $N \in \mathbb{N}_{\geq 4}$, as well as the asymptotic regularity conditions in Assumption 3. If $\tau_h : \tau_h = \tau$ is in the support of the prior distribution, then, as the size of the experiment increases indefinitely, the posterior probability of \mathcal{T}_h^* converges a.s. to 1:*

$$\frac{\int_{\tau_h \in \mathcal{T}_h^*} \hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\operatorname{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h) d\tau_h}{\int_{\tau_h \in \mathcal{T}_h} \hat{\phi}(\hat{\tau}(\mathbf{Z}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z}) / \sqrt{\operatorname{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})]}) r(\tau_h) d\tau_h} \xrightarrow{\text{a.s.}} 1.$$

Proof. First note that, under SUTVA and CRA in Assumptions 1 and 2,

$$\operatorname{Var}[\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})] = \frac{N}{n_1 n_0 (N-1)} \sum_{i=1}^N \left(y_i(\mathbf{Z}) - \tau_h Z_i - \frac{1}{N} \sum_{i=1}^N (y_i(\mathbf{Z}) - \tau_h Z_i) \right)^2$$

and

$$\begin{aligned} \operatorname{Var}[\sqrt{N} \hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})] &= \frac{N}{n_1} \frac{N}{n_0} \frac{1}{(N-1)} \sum_{i=1}^N \left(y_i(\mathbf{Z}) - \tau_h Z_i - \frac{1}{N} \sum_{i=1}^N (y_i(\mathbf{Z}) - \tau_h Z_i) \right)^2 \\ &= \frac{N^2(n_1-1)}{(N-1)n_1 n_0} \hat{S}_1^2 + \frac{n_0}{(N-1)} \left(\frac{1}{n_1} \sum_{i=1}^N Z_i (y_i(\mathbf{Z}) - \tau_h Z_i) - \frac{1}{n_0} \sum_{i=1}^N (1 - Z_i) y_i(\mathbf{Z}) \right)^2 \\ &\quad + \frac{N^2(n_0-1)}{(N-1)n_1 n_0} \hat{S}_0^2 + \frac{n_1}{(N-1)} \left(\frac{1}{n_1} \sum_{i=1}^N Z_i (y_i(\mathbf{Z}) - \tau_h Z_i) - \frac{1}{n_0} \sum_{i=1}^N (1 - Z_i) y_i(\mathbf{Z}) \right)^2 \end{aligned}$$

for each $N \in \mathbb{N}_{\geq 4}$.

Then, under the asymptotic regularity conditions in Assumption 3, Lemma A3 of Wu and Ding [60] and the CMT imply that

$$\text{Var}[\sqrt{N}\hat{\tau}(\mathbf{W}, \mathbf{y}(\mathbf{Z}) - \tau_h \mathbf{Z})] \xrightarrow{\text{a.s.}} \frac{1}{(1-v)} S_{1,\infty}^2 + \frac{1}{v} S_{0,\infty}^2 + (\bar{y}_{1,\infty} - \tau_h - \bar{y}_{0,\infty})^2 > 0,$$

where $\bar{y}_{1,\infty}$ and $\bar{y}_{0,\infty}$ are the limiting means of treated and control potential outcomes, respectively.

The proof then follows from steps identical to those of Theorem 1. □