Causally Sound Priors for Binary Experiments

Nicholas J. Irons* and Carlos Cinelli†

October 31, 2023

Abstract

We introduce the BREASE framework for the Bayesian analysis of randomized controlled trials with a binary treatment and a binary outcome. Approaching the problem from a causal inference perspective, we propose parameterizing the likelihood in terms of the **b**aseline **r**isk, **e**fficacy, and **a**dverse **s**ide **e**ffects of the treatment, along with a flexible, yet intuitive and tractable jointly independent beta prior distribution on these parameters, which we show to be a generalization of the Dirichlet prior for the joint distribution of potential outcomes. Our approach has a number of desirable characteristics when compared to current mainstream alternatives: (i) it naturally induces prior dependence between expected outcomes in the treatment and control groups; (ii) as the baseline risk, efficacy and risk of adverse side effects are quantities commonly present in the clinicians' vocabulary, the hyperparameters of the prior are directly interpretable, thus facilitating the elicitation of prior knowledge and sensitivity analysis; and (iii) we provide analytical formulae for the marginal likelihood, Bayes factor, and other posterior quantities, as well as exact posterior sampling via simulation, in cases where traditional MCMC fails. Empirical examples demonstrate the utility of our methods for estimation, hypothesis testing, and sensitivity analysis of treatment effects.

1 Introduction

Randomized controlled trials (RCTs) form the cornerstone of scientific research across numerous disciplines. In their most basic form, these trials compare the occurrence of an

*PhD Candidate, Department of Statistics, University of Washington, Seattle, WA, USA. Email: njirons@uw.edu. URL: njirons.github.io.

†Assistant Professor, Department of Statistics, University of Washington, Seattle, WA, USA. Email: cinelli@uw.edu. URL: carloscinelli.com.

adverse (or favorable) outcome between treatment and control groups. This is particularly evident in a drug or vaccine trial, in which the efficacy of an intervention is established by comparing the number of individuals who die or develop a disease in each arm of the study. We refer to this type of study design as a "binary experiment," wherein each participant is subjected to either a treatment or a control condition (a binary exposure), and we observe either the presence or absence of the adverse effect of interest (a binary outcome).

If participants of the trial are independent draws from a common (super-)population, statistical inference in binary experiments amounts to what is perhaps the simplest of tasks in statistics—the comparison of two binomial proportions. Indeed, from a Bayesian perspective, inference on the parameter of a binomial distribution dates back to at least as early as the origins of Bayesian inference itself, as evidenced by the seminal works of Bayes (1763) and Laplace (1774). The task comprises specifying a joint prior distribution for both binomial parameters, and computing the posterior distribution (or Bayes factors) of (relevant contrasts of) such parameters (e.g., the risk difference, or the risk ratio). Yet, perhaps surprisingly, despite this long tradition, their widespread occurrence in the sciences, and the apparent simplicity of the inferential task, mainstream approaches for prior specification in the analysis of binary experiments have several shortcomings.

As reviewed in Agresti and Min (2005) and Dablander et al. (2022) (and also evident from perusing popular textbooks[1]) the two predominant approaches for the Bayesian analysis of binary experiments consist of: (i) assigning independent beta priors to each of the binomial proportions, which are conjugate priors to the (also independent) binomials comprising the likelihood; and, (ii) what is essentially a logistic regression, i.e., applying a logit transformation to the binomial proportions, and assigning Gaussian priors to the average log odds and the log odds ratio. For all their popularity, these two approaches are unsatisfactory in several ways. For example, in the first case, the assumption of prior independence of the two proportions is often not credible—e.g., in most settings, one expects that learning about the mortality rate in the control group should inform our beliefs about the mortality rate in the treatment group. Moreover, while the logit approach addresses the problem of prior dependence, it does so at the sacrifice of clarity and interpretation—odds ratios are notoriously difficult to understand (Davies, Crombie, and Tavakoli, 1998), thus hindering the utility of this approach for prior elicitation and sensitivity analysis.

In this paper we demonstrate how causal logic can be used to address these challenges. Approaching the problem from a causal inference perspective, we first propose parameter-

[1]See, e.g., Gelman et al. (1995), Kruschke (2014), and McElreath (2020).

izing the likelihood in terms of three clinically meaningful counterfactual quantities: the baseline risk, efficacy, and risk of adverse side effects (BREASE) of the intervention. We then propose a flexible, yet intuitive and tractable jointly independent beta prior distribution on these parameters, which we show to be a generalization of the Dirichlet prior on the joint distribution of potential outcomes. Our approach has a number of desirable characteristics: (i) it naturally induces prior dependence between the two binomial proportions of the treatment and control arms of the study; (ii) as the baseline risk, efficacy and risk of adverse side effects are quantities familiar to clinicians, the hyperparameters of the prior are directly interpretable, thus facilitating the elicitation of prior knowledge and sensitivity analysis; and (iii) we derive analytical formulae for the marginal likelihood, Bayes factor, and other posterior quantities, as well as exact posterior sampling via simulation, in cases where traditional MCMC fails.

Related literature. When framed in the language of potential outcomes, causal inference can be seen as a missing data problem. Thus, our analysis is most closely related to the literature on contingency tables with missing or incomplete observations on certain cell counts. In fact, our proposed prior can be shown to induce a *generalized* Dirichlet distribution on the joint distribution of potential outcomes. This distribution has been studied in the 1970s and 1980s (Antelman, 1972; Kaufman and King, 1973; Dickey, 1983; Dickey, Jiang, and Kadane, 1987), though mostly in the context of survey sampling.[2] Perhaps due to the intractability of the integrals, the difficulty in interpretation of the original generalized Dirichlet parameterization, and the missing connection to formal causal inference, this prior has received little to no attention in the analysis of binary experiments.[3] Our analysis shows that the generalized Dirichlet distribution emerges naturally from the causal formulation of the problem, that the parameters of the distribution can be cast in intuitive clinical terms, and that statistical inference is manageable, with exact posterior sampling and analytical formulae for Bayes factors, which we derive in this paper.[4]

[2]Similar priors have also appeared in the analysis of diagnostic testing, such as in Branscum, Gardner, and Johnson (2005). This literature seems to be unaware of its connections with the generalized Dirichlet distribution, and some of the results we provide here, such as exact sampling, and analytical formulae for the marginal likelihood, could also be potentially applied to such settings (we leave this to future work).

[3]Related to our setup are studies that have used a *traditional* Dirichlet distribution on response types. This can be shown to be a special case of our proposal, and we discuss it in Sections 2.3 and 3.

[4]The history of statistical analysis of contingency tables is extensive; Killion and Zahn (1976) and Agresti and Hitchcock (2005) provide comprehensive reviews. Along the lines of relevant studies already mentioned, Tian, Ng, and Geng (2003) and Ng, Tang, et al. (2008), identify special cases of Dickey's generalized Dirichlet which admit alternative stochastic representations and simplified computation of posterior quantities. Less relevant to our proposed methodology, other priors used to model contingency

Outline of the paper. Section 2 introduces the statistical setup for the analysis of binary experiments and reviews existing methods for Bayesian inference in this setting. Section 3 introduces our proposal. It also derives key results for implementation, such as analytical formulae for the marginal likelihood and algorithms for exact posterior sampling. Section 4 demonstrates the utility of our method in three empirical examples. Section 5 concludes the paper, and suggests possible extensions for future research.

2 Preliminaries

In this section we set notation, the statistical setup, and briefly review the two main approaches currently used for the Bayesian analysis of binary experiments—the independent beta and logit transformation approaches. We also briefly introduce the response type parameterization of the joint distribution of potential outcomes, which is an important stepping stone for understanding our proposal.

2.1 Potential outcomes

Our analysis is situated within the potential outcomes framework of causal inference (Rubin, 1974; Neyman, 1990). Let N denote the total number of participants in the study, Z_i a binary treatment indicator and Y_i a binary outcome indicator for subject $i \in \{1, \ldots, N\}$. We denote by $Y_i(z)$ the potential outcome of subject i under the experimental condition $Z_i = z$, where $z = 0$ indicates the control and $z = 1$ the treatment condition. Under the standard consistency assumption, we have that the observed outcome of subject i equals the potential outcome associated to the experimental condition that subject i has actually received, i.e., $Y_i = Y_i(Z_i)$. Throughout the paper, we adopt the convention that $Y_i = 1$ denotes an adverse outcome, such as death or the contraction of a disease. We take a super-population perspective, and assume that subjects are independent and identically distributed (i.i.d.) draws from a common population. We assume complete randomization, which implies ignorability of the treatment assignment, $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp Z_i$.

2.2 Marginal parameterization

When subjects are independently drawn from a common super-population and the treatment is assigned at random, it follows that the observed *counts* of adverse outcomes in

table proportions have been proposed in Leonard (1972, 1975), Albert and Gupta (1982), Basu and Pereira (1982), Albert and Gupta (1983a,b, 1985), and Park and Brown (1994).

each treatment arm,

$$y_0 = \sum_{i:Z_i=0} Y_i, \qquad y_1 = \sum_{i:Z_i=1} Y_i,$$

follow independent binomial distributions:

$$y_0 \sim \text{Binomial}(N_0, \theta_0) \quad \perp\!\!\!\perp \quad y_1 \sim \text{Binomial}(N_1, \theta_1),$$

where here, $\theta_1 = \mathbb{P}(Y_i(1) = 1)$, $N_1 = \sum_i Z_i$ denote the probability of an adverse outcome and the sample size of the treatment group, and $\theta_0 = \mathbb{P}(Y_i(0) = 1)$, $N_0 = N - N_1$ are the analogous quantities for the control group.[5] We refer to the probabilities θ_0 and θ_1 as the *baseline risk* and *risk of treatment*, respectively.

This defines the likelihood under the marginal parameterization of a binary experiment—so called because the parameters (θ_0, θ_1) are defined in terms of the marginal distribution of the potential outcomes $Y_i(0)$ and $Y_i(1)$:

$$L(\mathcal{D}|\theta_0, \theta_1) = \binom{N_0}{y_0} \theta_0^{y_0}(1-\theta_0)^{N_0-y_0} \times \binom{N_1}{y_1} \theta_1^{y_1}(1-\theta_1)^{N_1-y_1}, \tag{2.1}$$

where hereafter we denote the observed data by $\mathcal{D} = (y_0, y_1, N_0, N_1)$. To determine the effect of treatment, if any, Bayesian inference is carried out using the posterior distribution of the parameters (θ_0, θ_1), which requires specification of a prior distribution for (θ_0, θ_1). There are two main parameterizations with accompanying priors currently in use, discussed extensively in Agresti and Min (2005) and Dablander et al. (2022)—these are the independent beta (IB) and logit transformation (LT) approaches, which we now discuss.

2.2.1 Independent beta (IB) approach

The independent beta (IB) approach (Jeffreys, 1935) assigns the prior[6]

$$\theta_0 \sim \text{Beta}(a_0, b_0) \quad \perp\!\!\!\perp \quad \theta_1 \sim \text{Beta}(a_1, b_1), \tag{2.2}$$

for some hyperparameters $a_0, b_0, a_1, b_1 > 0$. A common specification is $a_0 = b_0 = a_1 = b_1 = 1$, which assigns a uniform distribution to (θ_0, θ_1). This choice of flat priors is usually

[5] The likelihood of the observed outcomes, conditional on the treatment assignment vector $Z_1, \dots Z_N$, factorizes as $\mathbb{P}(Y_1 \dots, Y_N \mid Z_1 = z_1, \dots, Z_n = z_n) = \mathbb{P}(Y_1(z_1), \dots, Y_N(z_N) \mid Z_1 = z_1, \dots, Z_N = z_N) = \mathbb{P}(Y_1(z_1), \dots, Y_N(z_N)) = \prod_i \mathbb{P}(Y_i(z_i)) = \prod_{i:Z_i=1} \mathbb{P}(Y_i(1)) \prod_{i:Z_i=0} \mathbb{P}(Y_i(0))$, where the first equality is due to consistency, the second equality due to ignorability of the treatment assignment, and the third equality due the assumption that the subjects are i.i.d. draws from a common super-population. Therefore, the data can be seen as a sequence of independent Bernoulli trials, and the counts y_0, y_1 as independent binomials. Note this equivalence does not hold under a finite population perspective; see Ding and Miratrix (2019).

[6] Here $X \sim \text{Beta}(a, b)$ denotes the probability distribution on the unit interval $[0, 1]$ with Lebesgue density proportional to $x^{a-1}(1-x)^{b-1}$.

thought to encode ignorance of (θ_0, θ_1) *a priori*, though it makes strong implicit assumptions as we discuss next. We refer to (2.2) as the IB$(a; b)$ prior, where $a = (a_0, a_1), b = (b_0, b_1)$.[7]

The main advantage of the IB approach is its simplicity. As the beta prior is conjugate to the binomial likelihood, estimation and posterior simulation can be carried out exactly without resorting to approximate sampling algorithms, such as MCMC. Furthermore, marginal likelihoods and Bayes factors, which are widely used for Bayesian hypothesis testing and can be difficult to calculate in general (usually requiring numerical approximation or estimation via posterior simulation), can be calculated analytically (Kass and Raftery, 1995).

A significant drawback of the IB approach is the restrictive assumption of independence between θ_0 and θ_1. In most experimental settings, we would expect our knowledge about the risks in the control and treatment groups to be dependent. For example, if we known that the population prevalence of an infectious disease is approximately 1%, we would expect the prevalence of the disease among those receiving a vaccine to be concentrated around 1% or below, reflecting the common prior belief that it is unlikely that the vaccine would cause the disease. The IB prior fails to accommodate this natural dependence between risks in each arm of the trial. Furthermore, since independence in the prior and the likelihood implies independence *a posteriori*, this failure also extends to the posterior.

2.2.2 Logit Transformation (LT) approach

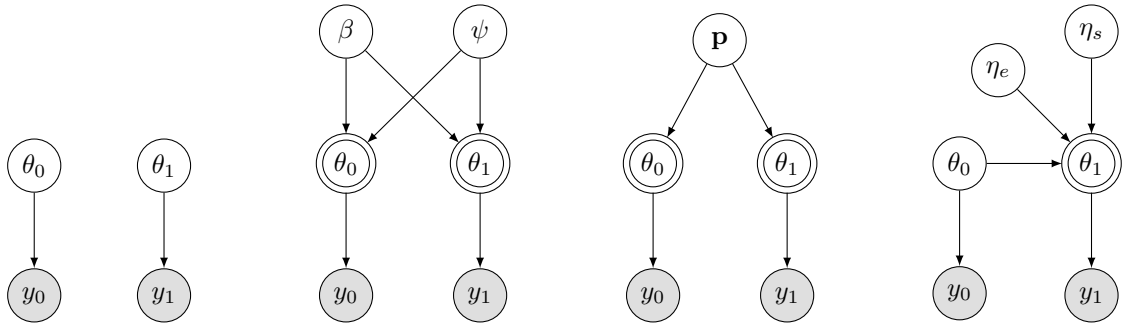
The logit transformation (LT) approach (Kass and Vaidyanathan, 1992; Agresti and Hitchcock, 2005; Dablander et al., 2022) reparameterizes the model in terms of the logit-transformed risks, by defining the parameters (β, ψ) satisfying

$$\log\left(\frac{\theta_0}{1-\theta_0}\right) = \beta - \frac{\psi}{2}, \qquad \log\left(\frac{\theta_1}{1-\theta_1}\right) = \beta + \frac{\psi}{2}.$$

Note this parameterization is equivalent to a logistic regression of the outcome on the treatment with the encoding $Z \in \{-1/2, 1/2\}$ (Gronau, Raj, and Wagenmakers, 2021). It then assigns an independent normal prior to (β, ψ):

$$\beta \sim \text{Normal}(\mu_\beta, \sigma^2_\beta) \quad \perp\!\!\!\perp \quad \psi \sim \text{Normal}(\mu_\psi, \sigma^2_\psi), \tag{2.3}$$

[7]Note that if we consider outcomes with multiple categories (e.g, as in Thall, Simon, and Estey, 1995), the analogous prior here is to assign independent Dirichlet distributions to the vector of probabilities of each arm of the study. This should not be conflated with assigning a Dirichlet prior to the joint distribution of potential outcomes, which we discuss in Section 2.3.



(a) Independent Beta (b) Logit Transform (c) Response Type (d) BREASE

Figure 1: Probabilistic graphical models for different parameterizations and prior setups. Gray nodes denote observed variables, white nodes denote latent parameters, and double borders indicate that a node is a deterministic function of its parents. (a) Independent beta priors are placed directly on θ_0 and θ_1; (b) Independent Gaussian priors are placed on the log odds quantities β and ψ; (c) A Dirichlet prior is placed on the response type probabilities $\mathbf{p}$; (d) Our proposal, independent beta priors are placed on θ_0, η_e, and η_s. In all cases, the observed data depends only on θ_0 and θ_1.

where $\mu = (\mu_\beta, \mu_\psi)$ and $\sigma = (\sigma_\beta, \sigma_\psi) > 0$ are hyperparameters. This prior encodes correlation between θ_0 and θ_1 through their shared dependence on β and ψ. We refer to (2.3) as the LT$(\mu; \sigma)$ prior. Figure 1 depicts probabilistic graphical models comparing the IB and LT parameterizations, as well as the other approaches we will introduce in this paper.

While the LT approach induces prior dependence between θ_0 and θ_1, this comes at the cost of a less intuitive parameterization. Here β is interpreted as the "grand log odds," i.e, the average of the log odds across treatment arms, whereas ψ is the log odds ratio. Odds ratios are notoriously difficult to understand, and thus reasoning about the plausible prior means and variances of log odds—two unbounded hyperparameters—is often challenging in practice. The LT approach also has other computational disadvantages relative to the IB prior. Unlike the IB approach, marginal likelihoods and Bayes factors for the LT approach are not available analytically, and posterior sampling must be carried out approximately.

2.3 Response type (RT) parameterization

The IB and LT approaches focus on the margins of the joint distribution of the potential outcomes $Y_i(0)$ and $Y_i(1)$. This focus is natural, because the observed data depends only upon the parameters θ_0 and θ_1. However, thinking in terms of their *joint* distribution reveals alternative ways of inducing prior dependence between these parameters. Specifically, the

	$Y_i(0) = 0$	$Y_i(0) = 1$	Row Sum
$Y_i(1) = 0$	$p_{00} = (1 - \eta_s)(1 - \theta_0)$	$p_{10} = \eta_e\theta_0$	$1 - \theta_1$
$Y_i(1) = 1$	$p_{01} = \eta_s(1 - \theta_0)$	$p_{11} = (1 - \eta_e)\theta_0$	θ_1
Column Sum	$1 - \theta_0$	θ_0	

Table 1: 2×2 contingency table of potential outcomes for a binary experiment. Only the margins of the table are identified from the observed data.

joint distribution of potential outcomes is fully characterized by four probabilities

$$p_{jk} = \mathbb{P}(Y_i(0) = j, Y_i(1) = k), \qquad j, k \in \{0, 1\}. \tag{2.4}$$

The probabilities $\mathbf{p} = \{p_{jk}\}_{j,k\in\{0,1\}}$ describe the frequencies of the four possible response types in the population (Copas, 1973; Greenland and Robins, 1986).[8] These include: (i) the "doomed" $\{Y_i(0) = 1, Y_i(1) = 1\}$, for whom the adverse outcome occurs regardless of treatment; (ii) the "immune" $\{Y_i(0) = 0, Y_i(1) = 0\}$, for whom the adverse outcome does not occur regardless of treatment; (iii) the "preventive" $\{Y_i(0) = 1, Y_i(1) = 0\}$, for whom treatment *prevents* the adverse outcome; and, (iv) the "causal" $\{Y_i(0) = 0, Y_i(1) = 1\}$, for whom treatment *causes* the adverse outcome. Here θ_0 and θ_1, which satisfy $\theta_0 = p_{10} + p_{11}$ and $\theta_1 = p_{01} + p_{11}$, define the margins of Table 1.

Whereas in the marginal parameterization, independence of the likelihood and prior imply that estimation of θ_0 is only informed by data in the control group (and similarly for θ_1), the response type (RT) parameterization intertwines the data from each arm of the study. The shared dependence of θ_0 and θ_1 on the response type proportions reveals the link between outcomes in the control and treated groups.

A Bayesian approach to modeling the response type probabilities $\mathbf{p}$ requires specification of a prior density supported on the probability simplex, making the Dirichlet distribution a natural candidate[9]

$$\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11}) \sim \text{Dirichlet}(a_{00}, a_{10}, a_{01}, a_{11}), \qquad a_{00}, a_{10}, a_{01}, a_{11} > 0. \tag{2.5}$$

Indeed, priors of this type have been used in the analysis of partially identified quantities in randomized trials with non-compliance, such as in Chickering and Pearl (1996).[10] As

[8] These probabilities are also known as "probabilities of causation" (Tian and Pearl, 2000; Pearl, 2009); for instance, $\mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$ is referred by Tian and Pearl (2000) as the probability that the treatment is both necessary and sufficient to prevent an adverse outcome.

[9] Here $(p_1, \ldots, p_k) \sim \text{Dirichlet}(a_1, \ldots, a_k)$ denotes the probability distribution on the simplex with Lebesgue density proportional to $\prod_{i=1}^{k} p_i^{a_i - 1}$.

[10] See also Imbens and Rubin (1997), Madigan (1999), and Hirano et al. (2000).

we show next, the Dirichlet prior is a special case of our proposal, and our analysis not only extends it, but also clarifies its advantages and limitations as a means to induce the desired joint prior distribution on the two binomial proportions (θ_0, θ_1).

3 The BREASE framework

In this section we introduce the BREASE framework for the analysis of binary experiments. We start by parameterizing the likelihood in terms of the baseline risk, efficacy, and risk of adverse side effects of the treatment. We then propose a jointly independent beta prior distributions on these three parameters, which we show to be a generalization of the Dirichlet prior on the response types. Our proposal has a number of advantages. From a statistical perspective, it induces dependence between the risks in the treatment and control groups, while also enabling exact posterior sampling, and marginal likelihood calculations. From a clinical perspective, this parameterization casts the model in terms of natural quantities appearing frequently in the clinician's vocabulary, thereby facilitating interpretability, elicitation of prior knowledge, and sensitivity analyses.

3.1 Baseline risk, efficacy and adverse side effects

To make things concrete, suppose $Y_i = 1$ denotes death. We define the *efficacy* of the treatment, η_e, as the probability that the treatment *prevents* the death of a patient that would have otherwise died without it:

$$\eta_e = \mathbb{P}(Y_i(1) = 0 | Y_i(0) = 1). \tag{3.1}$$

Similarly, we define the risk of *adverse side effects* of the treatment, η_s, as the probability that the treatment *causes* the death of a patient that would have otherwise been healthy:[11]

$$\eta_s = \mathbb{P}(Y_i(1) = 1 | Y_i(0) = 0). \tag{3.2}$$

These quantities can be interpreted as probabilities of sufficient causation (Tian and Pearl, 2000; Cinelli and Pearl, 2021), i.e., η_e is the probability that treatment is sufficient to save or cure a patient, while η_s is the probability that treatment is sufficient to kill or hurt a patient.

[11]Note these are severe adverse side effects that result in an outcome (e.g, death) opposite to the desired outcome of interest (i.e, survival). In the medical literature, this is sometimes called a "paradoxical reaction" (Smith, Hauben, and Aronson, 2012). Such events could be the result not only of severe adverse biological reactions, but also of other forms of iatrogenesis, such as medical errors.

They correspond directly to the counterfactual interpretation of what clinicians colloquially refer to as "efficacy" and "side effects" of a drug or vaccine. Indeed, not coincidentally, a commonly used measure in clinical trials called "efficacy", defined as $1 - \theta_1/\theta_0$, equals precisely η_e under the assumption that treatment causes no harm ($\eta_s = 0$).

Applying the law of total probability, we can decompose the risk of treatment in terms of the baseline risk, efficacy, and risk of adverse side effects (BREASE), as

$$\theta_1 = (1 - \eta_e)\theta_0 + \eta_s(1 - \theta_0). \tag{3.3}$$

Table 1 shows how the response type probabilities $\mathbf{p}$ can be written as products of θ_0, η_s, and η_e. As with the response type approach, this parameterization highlights the natural dependence between θ_0 and θ_1 that is nevertheless easy to miss without framing the problem in the language of potential outcomes. For example, note that θ_0 and θ_1 are functionally independent only under the strong assumption that $\eta_e = 1 - \eta_s$, i.e., the probability of treatment saving a patient is equal to the probability that it doesn't kill one.

3.1.1 Likelihood

Plugging in (3.3), we can rewrite the likelihood (2.1) in terms of $(\theta_0, \eta_e, \eta_s)$.

Theorem 1. *Under (2.1) and (3.1-3.3), the likelihood is*

$$\begin{aligned} L(\mathcal{D}|\theta_0, \eta_e, \eta_s) = &\binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{j=0}^{y_1}\sum_{k=0}^{N_1-y_1}\binom{y_1}{j}\binom{N_1-y_1}{k} \times \theta_0^{y_0+j+k}(1-\theta_0)^{N-(y_0+j+k)} \\ &\times \eta_e^k(1-\eta_e)^j \\ &\times \eta_s^{y_1-j}(1-\eta_s)^{N_1-y_1-k}, \qquad (\theta_0, \eta_e, \eta_s) \in [0,1]^3. \end{aligned} \tag{3.4}$$

Theorem 1 follows from applying the binomial theorem twice. As the likelihood (3.4) is polynomial in $(\theta_0, \eta_e, \eta_s)$, any prior distribution $\pi(\theta_0, \eta_e, \eta_s)$ for which the moments can be explicitly calculated yields an analytical expression for the marginal likelihood. In particular, if

$$\pi(\theta_0, \eta_e, \eta_s) \propto \theta_0^{\alpha_0-1}(1-\theta_0)^{\beta_0-1} \times \eta_e^{\alpha_e-1}(1-\eta_e)^{\beta_e-1} \times \eta_s^{\alpha_s-1}(1-\eta_s)^{\beta_s-1}$$

is a product of independent beta distributions, as we will see in the next section, then the marginal likelihood is a weighted sum of beta function values. Furthermore, the posterior distribution $\pi(\theta_0, \eta_e, \eta_s|\mathcal{D})$ will be a mixture of independent beta distributions, from which we can sample exactly via simulation.

3.1.2 Partial identification and monotonicity

The parameters η_e and η_s are only partially identified by the observed data. That is, without further assumptions, we have the following bounds,

$$\max\left\{0, 1-\frac{\theta_1}{\theta_0}\right\} \leq \eta_e \leq \min\left\{\frac{1-\theta_1}{\theta_0}, 1\right\}, \quad \max\left\{0, \frac{\theta_1-\theta_0}{1-\theta_0}\right\} \leq \eta_s \leq \min\left\{\frac{\theta_1}{1-\theta_0}, 1\right\}.$$

Thus, as the sample size increases, the posterior distribution of η_s and η_e will not concentrate in a point—rather, it will remain spread over its partially identified region (Richardson, Evans, and Robins, 2011; Gustafson, 2015). Notice, however, that this does not affect the behavior of the posterior distribution of (θ_0, θ_1). The BREASE parameterization thus explicitly separates the identified and partially identified parameters—(θ_0, θ_1) and (η_e, η_s), respectively. Even if interest does not lie in the counterfactual probabilities (η_s, η_e) *per se*, assigning a prior to those quantities can be thought of as a causally principled way to specify a joint prior on the identified target parameters (θ_0, θ_1).

Finally, a common assumption in the potential outcomes literature is called *monotonicity*, which states that the treatment does no harm. In our framework, this corresponds to the constraint $\eta_s = 0$. This assumption is reasonable in many clinical settings. Under monotonicity, the efficacy of the treatment is in fact point identified, and given by $\eta_e = 1 - \theta_1/\theta_0$. The quantity θ_1/θ_0 is known as the risk ratio, and the quantity $1 - \theta_1/\theta_0$ is indeed known as "efficacy" in the clinical trials literature. While the hard constraint $\eta_s = 0$ may not be credible in some settings, if side effects are expected to be small, the BREASE approach allows one to instead place an informative prior on η_s.

3.2 Prior specification

Bayesian inference with the likelihood (3.4) requires specifying a prior distribution on three separate and variation independent probabilities, $(\theta_0, \eta_e, \eta_s)$. We propose setting jointly independent beta prior distributions on these parameters:

$$\theta_0 \sim \text{Beta}^*(\mu_0, n_0) \quad \perp\!\!\!\perp \quad \eta_e \sim \text{Beta}^*(\mu_e, n_e) \quad \perp\!\!\!\perp \quad \eta_s \sim \text{Beta}^*(\mu_s, n_s), \tag{3.5}$$

where here $\text{Beta}^*(\mu, n)$ denotes a $\text{Beta}(a, b)$ distribution, with mean $\mu = a/(a+b)$ and prior "sample size" $n = a + b$. We refer to (3.5) as the BREASE$(\mu; n)$ prior, where $\mu = (\mu_0, \mu_e, \mu_s)$, $n = (n_0, n_e, n_s)$.

Since (3.5) defines a jointly independent beta prior on $(\theta_0, \eta_e, \eta_s)$, the discussion in Section 3.1.1 applies. In particular, the posterior of $(\theta_0, \eta_e, \eta_s)$ is a mixture of independent

betas, which permits exact sampling via simulation, and the marginal likelihood is available analytically as a weighted sum of beta functions, as we show in Sections 3.3 and 3.4.

Connections to the (generalized) Dirichlet. The prior (3.5) induces a *generalized* Dirichlet distribution (Dickey, 1983; Dickey, Jiang, and Kadane, 1987; Tian, Ng, and Geng, 2003) on the vector of potential outcomes probabilities $\mathbf{p}$—see Appendix B for derivation and further discussion. In particular, the generalized Dirichlet reduces to the traditional Dirichlet distribution (2.5) for the following restricted choice of prior sample sizes

$$n_e = \mu_0 n_0, \quad n_s = (1 - \mu_0) n_0. \tag{3.6}$$

Moreover, since $\theta_1 = p_{01} + p_{11}$, by the aggregation property of the Dirichlet (Ng, Tian, and Tang, 2011), marginally we have

$$\theta_1 \sim \text{Beta}^* \left((1 - \mu_e)\mu_0 + \mu_s(1 - \mu_0), n_0 \right), \tag{3.7}$$

which mirrors the decomposition (3.3). The BREASE approach thus reveals an implicit "equal confidence" assumption of the Dirichlet: the prior spread for θ_0 determines the spread of the distributions of η_e, η_s, and θ_1 *a priori*. Hence, the Dirichlet is underparameterized, and unsuitable for cases in which, say, we have ample knowledge of the baseline risk but relatively little information about the possible efficacy or side effects of the treatment (or vice-versa). Casting the likelihood in terms of the BREASE parameters makes such choices explicit, by allowing the hyperparameters governing θ_0, η_e and η_s to be set independently.

3.2.1 Induced prior distribution of (θ_0, θ_1)

As mentioned in Section 3.1.2, our goal with the BREASE approach is primarily to induce causally sound priors on the identified parameters of interest, the two binomial proportions (θ_0, θ_1). Thus we now discuss the induced marginal and conditional distribution of the risk of treatment, θ_1, under the BREASE prior (3.5).

From equation (3.3) we see that θ_1, conditionally on θ_0, is distributed as a convex combination of independent beta random variables *a priori*. This distribution was studied in Pham-Gia and Turkkan (1998) and is given in terms of Appell's first hypergeometric function F_1—in Appendix A we derive the explicit formula and provide further discussion. From here, the marginal prior on θ_1 can be obtained as $\pi(\theta_1) = \int_0^1 \pi(\theta_1|\theta_0)\pi(\theta_0)d\theta_0$. While the general formula for $\pi(\theta_1|\theta_0)$ may look unwieldy, and the integration in $\pi(\theta_1)$ prohibitive,

there are noteworthy specific cases.

Equal confidence. As noted in the previous discussion, under the equal confidence assumption, $n_e = \mu_0 n_0$, $n_s = (1-\mu_0)n_0$, the marginal prior induced on θ_1 is the beta distribution in (3.7). In particular, to obtain equal marginal priors for the treatment and control groups, i.e., $\theta_z \sim \text{Beta}(\mu_0, n_0)$ for $z \in \{0,1\}$, it suffices to set $\mu_s = (\mu_0/(1-\mu_0))\mu_e$, with $0 \leq \mu_e \leq \min(1, (1-\mu_0)/\mu_0)$. Choosing $\mu_0 = 1/2$, $n_0 = 2$, and $\mu_e = \mu_s = \mu$ results in marginal uniform priors with prior correlation $\text{Cor}(\theta_0, \theta_1) = 1 - 2\mu$.

Uniform prior. When at least one of η_e, η_s is uniformly distributed, the conditional prior $\pi(\theta_1|\theta_0)$ reduces to a simple expression in terms of the CDF of the beta distribution, which we derive in Appendix A. In particular, with a flat prior $(\theta_0, \eta_e, \eta_s) \sim \text{Uniform}(0,1)^3$, the marginal on θ_1 is $\pi(\theta_1) = -2\theta_1 \log \theta_1 - 2(1-\theta_1)\log(1-\theta_1)$.

Monotonicity. Under the "no harm" monotonicity assumption, $\eta_s = 0$, we have $\theta_1 = (1-\eta_e)\theta_0$, in which case θ_1 is a product of independent beta random variables *a priori*. Springer and Thompson (1970) derived the form of this distribution, with the density given as a Meijer G-function. In particular, if $n_e = \mu_0 n_0$, we can show that $\theta_1 \sim \text{Beta}((1-\mu_e)n_e, \mu_e n_e + (1-\mu_0)n_0)$. For another example, if $(\theta_0, \eta_e) \sim \text{Uniform}(0,1)^2$, we have $\pi(\theta_1) = -\log \theta_1$. Regarding the conditional prior $\pi(\theta_1|\theta_0)$ under the "no harm" assumption, it is clearly a scaled beta distribution, since $\theta_1 = (1-\eta_e)\theta_0$. If $\eta_e \sim \text{Uniform}(0,1)$, we have $\theta_1|\theta_0 \sim \text{Uniform}(0, \theta_0)$. Similarly, under the "no benefit" assumption $\eta_e = 0$, we have that $\theta_1 = \theta_0 + \eta_s(1-\theta_0)$, which is a scaled and shifted beta random variable conditional on θ_0. If $\eta_s \sim \text{Uniform}(0,1)$, then $\theta_1|\theta_0 \sim \text{Uniform}(\theta_0, 1)$.

Moments. The joint density $\pi(\theta_0, \theta_1)$ induced by the BREASE$(\mu; n)$ prior is generally complicated, but its moments are easily computed in terms of the hyperparameters (μ, n) as θ_1 is a polynomial in $(\theta_0, \eta_e, \eta_s)$, which are beta distributed *a priori*. For example, the prior covariance has a simple form, $\text{Cov}(\theta_0, \theta_1) = \frac{\mu_0(1-\mu_0)}{n_0+1}(1-\mu_e-\mu_s)$. This implies the following directions of the prior correlation,

$$\text{Cor}(\theta_0, \theta_1) \begin{cases} < 0, & \mu_e + \mu_s > 1, \\ = 0, & \mu_e + \mu_s = 1, \\ > 0, & \mu_e + \mu_s < 1. \end{cases} \tag{3.8}$$

In words, θ_0 and θ_1 are positively correlated *a priori* when the expected harm and benefit of treatment are small, and negatively correlated otherwise.

Default prior. While we encourage the use of informative priors, it is useful to have reasonable defaults to start the analysis. If we would like to put θ_0 and θ_1 on equal footing, the BREASE$(1/2, \mu, \mu; 2, 1, 1)$ is thus the natural choice, with the following properties: (i) puts flat uniform priors on θ_0 and θ_1 (as with the IB approach); (ii) induces prior correlation between parameters (as with the LT approach); (iii) assumes no effect of treatment, on average (as with the IB and LT approaches); and, (iv) depends on a single, easily interpretable parameter μ denoting the expected benefits (efficacy) or harm (side effects) of the treatment. When $\mu > 1/2$, θ_1 and θ_0 become anti-correlated, and thus for most cases, $\mu \leq 1/2$ is a reasonable choice. Our preferred specification uses $\mu = 0.3$ as the default. As Figure 6 in the appendix shows, this (weakly) encodes the expectation of moderate effects and concentrates mass on the diagonal $\theta_0 = \theta_1$. This quality is useful in the context of Bayesian hypothesis testing. When testing a null hypothesis H_0 (e.g., no effect of treatment on average, $H_0 : \theta_0 = \theta_1$) nested within an alternative H_1, it is desirable for the prior under H_1 to concentrate mass around the null model (Jeffreys, 1961; Gunel and Dickey, 1974; Casella and Moreno, 2009).

3.3 Posterior sampling

The posterior under (3.5) is given by the following mixture of independent betas[12]

$$\begin{aligned}\pi(\theta_0, \eta_e, \eta_s|\mathcal{D}) \propto \sum_{j=0}^{y_1} \sum_{k=0}^{N_1-y_1} \binom{y_1}{j} \binom{N_1 - y_1}{k} \times \theta_0^{y_0+j+k+\mu_0 n_0} (1-\theta_0)^{N-(y_0+j+k)+(1-\mu_0)n_0} \\ \times \eta_e^{\eta_e; k+\mu_e n_e} (1-\eta_e)^{j+(1-\mu_e)n_e} \\ \times \eta_s^{y_1-j+\mu_s n_s} (1-\eta_s)^{N_1-y_1-k+(1-\mu_s)n_s}. \end{aligned} \tag{3.9}$$

As with the prior, this posterior falls into the family of generalized Dirichlet distributions on the vector of potential outcomes probabilities $\mathbf{p}$. While some posterior quantities can be obtained analytically (see Appendix D), working with the posterior density can often be cumbersome; thus, we now describe how to sample exactly from the posterior via simulation.[13]

Theorem 2. *Let $(\theta_0, \eta_e, \eta_s)$ be random variables drawn according to Algorithm 1. Then $(\theta_0, \eta_e, \eta_s)$ are distributed according to the BREASE posterior (3.9).*

Sketch of proof. Let $I_0 = \{1, \ldots, N_0\}, I_1 = \{N_0 + 1, \ldots, N_0 + N_1\}$ denote the indices of

[12]Here Beta$(x; a, b)$ denotes the density of the Beta(a, b) distribution evaluated at $x \in [0, 1]$.
[13]See Appendix C.1 for a full derivation of Theorem 2.

Algorithm 1 BREASE posterior sampling algorithm

Input: Data $\mathcal{D} = (y_0, y_1, N_0, N_1)$, hyperparameters $(\mu_0, \mu_e, \mu_s, n_0, n_e, n_s)$, and desired number of posterior samples T.

Iterate: For sample $t \in \{1, \ldots, T\}$,

(i) Sample $x_1(1) \in \{0, \ldots, N_1 - y_1\}$ conditional on $\mathcal{D}$ with probability, according to (3.11),

$$\pi(x_1(1)|\mathcal{D}) = \sum_{y_1(0)=0}^{y_1} \pi(y_1(0), x_1(1)|\mathcal{D}).$$

(ii) Sample $y_1(0) \in \{0, \ldots, y_1\}$ conditional on $(x_1(1), \mathcal{D})$ with probability, according to (3.11),

$$\pi(y_1(0)|x_1(1), \mathcal{D}) \propto \pi(y_1(0), x_1(1)|\mathcal{D}).$$

(iii) Sample $(\theta_0, \eta_e, \eta_s)$ conditional on $(y_1(0), x_1(1), \mathcal{D})$ from the distribution (3.12).

Output: Posterior samples $\{(\theta_0^{(t)}, \eta_e^{(t)}, \eta_s^{(t)})\}_{t\in\{1,\ldots,T\}}$.

subjects in the control and treatment groups, respectively. For $j, k \in \{0, 1\}$, we define the counterfactual counts

$$y_j(k) = \sum_{i\in I_j} I(Y_i(j) = 1, Y_i(1-j) = k), \qquad x_j(k) = \sum_{i\in I_j} I(Y_i(j) = 0, Y_i(1-j) = k),$$

which are unobserved quantities. For example, $y_1(0)$ is the number of subjects in the treatment group who died but would not have if untreated. Similarly, $x_1(1)$ is the number of subjects in the treatment group who did not die but would have if untreated. The BREASE posterior can then be expressed as a mixture distribution:

$$\pi(\theta_0, \eta_e, \eta_s|\mathcal{D}) = \sum_{y_1(0)=0}^{y_1} \sum_{x_1(1)=0}^{N_1-y_1} \pi(\theta_0, \eta_e, \eta_s|y_1(0), x_1(1), \mathcal{D}) \times \pi(y_1(0), x_1(1)|\mathcal{D}). \tag{3.10}$$

Hence, we can sample from the posterior by first drawing from the distribution of unobserved counts $(y_1(0), x_1(1))$ conditional on the observed data $\mathcal{D}$. This distribution has probability mass function

$$\begin{aligned}\pi(y_1(0), x_1(1)|\mathcal{D}) \propto & \binom{y_1}{y_1(0)} \binom{N_1 - y_1}{x_1(1)} \mathrm{B}(x_1(1) + \mu_e n_e, y_1 - y_1(0) + (1-\mu_e)n_e) \\ & \times \mathrm{B}(y_0 + y_1 - y_1(0) + x_1(1) + \mu_0 n_0, N - (y_0 + y_1 - y_1(0) + x_1(1)) + (1-\mu_0)n_0) \\ & \times \mathrm{B}(y_1(0) + \mu_s n_s, N_1 - y_1 - x_1(1) + (1-\mu_s)n_s). \end{aligned} \tag{3.11}$$

We then sample the parameters $(\theta_0, \eta_e, \eta_s)$, which have an independent beta distribution

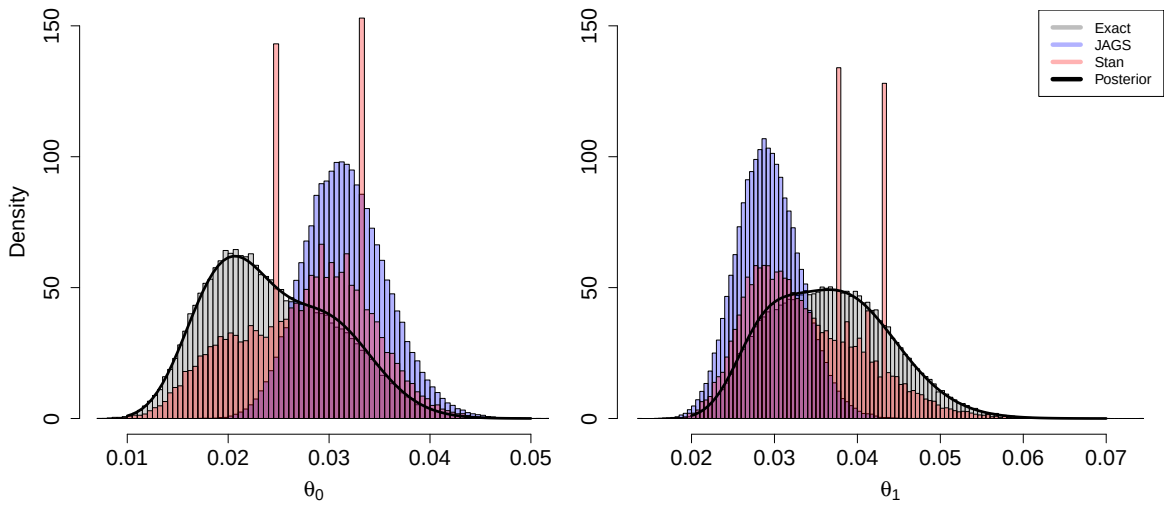


Figure 2: Pathological MCMC posterior sampling exhibited in posterior histograms of the baseline risk θ_0 (left) and treatment risk θ_1 (right). The marginal posterior of θ_1 (black curve) was approximated using numerical integration.

conditional on the augmented data $(y_1(0), x_1(1), \mathcal{D})$:

$$\begin{aligned}\pi(\theta_0, \eta_e, \eta_s | y_1(0), x_1(1), \mathcal{D}) &= \text{Beta}(\eta_e; x_1(1) + \mu_e n_e, y_1 - y_1(0) + (1-\mu_e)n_e) \\ &\times \text{Beta}(\theta_0; y_0 + y_1 - y_1(0) + x_1(1) + \mu_0 n_0, N - (y_0 + y_1 - y_1(0) + x_1(1)) + (1-\mu_0)n_0) \\ &\times \text{Beta}(\eta_s; y_1(0) + \mu_s n_s, N_1 - y_1 - x_1(1) + (1-\mu_s)n_s). \qquad (3.12)\end{aligned}$$

Note this provides a counterfactual interpretation of the mixture weights resulting from the normalization of the kernels in (3.9). □

To demonstrate the utility of exact posterior simulation, we now turn to an example for which RJAGS (Plummer, 2023) and RStan (Stan Development Team, 2023), two popular MCMC software packages, fail to sample from the BREASE posterior. We use the data $y_0 = 20$, $N_0 = 1000$, $y_1 = 40$, $N_1 = 1000$, and the hyperparameters $\mu_0 = 0.5$, $n_0 = 2$, $\mu_e = 0.5$, $n_e = 2, \mu_s = 0.01$, $n_s = 1$. The prior distributions for θ_0 and η_e are vague independent Uniform$(0, 1)$ distributions. On the other hand, the prior on the risk of side effects η_s is concentrated near 0 with mean $\mu_s = 0.01$. This prior encodes a quasi-monotonicity assumption on the treatment that is clearly in conflict with the data.

Prior-data conflict, which arises when the prior is concentrated on parameter values that are unlikely given the data, is a common culprit when diagnosing pathological MCMC sampling (Evans and Moshonov, 2006). This example is no exception. Figure 2 shows histograms of 100,000 posterior samples of θ_0 and θ_1 drawn using Algorithm 1 (grey), JAGS (blue), and Stan (red). The marginal posterior density is plotted in black for reference.

The posterior of θ_0 is a mixture of beta distributions and its multimodality is exhibited in the left panel of Figure 2. While Algorithm 1 produces exact posterior samples that fully capture the distribution, JAGS and Stan fail to adequately explore the left-hand mode. Although Stan manages to deviate from the right-hand mode as compared to JAGS, its chains get stuck at $\theta_0 \approx 0.024$ and $\theta_0 \approx 0.033$ when the sampler rejects numerous proposal draws. The story is much the same for θ_1.

Monotonicity. Posterior sampling under monotonicity constraints, such as setting $\eta_s = 0$ or $\eta_e = 0$, can be obtained with similar procedures, and we thus defer their discussion to the appendix. See Theorems 4-5 of Appendix C.

3.4 Marginal likelihoods and Bayes factors

From a Bayesian perspective, hypothesis testing is essentially a model comparison exercise (Jeffreys, 1961; Dickey and Lientz, 1970; Kass and Raftery, 1995). Consider two competing hypothesis, H_0 and H_1. For each hypothesis H_k, $k \in \{0, 1\}$, the Bayesian approach requires postulating a fully specified model M_k, with likelihood $L_k(\mathcal{D}|\theta)$ and prior $\pi_k(\theta)$, respecting the constraints of the hypothesis the model is intended to represent. Evidence in favor of H_1 relative to H_0 is then quantified using the Bayes factor BF_{10}, given by the ratio of the marginal likelihoods of the observed data under each model, $\text{BF}_{10} = L_1(\mathcal{D})/L_0(\mathcal{D})$, where $L_k(\mathcal{D}) = \int L_k(\mathcal{D}|\theta)\pi_k(\theta)d\theta$. Given prior model probabilities $\mathbb{P}(M_0)$, $\mathbb{P}(M_1)$, the posterior odds of M_1 and M_0 are then $\mathbb{P}(M_1|\mathcal{D})/\mathbb{P}(M_0|\mathcal{D}) = \text{BF}_{10} \times \mathbb{P}(M_1)/\mathbb{P}(M_0)$. In this section we show how to formulate such models instantiating a number of relevant statistical hypotheses with the BREASE approach, and provide analytical formulae for the marginal likelihoods. For all models considered here the likelihood is the same, so we focus the discussion on the formulation of the prior.

Let us first consider testing the null hypothesis $H_0 : \theta_1 = \theta_0$ against the alternative hypothesis $H_1 : \theta_1 \neq \theta_0$. For H_1, we propose using the unconstrained model M_1, with the BREASE prior in (3.5) and equation (3.3),

$$M_1 : (\theta_0, \eta_e, \eta_s) \sim \text{BREASE}(\mu; n), \qquad \theta_1 = (1 - \eta_e)\theta_0 + \eta_s(1 - \theta_0). \tag{3.13}$$

As for the null hypothesis $H_0 : \theta_1 = \theta_0$, we instantiate it with the null model,

$$M_0 : \theta_0 \sim \text{Beta}^*(\mu_0, n_0), \qquad \theta_1 = \theta_0. \tag{3.14}$$

One benefit of M_0 is that its prior is logically consistent with the marginal distribution of θ_0 under M_1, both implying $\theta_0 \sim \text{Beta}^*(\mu_0, n_0)$ *a priori*. Note that the prior (3.14) emerges naturally from M_1 in at least two ways: (i) when postulating that the treatment does not work at all, by setting $\eta_s = \eta_s = 0$; or, (ii) by noting that, if the treatment has no effect on average (i.e, the efficacy of the treatment precisely offsets its side effects), one can side-step thinking about η_s and η_e altogether. In both cases, we borrow the prior of θ_0 from M_1, and simply set θ_1 equal to θ_0. We discuss alternative prior formulations for H_0 in Appendix E.1.

Other relevant hypothesis one may wish to test are that the treatment is beneficial $H_- : \theta_1 < \theta_0$ or that the treatment is harmful $H_+ : \theta_1 > \theta_0$, on average. A straightforward approach to specify models for such hypotheses is to note that M_1 already induces positive probabilities to the events postulated in H_- and H_+. Thus, we can borrow this knowledge, already elicited when forming M_1, to define the priors π_- and π_+,

$$\pi_-(\theta_0, \eta_e, \eta_s) := \pi_1(\theta_0, \eta_e, \eta_s | \theta_1 < \theta_0), \quad \pi_+(\theta_0, \eta_e, \eta_s) := \pi_1(\theta_0, \eta_e, \eta_s | \theta_1 > \theta_0). \tag{3.15}$$

The priors π_- and π_+ result in the models M_- and M_+, for H_- and H_+ respectively. Similarly to M_0, one benefit of these models is that the induced priors on $(\theta_0, \eta_e, \eta_s)$ are logically consistent with the beliefs expressed in M_1, under the constraints H_- and H_+. Note that the same strategy employed here can be used for interval hypotheses of the type $H_0^\delta : |\theta_1 - \theta_0| \leq \delta$, with $\delta > 0$ (or, more generally, for any event with nonzero probability under M_1). Alternative models for H_- and H_+, leveraging instead monotonicity constraints, such as $\eta_s = 0$ or $\eta_e = 0$, are discussed in Appendix E.2.

In all cases above, the marginal likelihood can be obtained using analytical formulae and simple Monte Carlo approximation, thereby facilitating the computation of Bayes factors.

Theorem 3. *The marginal likelihood of the data under M_0 is given by a beta-binomial distribution. Under M_1, it is given by a weighted sum of beta functions:*[14]

$$\begin{aligned} L_1(\mathcal{D}) = \binom{N_0}{y_0}\binom{N_1}{y_1} \sum_{j=0}^{y_1} \sum_{k=0}^{N_1 - y_1} \binom{y_1}{j}\binom{N_1 - y_1}{k} \times \frac{B(k + \mu_e n_e, j + (1-\mu_e)n_e)}{B(\mu_e n_e, (1-\mu_e)n_e)} \\ \times \frac{B(y_0 + j + k + \mu_0 n_0, N - (y_0 + j + k) + (1-\mu_0)n_0)}{B(\mu_0 n_0, (1-\mu_0)n_0)} \\ \times \frac{B(y_1 - j + \mu_s n_s, N_1 - y_1 - k + (1-\mu_s)n_s)}{B(\mu_s n_s, (1-\mu_s)n_s)}. \end{aligned} \tag{3.16}$$

[14]Here $\text{B}(a, b)$ denotes the beta function evaluated at (a, b).

Under M_- and M_+, it can be obtained from $L_1(\mathcal{D})$ as follows,

$$L_-(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 < \theta_0|\mathcal{D})}{\pi_1(\theta_1 < \theta_0)}, \qquad L_+(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 > \theta_0|\mathcal{D})}{\pi_1(\theta_1 > \theta_0)}. \tag{3.17}$$

Proof. The result for M_0 is well-known. $L_1(\mathcal{D})$ in (3.16) follows directly from integration of (3.4) under the prior (3.5). $L_-(\mathcal{D})$ and $L_+(\mathcal{D})$ in (3.17) follow from Bayes' rule. □

Remark 1. The prior and posterior probabilities $\pi_1(\theta_1 < \theta_0)$ and $\pi_1(\theta_1 < \theta_0|\mathcal{D})$ can be approximated using Monte Carlo integration with exact samples, as per Section 3.3.

Remark 2. As per (3.17), if one postulates prior model probabilities $\mathbb{P}(M_-|M_1) = \pi_1(\theta_1 < \theta_0)$ and $\mathbb{P}(M_+|M_1) = \pi_1(\theta_1 > \theta_0)$, the Bayes factor testing $H_0 : \theta_1 = \theta_0$ against $H_1 : \theta_1 \neq \theta_0$ (using M_1) conveniently decomposes into the weighted average of the Bayes factors testing H_0 against H_- (using M_-) and H_0 against H_+ (using M_+)—though, of course, users can postulate prior probabilities for the models M_- and M_+ as they wish.

As noted by Campbell and Gustafson (2022), if one reports a Bayes factor comparing models, it is advisable to also report posterior estimates accounting for model uncertainty, i.e., using the implied mixture prior given by the weighted combination of the priors of all models being compared, $\pi(\theta) = \sum_k \mathbb{P}(M_k)\pi_k(\theta)$. In this case, samples from the mixture posterior can be readily obtained by sampling from the posterior of each model (as detailed in Section 3.3) proportionally to each model's posterior probability, $\pi(\theta|\mathcal{D}) = \sum_k \mathbb{P}(M_k|\mathcal{D})\pi_k(\theta|\mathcal{D})$.

4 Empirical Examples

We now demonstrate the utility of our approach in three empirical examples. We show how the BREASE framework can be used to facilitate Bayesian estimation, hypothesis testing, and sensitivity analysis of the results of binary experiments. Concretely, the examples illustrate how our proposal can: (i) help analysts distinguish robust from fragile findings; (ii) clarify what one needs to believe in order to claim that a treatment is effective; and (iii) reconcile disparate results obtained from different methods.

4.1 The effect of aspirin on myocardial infarction

We revisit the aspirin component of the Physicians' Health Study, a large-scale randomized, placebo-controlled trial designed, in part, to investigate whether low-dose aspirin decreases

the risk of cardiovascular mortality (Physicians' Health Study Research Group, 1989). During the study, $y_0 = 26$ out of $N_0 = 11{,}034$ subjects in the placebo group experienced fatal myocardial infarction compared to $y_1 = 10$ out of $N_1 = 11{,}037$ prescribed aspirin. Using maximum likelihood estimation, the estimated risk ratio θ_1/θ_0 is 0.38, with 95% confidence interval (based on inverting Fisher's exact test) CI(95%) = $[0.17, 0.82]$. Consequently, we reject the null hypothesis of zero effect, $H_0 : \theta_1 = \theta_0$, with p-value 0.008. Results based on asymptotic Wald and Pearson tests are nearly identical. Hence, a frequentist would confidently conclude that low-dose aspirin significantly reduces cardiovascular mortality.

Traditional Bayesian estimation under the alternative hypothesis (i.e, with a prior that gives zero probability to the null hypothesis of zero effect) yields qualitatively similar, though more conservative, answers. Using our default prior, BREASE$(1/2, .3, .3; 2, 1, 1)$, the posterior median of the risk ratio is 0.44 with a wider 95% credible interval of CrI(95%) = $[0.2, 0.96]$. The results for the LT and IB approach are similar.[15]

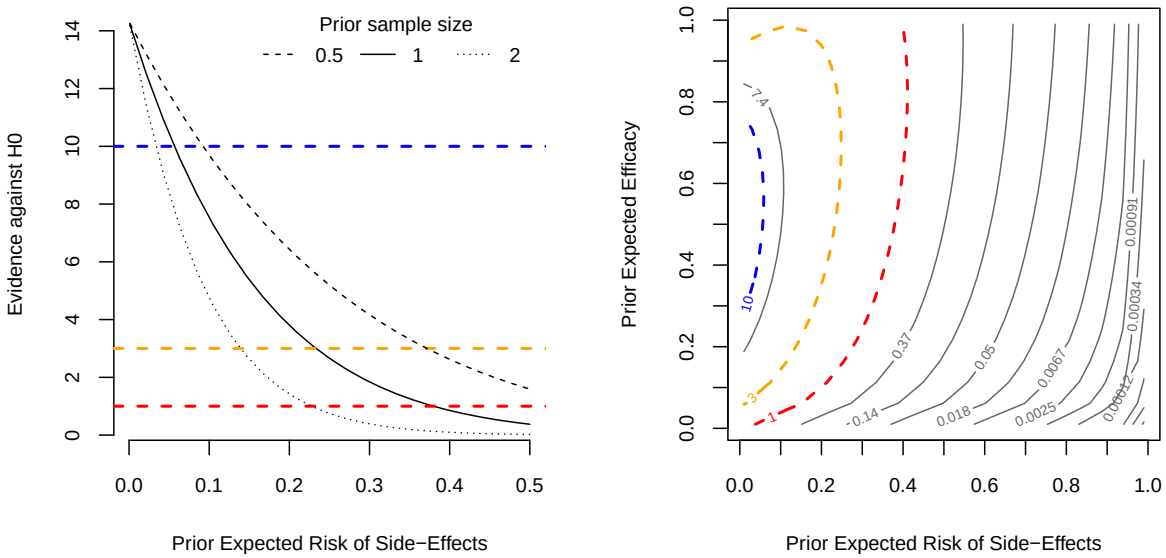
Traditional estimation, however, does not give the null hypothesis of zero effect a fighting chance, as it is assumed to be false *a priori*. One may thus be interested in performing a Bayesian hypothesis test assigning nonzero prior probability to H_0.[16] Perhaps surprisingly, a test based on the IB approach yields a Bayes factor $\text{BF}_{01} = 20.27$, suggesting that the data provide strong evidence *in favor* of H_0. On the other hand, the Bayes factor under the LT approach is $\text{BF}_{10} = 5.24$, which suggests moderate evidence in favor of $H_1 : \theta_1 \neq \theta_0$.[17] Finally, the default BREASE prior results in $\text{BF}_{10} = 1.2$ providing essentially little evidence in favor of one hypothesis or the other. How can we make sense of these disparate results? As is well known, Bayes factors are sensitive to the prior distribution (Kass and Raftery, 1995). It is important, then, that prior assumptions are encoded in a way that practitioners can understand, both to examine the reasonableness of the prior, as well as to explore how robust inferences are to sensible perturbations of the prior (Leamer, 1978; Gunel, 1984; Kass and Raftery, 1995).

One benefit of the BREASE approach is that it allows one to clearly encode prior assumptions in terms of the expected efficacy and side effects of aspirin, and to examine how sensitive the BF is to those assumptions. For example, aspirin is an over-the-counter medicine, with ample usage, and it would thus be unreasonable to expect that aspirin would *cause* myocardial infarction in a large fraction of otherwise healthy patients. Figure 3a

[15] LT(0,0;1,1): median = 0.48 and CrI(95%) = $[0.25, 0.87]$. IB(1,1;1,1): median = 0.4 and CrI(95%) = $[0.18, 0.79]$.

[16] Here we focus on the exact null, but we note that researchers can also specify an interval null hypothesis, such as $|\theta_1 - \theta_0| < \delta$, as per discussion of Section 3.4.

[17] See Appendix F for details on the calculation of Bayes factors for the IB and LT approaches.



(a) Sensitivity plot varying μ_s and n_s.

(b) Sensitivity contours varying μ_s and μ_e.

Figure 3: Sensitivity analysis of BF_{10} for the aspirin trial.

inspects how the Bayes factor is affected as we vary the prior expectation of side effects, ranging from 0.01% to 50%, while still keeping relatively vague priors on the baseline risk and efficacy. The dashed red, orange, and blue lines denote (slightly modified) Jeffreys' thresholds for weak ($1 \leq \text{BF}_{10} \leq 3$), moderate ($3 \leq \text{BF}_{10} \leq 10$), and strong ($\text{BF}_{10} \geq 10$) evidence against H_0, respectively (Jeffreys, 1961; Kass and Raftery, 1995). Indeed, as the plot shows, the results are extremely sensitive to the choice of μ_s. Setting the expected value of side effects to 1% results in $\text{BF}_{10} = 13.45$, yielding strong evidence in favor of H_1, while setting it to 50% results in $\text{BF}_{01} = 2.66$, yielding weak evidence in favor of H_0. Translating these to posterior probabilities, we have the wide range of 27% to 93% probability of the existence of an effect (assuming equal prior odds for H_0 and H_1).

One may also want to conduct a sensitivity analysis with respect to both hyperparameters simultaneously for the $\text{BREASE}(1/2, \mu_e, \mu_s; 2, 1, 1)$ prior. Figure 3b shows the contour lines of BF_{10} as a function of $(\mu_e, \mu_s) \in (0, 1)^2$ over their full range of possible values, while keeping $n_e = n_s = 1$ fixed. In general, the results seem more sensitive to plausible variations of the expected risk of side effects μ_s than to plausible variations of the expected efficacy μ_e of aspirin. Overall, only when (i) side effects are expected to be small ($< 1\%$), and (ii) the efficacy is expected to be relatively large (between 30% and 70%), does the Bayes factor provide strong evidence against the null of no effect. For all other combinations of prior hyperparameters, the evidence is either moderate, weak, or favors the null. In this light, the results of the trial are ambiguous, and the conclusion that aspirin prevents heart

attack strongly depends on the prior. Note that this need not always be the case, as we show next in a reanalysis of the Pfizer-BioNTech COVID-19 vaccine trial.

4.2 The Pfizer-BioNTech COVID-19 vaccine trial

We now reexamine the results of the Pfizer-BioNTech mRNA COVID-19 vaccine study (Polack et al., 2020). The experiment was a global multi-phase randomized placebo-controlled trial designed, in part, to evaluate the efficacy of the BNT162b2 vaccine candidate in preventing COVID-19. Vaccine development and evaluation were carried out in rapid response to the emerging SARS-CoV-2 pandemic. The results of the trial were definitive and precipitated the U.S. Food and Drug Administration's emergency use authorization for widespread dissemination of the vaccine (U.S. Food and Drug Administration, 2020).

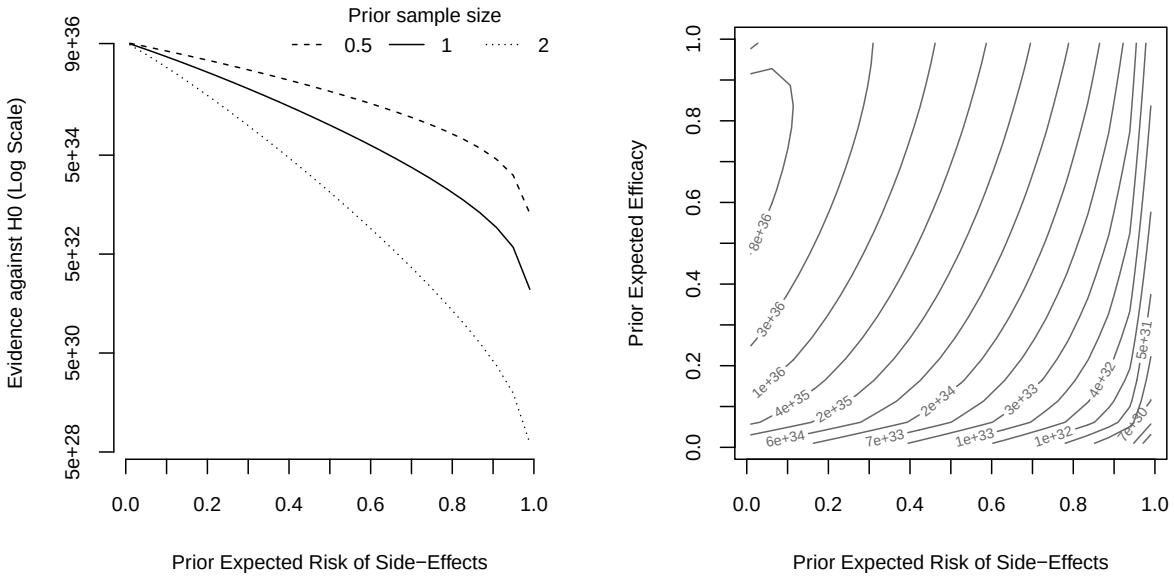
During the study, $y_1 = 9$ out of $N_1 = 19{,}965$ subjects contracted COVID-19 subsequent to the second dose of the vaccine, while there were $y_0 = 169$ cases out of $N_0 = 20{,}172$ subjects receiving placebo injections. In their paper, Polack et al. adopted a Bayesian approach, focusing particularly on evaluating the vaccine's efficacy (defined in the study as the estimand $1 - \theta_1/\theta_0$). The efficacy of the vaccine was estimated at 0.95, with credible interval CrI(95%) $= [0.90, 0.97]$. Frequentist estimates are similar, with a point estimate of 0.95, confidence interval CI(95%) $= [0.90, 0.97]$, and a p-value for testing the null hypothesis of zero effect of the order 6×10^{-33}.

Polack et al. (2020) estimate $1 - \theta_1/\theta_0$ as the efficacy of the vaccine, but, as per Section 3.1.2, this only has the counterfactual interpretation of efficacy (i.e., $\eta_e = 1 - \theta_1/\theta_0$) under the assumption of monotonicity. Using the BREASE approach we can easily encode the monotonicity assumption by setting $\eta_s = 0$ and then proceed with estimation. The default BREASE prior, with the monotonicity constraint, results in posterior median and 95% credible interval for $\eta_e = 1 - \theta_1/\theta_0$ that are essentially the same as the previous results, namely, 0.94 and CrI(95%) $= [0.90, 0.97]$. In the absence of the monotonicity assumption, we have that $1 - \theta_1/\theta_0$ is in fact a lower bound on η_e. Again using the default BREASE prior, results are virtually unchanged, with posterior median and 95% credible interval for $1 - \theta_1/\theta_0$ of 0.94 and CrI(95%) $= [0.90, 0.97]$.[18] Conclusions using the IB and LT priors are practically equivalent.[19]

Turning to hypothesis testing, differently from the aspirin study, here all approaches

[18] Corresponding values for η_e are 0.96 and CrI(95%) $= [0.90, 0.99]$. In this case, however, since η_e is not identified, the posterior of η_e is sensitive to the prior, and it remains spread in the partially identified region of η_e regardless of sample size.

[19] LT(0,0;1,1): med = 0.91, CrI(95%) $= [0.86, 0.95]$. IB(1,1;1,1): med = 0.94, CrI(95%) $= [0.90, 0.97]$.



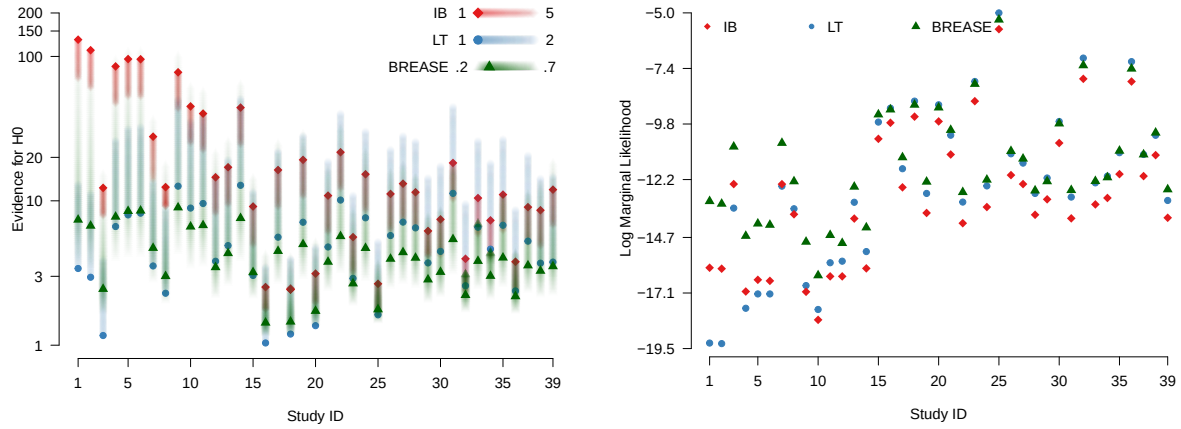
(a) Sensitivity plot varying μ_s and n_s. (b) Sensitivity contours varying μ_s and μ_e.

Figure 4: Sensitivity analysis of BF_{10} for the COVID-19 vaccine trial.

point to the same direction, with overwhelming evidence against H_0. The Bayes factors against the null hypothesis of zero effect are 9×10^{33}, 5×10^{34} and 4×10^{35} for the IB, LT and BREASE default priors, respectively. Further, sensitivity analyses reveal the Bayes factor is in fact robust to variations in the hyperparameters across the whole range of prior expected efficacy and side effects of the vaccine, i.e., $(\mu_e, \mu_s) \in (0, 1)^2$. Figure 4 replicates the same sensitivity plots of the aspirin study for the COVID-19 trial. Notice that, in all scenarios, the posterior probability of the null hypothesis is essentially zero even if we posit equal prior odds for H_0 and H_1. Therefore, in this case, credible intervals constructed under H_1, neglecting H_0, are identical to credible intervals constructed using the mixture prior assigning a point mass of 0.5 to H_0. The trial provides unequivocal evidence that the vaccine is highly efficacious.

4.3 Null results in the *New England Journal of Medicine*

Dablander et al. (2022) conducted a Bayesian reanalysis of 39 binary experiments reporting null results (claiming absence or nonsignificance of an effect of treatment) in the *New England Journal of Medicine* (NEJM). They were particularly concerned with distinguishing between *absence of evidence* and *evidence of absence* of an effect when outcomes in the treatment and control groups are similar. Finding that Bayes factors calculated using the IB approach often strongly favored the null hypothesis (leaning heavily toward *evidence of*



(a) Bayes factors.

(b) Log marginal likelihoods.

Figure 5: Comparisons of log marginal likelihoods and Bayes factors across 39 NEJM studies, for the IB, LT and BREASE priors.

absence) whereas LT Bayes factors were generally equivocal, Dablander et al. concluded that the LT approach should be preferred for Bayesian tests for an equality of proportions. In our final empirical example, we expand their reanalysis to include the BREASE approach, and we show how it can easily address the concerns of Dablander et al. while also providing a better fit to the data in most cases.

Figure 5a contrasts the Bayes factors in favor of the null hypothesis using: (i) the IB$(a, a; a, a)$ prior varying $a \in [1, 5]$ (red diamonds); (ii) the LT$(0, 0; 1, \sigma_\psi)$ prior varying $\sigma_\psi \in [1, 2]$ (blue circles); and, the BREASE$(1/2, \mu, \mu; 2, 1, 1)$ prior varying $\mu \in [.2, .7]$ (green triangles). The solid color stands for the proposed default values of each method, namely $a = 1$ for the IB, $\sigma_\psi = 1$ for the LT and $\mu = .3$ for the BREASE. Note that the Bayes factors of the BREASE and LT default priors (solid triangle and circles) are similar across studies. Moreover, Dablander et al. (2022) noted that, in many examples, the Bayes factors of the IB and LT approaches could not be easily reconciled, even when reasonably varying their hyperparameters. The BREASE approach shows that this behavior is a mere artifact of those parameterizations. Indeed, for all studies, the BREASE prior easily interpolates between the two regimes, thus solving the apparent contradiction between the results of the LT and IB approaches, by transparently revealing how sensitive inferences are to the prior expected efficacy and side effects of the treatment μ. Finally, Figure 5b compares the predictive performance of the default IB, LT, and BREASE priors via the log marginal likelihood. The BREASE prior exhibits superior performance in *every study*

when compared to the IB prior, and in more than 74% of the studies when compared to the LT prior.[20] Thus, in this setting, our proposed default prior seems to provide not only a more sensible parameterization, but also a better fit to the data.

5 Conclusion

We have introduced the BREASE framework for the Bayesian analysis of randomized controlled trials with a binary treatment and outcome. Framing the problem in the language of potential outcomes, we reparameterized the likelihood in terms of clinically meaningful quantities—the baseline risk, efficacy, and risk of adverse side effects of the treatment—and proposed a simple, yet flexible jointly independent beta prior distribution on these parameters. We provided algorithms for exact posterior sampling, as well as analytical formulae for marginal likelihoods, Bayes factors, and other quantities. Finally, we showed with empirical examples how our proposal facilitates estimation, hypothesis testing, elicitation of prior knowledge and sensitivity analysis of treatment effects in binary experiments.

Many interesting extensions of this framework are possible. One possibility is to extend the method to pool evidence across multiple trials. The problem of aggregating evidence is important in its own right, and data from multiple sites may also allow to point identify, or at least narrow the bounds on the fraction of people who benefit or are harmed by the intervention. In a similar vein, another possibility is to extend our framework to the analysis of crossover trials. Under certain assumptions of temporal homogeneity, the efficacy and side effects may be identifiable, making our parameterization and prior proposal natural candidates to the study of treatment effects in such designs.

Acknowledgments. Irons's research was supported by a Shanahan Endowment Fellowship and a Eunice Kennedy Shriver National Institute of Child Health and Human Development training grant, T32 HD101442-01, to the Center for Studies in Demography & Ecology at the University of Washington.

References

Agresti, Alan and David B. Hitchcock (2005). "Bayesian inference for categorical data analysis". In: *Statistical Methods and Applications* 14.3, pp. 297–330.

[20] We use RJAGS (Plummer, 2023) to generate MCMC samples from the LT posterior and THAMES (Metodiev et al., 2023) to estimate the LT marginal likelihood using the samples.

Agresti, Alan and Yongyi Min (2005). "Frequentist Performance of Bayesian Confidence Intervals for Comparing Proportions in 2 × 2 Contingency Tables". In: *Biometrics* 61.2, pp. 515–523.

Albert, James H. and Arjun K. Gupta (1983a). "Bayesian Estimation Methods for 2 × 2 Contingency Tables Using Mixtures of Dirichlet Distributions". In: *Journal of the American Statistical Association* 78.383, pp. 708–717.

— (1985). "Bayesian methods for binomial data with applications to a nonresponse problem". In: *J. Amer. Statist. Assoc.* 80, pp. 167–174.

— (1983b). "Estimation in contingency tables using prior information". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 45.1, pp. 60–69.

— (1982). "Mixtures of Dirichlet Distributions and Estimation in Contingency Tables". In: *The Annals of Statistics* 10.4, pp. 1261–1268.

Antelman, Gordon R. (1972). "Interrelated Bernoulli Processes". In: *Journal of the American Statistical Association* 67.340, pp. 831–841.

Basu, D. and C. A. de B. Pereira (1982). "On the Bayesian analysis of categorical data: the problem of nonresponse". In: *J. Statist. Plann. Inference* 6, pp. 345–362.

Bayes, Thomas (1763). "An essay toward solving a problem in the doctrine of chances, with Richard Price's foreword and discussion". In: *Philos. Trans. R. Soc. London* 53, pp. 370–418.

Branscum, AJ, IA Gardner, and WO Johnson (2005). "Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling". In: *Preventive veterinary medicine* 68.2-4, pp. 145–163.

Campbell, Harlan and Paul Gustafson (2022). "Bayes Factors and Posterior Estimation: Two Sides of the Very Same Coin". In: *The American Statistician* 0.0, pp. 1–11.

Casella, George and Elías Moreno (2009). "Assessing Robustness of Intrinsic Tests of Independence in Two-Way Contingency Tables". In: *Journal of the American Statistical Association* 104.487, pp. 1261–1271.

Chickering, David M. and Judea Pearl (1996). "A Clinician's Tool for Analyzing Noncompliance". In: *Proceedings of the AAAI Conference on Artificial Intelligence, 13*.

Cinelli, Carlos and Judea Pearl (2021). "Generalizing experimental results by leveraging knowledge of mechanisms". In: *European Journal of Epidemiology* 36, pp. 149–164.

Copas, J. B. (1973). "Randomization models for the Matched and Unmatched 2 × 2 Tables". In: *Biometrika* 60.3, pp. 467–476. ISSN: 00063444. URL: `http://www.jstor.org/stable/2334995` (visited on 10/24/2023).

Dablander, Fabian et al. (2022). "A puzzle of proportions: Two popular Bayesian tests can yield dramatically different conclusions". In: *Statistics in Medicine* 41.8, pp. 1319–1333.

Davies, H.T., I.K. Crombie, and M. Tavakoli (1998). "When can odds ratios mislead?" In: *BMJ* 316.7136, pp. 989–991.

Dickey, J. M., J. M. Jiang, and J. B. Kadane (1987). "Bayesian methods for censored categorical data". In: *Journal of the American Statistical Association* 82, pp. 773–781.

Dickey, James M. (1983). "Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses". In: *Journal of the American Statistical Association* 78.383, pp. 628–637.

Dickey, James M. and B. P. Lientz (1970). "The Weighted Likelihood Ratio, Sharp Hypotheses about Chances, the Order of a Markov Chain". In: *The Annals of Mathematical Statistics* 41.1, pp. 214–226.

Ding, Peng and Luke W. Miratrix (2019). "Model-free causal inference of binary experimental data". In: *Scandinavian Journal of Statistics* 46.1, pp. 200–214.

Evans, Michael and Hadas Moshonov (2006). "Checking for Prior-Data Conflict". In: *Bayesian Analysis* 1.4, pp. 893–914.

Gelman, Andrew et al. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Greenland, Sander and James Robins (1986). "Identifiability, Exchangeability, and Epidemiological Confounding". In: *International Journal of Epidemiology* 15.3, pp. 413–419.

Gronau, Quentin F., K. N. Akash Raj, and Eric-Jan Wagenmakers (2021). "Informed Bayesian Inference for the A/B Test". In: *Journal of Statistical Software* 100.17, pp. 1–39.

Gunel, E. (1984). "A Bayesian analysis of the multinomial model for a dichotomous response with nonrespondents". In: *Comm. Statist. Theory Methods* 13, pp. 737–751.

Gunel, Erdogan and James Dickey (1974). "Bayes Factors for Independence in Contingency Tables". In: *Biometrika* 61.3, pp. 545–557.

Gustafson, Paul (2015). *Bayesian inference for partially identified models: Exploring the limits of limited data*. Vol. 140. CRC Press.

Hirano, Keisuke et al. (2000). "Assessing the effect of an influenza vaccine in an encouragement design". In: *Biostatistics* 1.1, pp. 69–88.

Imbens, Guido W. and Donald B. Rubin (1997). "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance". In: *The Annals of Statistics* 25.1, pp. 305–327.

Jeffreys, Harold (1935). "Some Tests of Significance, Treated by the Theory of Probability". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 31.2, pp. 203–222.

— (1961). *Theory of Probability*. 3rd. Oxford, UK: Oxford University Press.

Karson, M.J. and W.J. Wrobleski (1970). "A Bayesian Analysis of a Binomial Model with a Partially Informative Category". In: *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 532–534.

Kass, Robert E. and Adrian E. Raftery (1995). "Bayes Factors". In: *Journal of the American Statistical Association* 90.430, pp. 773–795.

Kass, Robert E. and Suresh K. Vaidyanathan (1992). "Approximate Bayes Factors and Orthogonal Parameters, with Application to Testing Equality of Two Binomial Proportions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 54.1, pp. 129–144.

Kass, Robert E. and Larry Wasserman (1995). "A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz Criterion". In: *Journal of the American Statistical Association* 90.431, pp. 928–934.

Kaufman, G. M. and Benjamin King (1973). "A Bayesian Analysis of Nonresponse in Dichotomous Processes". In: *Journal of the American Statistical Association* 68.343, pp. 670–678.

Killion, Ruth A. and Douglas A. Zahn (1976). "A Bibliography of Contingency Table Literature: 1900 to 1974". In: *International Statistical Review* 44.1, pp. 71–112.

Kruschke, John (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Elsevier Science & Technology.

Laplace, Pierre Simon (1774). "Mémoire sur la probabilité de causes par les évenements". In: *Mémoire de l'académie royale des sciences*.

Leamer, Edward E. (1978). *Specification searches: ad hoc inference with nonexperimental data*. Wiley series in probability and mathematical statistics. New York: Wiley. ISBN: 0471015202.

Leonard, Tom (1975). "Bayesian estimation methods for two-way contingency tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 37.1, pp. 23–37.

— (1972). "Bayesian Methods for Binomial Data". In: *Biometrika* 59.3, pp. 581–589.

Madigan, David (1999). "Bayesian graphical models, intention-to-treat, and the Rubin causal model". In: *Seventh International Workshop on Artificial Intelligence and Statistics*. PMLR.

McElreath, Richard (2020). *Statistical rethinking : a Bayesian course with examples in R and Stan*. Second edition. Texts in statistical science. Chapman & Hall/CRC.

Metodiev, Martin et al. (2023). *Easily Computed Marginal Likelihoods from Posterior Simulation Using the THAMES Estimator*. arXiv: 2305.08952 [stat.ME].

Neyman, Jerzy (1990). "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." In: *Statistical Science* 5.4. Translated from the 1923 Polish original and edited by D. M. Dabrowska and T. P. Speed, pp. 465–472.

Ng, Kai Wang, Man-Lai Tang, et al. (2008). "Grouped Dirichlet distribution: A new tool for incomplete categorical data analysis". In: *Journal of Multivariate Analysis* 99.3, pp. 490–509.

Ng, Kai Wang, Guo-Liang Tian, and Man-Lai Tang (2011). *Dirichlet and related distributions: Theory, methods and applications*. Wiley.

Park, T. and M. B. Brown (1994). "Models for categorical data with nonignorable nonresponse". In: *J. Amer. Statist. Assoc.* 89, pp. 44–52.

Pearl, Judea (2009). *Causality*. Cambridge University Press.

Pham-Gia, T. and N. Turkkan (1998). "Distribution of the linear combination of two general beta variables and applications". In: *Communications in Statistics - Theory and Methods* 27.7, pp. 1851–1869.

Physicians' Health Study Research Group, Steering Committee of the (1989). "Final Report on the Aspirin Component of the Ongoing Physicians' Health Study". In: *New England Journal of Medicine* 321.3, pp. 129–135.

Plummer, Martyn (2023). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-14.

Polack, Fernando P. et al. (2020). "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine". In: *New England Journal of Medicine* 383.27, pp. 2603–2615.

Richardson, Thomas S., Robin J. Evans, and James M. Robins (2011). "Transparent Parametrizations of Models for Potential Outcomes". In: *Bayesian Statistics*. Ed. by José M. Bernardo et al. Vol. 9. Oxford, UK: Oxford University Press.

Robbins, Herbert E. (1992). "An Empirical Bayes Approach to Statistics". In: *Breakthroughs in Statistics: Foundations and Basic Theory*. Ed. by Samuel Kotz and Norman L. Johnson. New York, NY: Springer New York, pp. 388–394.

Rubin, Donald B (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." In: *Journal of educational Psychology* 66.5, p. 688.

Smith, Silas W, Manfred Hauben, and Jeffrey K Aronson (2012). "Paradoxical and bidirectional drug effects". In: *Drug safety* 35, pp. 173–189.

Springer, M. D. and W. E. Thompson (1970). "The Distribution of Products of Beta, Gamma and Gaussian Random Variables". In: *SIAM Journal on Applied Mathematics* 18.4, pp. 721–737.

Stan Development Team (2023). *RStan: the R interface to Stan*. R package version 2.21.8.

Thall, Peter F, Richard M Simon, and Elihu H Estey (1995). "Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes". In: *Statistics in medicine* 14.4, pp. 357–379.

Tian, Guo-Liang, Kai Wang Ng, and Zhi Geng (2003). "Bayesian computation for contingency tables with incomplete cell-counts". In: *Statistica Sinica* 13.1, pp. 189–206.

Tian, Jin and Judea Pearl (2000). "Probabilities of causation: Bounds and identification". In: *Annals of Mathematics and Artificial Intelligence* 28.1, pp. 287–313.

U.S. Food and Drug Administration (Oct. 2020). *Guidance for industry: emergency use authorization for vaccines to prevent COVID-19*. URL: `https://www.fda.gov/media/142749/download`. Accessed 2023/07/12.

Wagenmakers, Eric-Jan et al. (2010). "Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method". In: *Cognitive Psychology* 60.3, pp. 158–189.

Appendix for "Causally Sound Priors for Binary Experiments"

Nicholas J. Irons & Carlos Cinelli

A Implied prior on θ_1

Let the prior of $(\theta_0, \eta_e, \eta_s)$ consist of independent beta distributions with PDFs denoted by $\theta_0 \sim \pi_{\theta_0}(\theta_0)$, $\eta_s \sim \pi_s(\eta_s)$, and $\eta_e \sim \pi_e(\eta_e)$. By the law of total probability, the conditional distribution of θ_1 given θ_0 can be written as

$$\pi(\theta_1 \mid \theta_0) = \int_0^1 \pi(\theta_1 \mid \theta_0, \eta_e)\pi_e(\eta_e)d\eta_e, \tag{A.1}$$

where here we make use of the fact that η_e and θ_0 are *a priori* independent. Note that, conditional on θ_0 and η_e, θ_1 is simply a linear transformation of η_s, namely $\theta_1 = \theta_0(1 - \eta_e) + (1 - \theta_0)\eta_s$. We can thus write the density of θ_1 in terms of the density of η_s as

$$\pi(\theta_1 \mid \theta_0, \eta_e) = \left(\frac{1}{1-\theta_0}\right) \pi_s \left(\frac{\theta_1 - \theta_0(1-\eta_e)}{1-\theta_0}\right),$$

where we make use of the fact that $\dfrac{d\eta_s}{d\theta_1} = \dfrac{1}{1-\theta_0}$. Substituting this back into Eq. A.1, we have the following integral

$$\pi(\theta_1 \mid \theta_0) = \left(\frac{1}{1-\theta_0}\right) \int_0^1 \pi_s \left(\frac{\theta_1 - \theta_0(1-\eta_e)}{1-\theta_0}\right) \pi_e(\eta_e)d\eta_e. \tag{A.2}$$

For the special case where η_e is uniformly distributed, $\pi_e(\eta_e) = 1$, the integral simplifies,

$$\pi(\theta_1 \mid \theta_0) = \left(\frac{1}{1-\theta_0}\right) \int_0^1 \pi_s \left(\frac{\theta_1 - \theta_0(1-\eta_e)}{1-\theta_0}\right) d\eta_e \tag{A.3}$$

$$= \left(\frac{1}{\theta_0}\right) \int_{\frac{\theta_1-\theta_0}{1-\theta_0}}^{\frac{\theta_1}{1-\theta_0}} \pi_s\left(\eta_s\right) d\eta_s \tag{A.4}$$

$$= \left(\frac{1}{\theta_0}\right) \left(F_s\left(\frac{\theta_1}{1-\theta_0}\right) - F_s\left(\frac{\theta_1-\theta_0}{1-\theta_0}\right)\right), \tag{A.5}$$

where the second equality follows from change of variables, noting $d\eta_e = (1-\theta_0)/\theta_0 d\eta_s$. Here $F_s(\cdot)$ denotes the CDF of the beta distribution, which is given by the the regularized incomplete beta function.

For special cases the expression above simplifies. For instance, when η_s is also uniformly

distributed, we have that $F_s(x) = x$, and we obtain a simple closed form expression for the conditional density. Specifically, for $\theta_0 \leq 1/2$,

$$\pi(\theta_1 \mid \theta_0) = \begin{cases} \dfrac{\theta_1}{\theta_0(1-\theta_0)} & \text{if } 0 \leq \theta_1 < \theta_0, \\ \dfrac{1}{1-\theta_0} & \text{if } \theta_0 \leq \theta_1 < 1-\theta_0, \\ \dfrac{1-\theta_1}{\theta_0(1-\theta_0)} & \text{if } 1-\theta_0 \leq \theta_1 \leq 1, \end{cases} \tag{A.6}$$

and zero, otherwise. Analogously, for $\theta_0 \geq 1/2$,

$$\pi(\theta_1 \mid \theta_0) = \begin{cases} \dfrac{\theta_1}{\theta_0(1-\theta_0)} & \text{if } 0 \leq \theta_1 < 1-\theta_0, \\ \dfrac{1}{\theta_0} & \text{if } 1-\theta_0 \leq \theta_1 < \theta_0, \\ \dfrac{1-\theta_1}{\theta_0(1-\theta_0)} & \text{if } \theta_0 \leq \theta_1 \leq 1, \end{cases} \tag{A.7}$$

and zero, otherwise. Notice this is a piece-wise linear function of θ_1. Remarkably, however, integrating each region over θ_0 results in the following marginal distribution of $\pi(\theta_1)$,

$$\pi(\theta_1) = 2(-\theta_1 \log \theta_1 - (1-\theta_1)\log(1-\theta_1)),$$

for $\theta_1 \in [0,1]$, and zero otherwise, which is twice the entropy of the Bernoulli(θ_1) distribution.

More generally, the distribution of linear combinations of beta random variables was studied in Pham-Gia and Turkkan (1998) and is given in terms of Appell's first hypergeometric function F_1, which is an infinite series in two variables:

$$F_1(x, y; a; b_1, b_2; c) = \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \frac{\Gamma(a+m_1+m_2)\Gamma(b_1+m_1)\Gamma(b_2+m_2)\Gamma(c)}{\Gamma(a)\Gamma(b_1)\Gamma(b_2)\Gamma(c+m_1+m_2)} \frac{x^{m_1}}{m_1!} \frac{y^{m_2}}{m_2!}. \tag{A.8}$$

Appell's function also has an integral representation given by

$$F_1(x, y; a; b_1, b_2; c) = \mathrm{B}(a, c-a)^{-1} \int_0^1 u^{a-1}(1-u)^{c-a-1}(1-ux)^{-b_1}(1-uy)^{-b_2} du. \tag{A.9}$$

Applying the results of Pham-Gia and Turkkan (1998) to our setup, the prior on θ_1 conditional on θ_0 induced by the BREASE prior can be obtained as the following piecewise function: (i) for $\theta_0 \leq 1/2$, we have

$$
\begin{aligned}
\pi(\theta_1|\theta_0) = {} & I(0 \le \theta_1 \le \theta_0) \\
& \times \frac{\theta_1^{(1-\mu_e)n_e+\mu_s n_s-1}(\theta_0-\theta_1)^{\mu_e n_e-1}\mathrm{B}(\mu_s n_s,(1-\mu_e)n_e)}{\theta_0^{n_e-1}(1-\theta_0)^{\mu_s n_s}\mathrm{B}(\mu_s n_s,(1-\mu_s)n_s)\mathrm{B}((1-\mu_e)n_e,\mu_e n_e)} \\
& \times F_1\left(\frac{-\theta_1}{\theta_0-\theta_1},\frac{\theta_1}{1-\theta_0};\mu_s n_s;1-\mu_e n_e,1-(1-\mu_s)n_s;(1-\mu_e)n_e+\mu_s n_s\right) \\
& + I(\theta_0 \le \theta_1 \le 1-\theta_0) \\
& \times \frac{(\theta_1-\theta_0)^{\mu_s n_s-1}(1-\theta_1)^{(1-\mu_s)n_s-1}}{(1-\theta_0)^{n_s-1}\mathrm{B}(\mu_s n_s,(1-\mu_s)n_s)} \\
& \times F_1\left(\frac{-\theta_0}{\theta_1-\theta_0},\frac{\theta_0}{1-\theta_1};\mu_e n_e;1-\mu_s n_s,1-(1-\mu_s)n_s;n_e\right) \\
& + I(1-\theta_0 \le \theta_1 \le 1) \\
& \times \frac{(1-\theta_1)^{\mu_e n_e+(1-\mu_s)n_s-1}(\theta_1-\theta_0)^{\mu_s n_s-1}\mathrm{B}(\mu_e n_e,(1-\mu_s)n_s)}{\theta_0^{\mu_e n_e}(1-\theta_0)^{n_s-1}\mathrm{B}(\mu_s n_s,(1-\mu_s)n_s)\mathrm{B}((1-\mu_e)n_e,\mu_e n_e)} \\
& \times F_1\left(\frac{1-\theta_1}{\theta_0},\frac{\theta_1-1}{\theta_1-\theta_0};\mu_e n_e;1-(1-\mu_e)n_e,1-\mu_s n_s;\mu_e n_e+(1-\mu_s)n_s\right).
\end{aligned}
\tag{A.10}
$$

Similarly, (ii) for $\theta_0 \ge 1/2$, we have

$$
\begin{aligned}
\pi(\theta_1|\theta_0) = {} & I(0 \le \theta_1 \le 1-\theta_0) \\
& \times \frac{\theta_1^{(1-\mu_e)n_e+\mu_s n_s-1}(1-\theta_0-\theta_1)^{(1-\mu_s)n_s-1}\mathrm{B}((1-\mu_e)n_e,\mu_s n_s)}{(1-\theta_0)^{n_s-1}\theta_0^{(1-\mu_e)n_e}\mathrm{B}((1-\mu_e)n_e,\mu_e n_e)\mathrm{B}(\mu_s n_s,(1-\mu_s)n_s)} \\
& \times F_1\left(\frac{-\theta_1}{1-\theta_0-\theta_1},\frac{\theta_1}{\theta_0};(1-\mu_e)n_e;1-(1-\mu_s)n_s,1-\mu_e n_e;(1-\mu_e)n_e+\mu_s n_s\right) \\
& + I(1-\theta_0 \le \theta_1 \le \theta_0) \\
& \times \frac{(\theta_1-(1-\theta_0))^{(1-\mu_e)n_e-1}(1-\theta_1)^{\mu_e n_e-1}}{\theta_0^{n_e-1}\mathrm{B}((1-\mu_e)n_e,\mu_e n_e)} \\
& \times F_1\left(\frac{-(1-\theta_0)}{\theta_1-(1-\theta_0)},\frac{1-\theta_0}{1-\theta_1};(1-\mu_s)n_s;1-(1-\mu_e)n_e,1-\mu_e n_e;n_s\right) \\
& + I(\theta_0 \le \theta_1 \le 1) \\
& \times \frac{(1-\theta_1)^{\mu_e n_e+(1-\mu_s)n_s-1}(\theta_1-(1-\theta_0))^{(1-\mu_e)n_e-1}\mathrm{B}((1-\mu_s)n_s,\mu_e n_e)}{(1-\theta_0)^{(1-\mu_s)n_s}\theta_0^{n_e-1}\mathrm{B}((1-\mu_e)n_e,\mu_e n_e)\mathrm{B}(\mu_s n_s,(1-\mu_s)n_s)} \\
& \times F_1\left(\frac{1-\theta_1}{1-\theta_0},\frac{\theta_1-1}{\theta_1-(1-\theta_0)};(1-\mu_s)n_s;1-\mu_s n_s,1-(1-\mu_e)n_e;\mu_e n_e+(1-\mu_s)n_s\right).
\end{aligned}
\tag{A.11}
$$

Monotonicity. Under the "no harm" monotonicity assumption $\eta_s = 0$ we have $\theta_1 = (1-\eta_e)\theta_0$, in which case θ_1 is a product of independent beta random variables *a priori*. Springer and Thompson (1970) derived the form of this distribution, with the density given as a Meijer G-function. In general, this function is expressed as a contour integral in the complex plane. However, when $a_e = \mu_e n_e$, $b_e = (1-\mu_e)n_e$, $a_0 = \mu_0 n_0$, and $b_0 = (1-\mu_0)n_0$ are integers, the prior on θ_1 can be expressed in closed form as

$$\pi(\theta_1) = \frac{\Gamma(n_0)\Gamma(n_e)}{\Gamma(\mu_0 n_0)\Gamma((1-\mu_e)n_e)} \sum_{k=1}^{m} \sum_{j=0}^{e_k-1} \frac{K_{kj}\theta_1^{d_k-1}(-\log\theta_1)^{e_k-j-1}}{\Gamma(e_k-j)\Gamma(j+1)},$$

where $\{d_1, \ldots, d_m\}$ denote the m different integers occurring with multiplicity $\{e_1, \ldots, e_m\}$, respectively, among the sets $\{a_0 - 1, \ldots, a_0 + b_0 - 2\}$ and $\{a_e - 1, \ldots, a_e + b_e - 2\}$, and

$$K_{kj} = \sum_{r=0}^{j} \sum_{q\in\{1,\ldots,m\}, q\neq k} (-1)^{r+1} \binom{j}{r} \frac{\Gamma(r+1)e_q}{(d_q - d_k)^{r+1}}.$$

In particular, if $a_e + b_e = a_0$ (equivalently $n_e = \mu_0 n_0$, an implicit assumption of the Dirichlet prior), we have

$$\theta_1 \sim \text{Beta}((1-\mu_e)n_e, \mu_e n_e + (1-\mu_0)n_0).$$

For another example, if $(\theta_0, \eta_e) \sim \text{Uniform}(0,1)^2$, we have

$$\pi(\theta_1) = -\log\theta_1.$$

Regarding the conditional prior $\pi(\theta_1|\theta_0)$ under the "no harm" assumption, it is clearly a scaled beta distribution, since $\theta_1 = (1-\eta_e)\theta_0$. If $\eta_e \sim \text{Uniform}(0,1)$, we then have that $\theta_1|\theta_0 \sim \text{Uniform}(0, \theta_0)$. Similarly, under the "no benefit" assumption $\eta_e = 0$, we have that $\theta_1 = \theta_0 + \eta_s(1-\theta_0)$, which is a scaled and shifted beta random variable conditional on θ_0. If $\eta_s \sim \text{Uniform}(0,1)$, then $\theta_1|\theta_0 \sim \text{Uniform}(\theta_0, 1)$.

As for the moments, applying the law of total covariance to the terms involving θ_1 by

conditioning on θ_0 and making use of equation (3.3), we obtain

$$\begin{aligned}
\mathrm{Cov}(\theta_0, \theta_1) &= \frac{\mu_0(1-\mu_0)}{n_0+1}(1-\mu_e-\mu_s), \\
\mathrm{Var}(\theta_0) &= \frac{\mu_0(1-\mu_0)}{n_0+1}, \\
\mathrm{Var}(\theta_1) &= \frac{\mu_0(1-\mu_0)}{n_0+1}(1-\mu_e-\mu_s)^2 \\
&\quad + \frac{\mu_e(1-\mu_e)}{n_e+1}\left\{\frac{\mu_0(1-\mu_0)}{n_0+1}+\mu_0^2\right\} \\
&\quad + \frac{\mu_s(1-\mu_s)}{n_s+1}\left\{\frac{\mu_0(1-\mu_0)}{n_0+1}+(1-\mu_0)^2\right\}.
\end{aligned}$$

This can be used to obtain the prior correlation,

$$\mathrm{Cor}(\theta_0, \theta_1) = \frac{\mathrm{Cov}(\theta_0, \theta_1)}{\sqrt{\mathrm{Var}(\theta_0)\mathrm{Var}(\theta_1)}}.$$

B The generalized Dirichlet distribution on p

Given a vector of probabilities $\mathbf{p} = (p_1, \ldots, p_k)$, such that $\sum_{i=1}^k p_i = 1$, the generalized Dirichlet distribution (Tian, Ng, and Geng, 2003) is defined as,

$$\pi(\mathbf{p}) \propto \prod_{i=1}^{k} p_i^{a_i-1} \prod_{j=1}^{m} \left(\sum_{i=1}^{k} \gamma_{ij} p_i\right)^{b_j-1} \tag{B.1}$$

where $\Gamma = (\gamma_{ij})$ is a $k \times m$ known scale matrix. We refer to (B.1) as $\mathrm{GD}(a, b, \Gamma)$. Now consider the vector of potential outcomes $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11})$. By change of variables arguments, if $(\theta_0, \eta_e, \eta_s) \sim \mathrm{BREASE}(\mu; n)$ as in (3.5), it is easy to show that $\mathbf{p}$ has density

$$\pi(\mathbf{p}) \propto p_{00}^{(1-\mu_s)n_s-1} p_{01}^{\mu_s n_s-1} p_{10}^{\mu_e n_e-1} p_{11}^{(1-\mu_e)n_e-1} (p_{00}+p_{01})^{(1-\mu_0)n_0-n_s} (p_{10}+p_{11})^{\mu_0 n_0-n_e}, \tag{B.2}$$

which is a $\mathrm{GD}(a, b, \Gamma)$ distribution with parameters

$$\begin{aligned}
a &= (\mu_s n_s, (1-\mu_s)n_s, \mu_e n_e, (1-\mu_e)n_e), \\
b &= ((1-\mu_0)n_0 - n_s + 1, \mu_0 n_0 - n_e + 1, 1, 1), \\
\Gamma &= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.
\end{aligned}$$

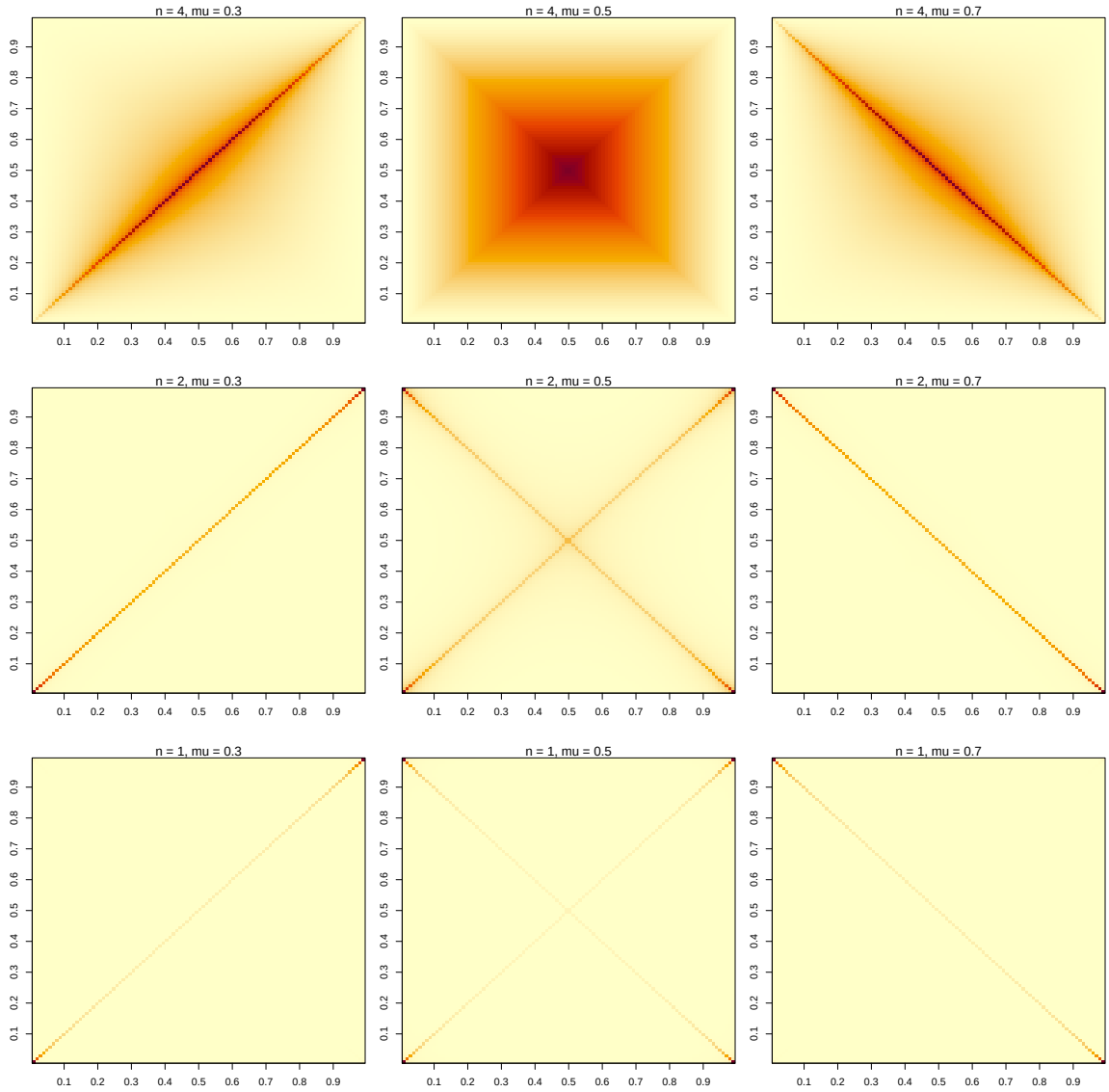


Figure 6: Heatmaps of the joint density of (θ_0, θ_1) under the BREASE$(1/2, \mu, \mu; n, n/2, n/2)$ prior varying n and μ. Our proposed default prior takes $n = 2$ and $\mu = .3$. As the plot shows, this: (i) leads to uniform marginals on θ_0 and θ_1; (ii) assumes zero treatment effect on average; (iii) concentrates mass on the diagonal $\theta_0 = \theta_1$; (iv) favors small (or large) proportions, instead of proportions around the center, which is expected when one quantifies rare outcomes such as death (proportions would be small) or, its complement, survival (in which case proportions would be large).

The prior (B.2) is also a grouped Dirichlet distribution, as defined in Tian, Ng, and Geng (2003) and Ng, Tang, et al. (2008) (which is a special case of the generalized Dirichlet). Similarly, the posterior in (3.9) induces the following posterior distribution on the vector $\mathbf{p}$,

$$\begin{aligned}\pi(\mathbf{p}|\mathcal{D}) \propto\ & p_{00}^{(1-\mu_s)n_s-1} p_{01}^{\mu_s n_s-1} p_{10}^{\mu_e n_e-1} p_{11}^{(1-\mu_e)n_e-1} \\ & \times (p_{00}+p_{01})^{N_0-y_0+(1-\mu_0)n_0-n_s} (p_{10}+p_{11})^{y_0+\mu_0 n_0-n_e} \\ & \times (p_{00}+p_{10})^{N_1-y_1} (p_{01}+p_{11})^{y_1},\end{aligned}$$

which is again a generalized Dirichlet distribution, $\mathrm{GD}(a, b', \Gamma)$, with parameters a and Γ as in the prior, and updated parameter b' given by

$$b' = (N_0 + y_0 + (1-\mu_0)n_0 - n_s + 1,\ y_0 + \mu_0 n_0 - n_e + 1,\ N_1 - y_1 + 1,\ y_1 + 1).$$

The generalized Dirichlet distribution of Dickey (1983), as well as special cases, such as the grouped Dirichlet and Dirichlet-beta, have been proposed for the Bayesian analysis of categorical data and contingency tables with missing observations (Antelman, 1972; Dickey, Jiang, and Kadane, 1987; Gunel, 1984; Karson and Wrobleski, 1970; Kaufman and King, 1973; Ng, Tang, et al., 2008; Tian, Ng, and Geng, 2003). These studies largely focused on the derivation of closed-form expressions (when available) and accurate approximations for posterior moments and predictive probabilities used in estimation and inference. They did not address the parameterization and interpretation of the generalized Dirichlet in terms of the baseline risk, efficacy, and side effects; algorithms for exact posterior simulation; testing for an effect of treatment and sensitivity analysis using analytical formulae; or the specific application to and prior elicitation for binary experiments.

The Dirichlet as a product of independent betas. To better understand the connection of the BREASE prior with the traditional Dirichlet distribution, it is instructive to first derive the distribution of $(\theta_0, \eta_e, \eta_s)$ induced by a Dirichlet prior on the response type probabilities $\mathbf{p}$. The BREASE parameters can be expressed as

$$\theta_0 = p_{10} + p_{11}, \qquad \eta_e = \frac{p_{10}}{p_{10}+p_{11}}, \qquad \eta_s = \frac{p_{01}}{p_{00}+p_{01}}.$$

Elementary properties of the Dirichlet distribution then imply that these quantities are mutually independent beta random variables (Ng, Tian, and Tang, 2011)

$$\theta_0 \sim \mathrm{Beta}(a_{10}+a_{11}, a_{00}+a_{01}) \quad \perp\!\!\!\perp \quad \eta_e \sim \mathrm{Beta}(a_{10}, a_{11}) \quad \perp\!\!\!\perp \quad \eta_s \sim \mathrm{Beta}(a_{01}, a_{00}). \tag{B.3}$$

Similarly, since $\theta_1 = p_{01} + p_{11}$, we also have that $\theta_1 \sim \mathrm{Beta}(a_{01}+a_{11}, a_{00}+a_{10})$ marginally.

While the Dirichlet density seems like a natural choice for the probability vector $\mathbf{p}$, the implied distribution on $(\theta_0, \eta_e, \eta_s)$ reveals some implicit assumptions. In particular, this prior has the peculiar (and potentially undesirable) feature that once we have decided on the parameters underlying the marginal distribution of the efficacy and side effects of treatment (η_e, η_s)—which requires specifying $(a_{00}, a_{10}, a_{01}, a_{11})$—we have fully determined the joint prior on $(\theta_0, \eta_e, \eta_s)$. In this sense, the Dirichlet distribution is underparametrized.

This underparameterization becomes clearer with an alternative representation of the beta distribution, in terms of the prior mean and prior "sample size." For $\mu = a/(a+b)$ and $n = a + b$, we write $\text{Beta}^*(\mu, n)$ to denote a $\text{Beta}(a, b)$ distribution, with mean μ and sample size n. The Dirichlet joint prior on $(\theta_0, \eta_e, \eta_s)$ has then the following alternative stochastic representation,

$$\theta_0 \sim \text{Beta}^*(\mu_0, n_0) \quad \perp\!\!\!\perp \quad \eta_e \sim \text{Beta}^*(\mu_e, \mu_0 n_0) \quad \perp\!\!\!\perp \quad \eta_s \sim \text{Beta}^*(\mu_s, (1-\mu_0)n_0), \tag{B.4}$$

which is equivalent to the BREASE prior imposing a restriction on the choice of prior sample sizes n_e and n_s. Marginally, we also have

$$\theta_1 \sim \text{Beta}^*\left((1-\mu_e)\mu_0 + \mu_s(1-\mu_0), n_0\right), \tag{B.5}$$

which mirrors the decomposition (3.3).

C Posterior sampling

C.1 Proof of Theorem 2

We now describe in greater detail how to sample exactly from the BREASE posterior distribution via simulation.

Proof of Theorem 2. Let $I_0 = \{1, \ldots, N_0\}, I_1 = \{N_0 + 1, \ldots, N_0 + N_1\}$ denote the indices of subjects in the control and treatment groups, respectively. Define the counterfactual counts

$$\begin{aligned} y_j(k) &= \sum_{i \in I_j} I(Y_i(j) = 1, Y_i(1-j) = k), \\ x_j(k) &= \sum_{i \in I_j} I(Y_i(j) = 0, Y_i(1-j) = k), \qquad j, k \in \{0, 1\}. \end{aligned}$$

For example, $y_1(0)$ is the number of subjects in the treatment group who died but would not have if untreated. Similarly, $x_1(1)$ is the number of subjects in the treatment group

who did not die but would have if untreated. The posterior can then be expressed as a mixture distribution:

$$\pi(\theta_0, \eta_e, \eta_s|\mathcal{D}) = \sum_{y_1(0)=0}^{y_1} \sum_{x_1(1)=0}^{N_1-y_1} \pi(\theta_0, \eta_e, \eta_s|y_1(0), x_1(1), \mathcal{D}) \times \pi(y_1(0), x_1(1)|\mathcal{D}). \tag{C.1}$$

We will derive each term in the sum. A straightforward calculation shows that

$$\begin{aligned}
&(y_0(0), y_0(1), x_0(0), x_0(1))|(\theta_0, \eta_e, \eta_s, N_0) \sim \\
&\qquad \text{Multinomial}_{N_0}(\theta_0\eta_e, \theta_0(1-\eta_e), (1-\theta_0)(1-\eta_s), (1-\theta_0)\eta_s), \\
&(y_1(0), y_1(1), x_1(0), x_1(1))|(\theta_0, \eta_e, \eta_s, N_1) \sim \\
&\qquad \text{Multinomial}_{N_1}((1-\theta_0)\eta_s, \theta_0(1-\eta_e), (1-\theta_0)(1-\eta_s), \theta_0\eta_e),
\end{aligned}$$

and the two distributions are independent. Since

$$y_1 = y_1(0) + y_1(1) \qquad \text{and} \qquad N_1 - y_1 = x_1(0) + x_1(1),$$

it follows that

$$\begin{aligned}
y_1(0)|(y_1, \theta_0, \eta_e, \eta_s) &\sim \text{Binomial}\left(y_1, \frac{(1-\theta_0)\eta_s}{\theta_1}\right), \\
x_1(1)|(y_1, N_1, \theta_0, \eta_e, \eta_s) &\sim \text{Binomial}\left(N_1 - y_1, \frac{\theta_0\eta_e}{1-\theta_1}\right),
\end{aligned}$$

independently. Consequently, we have

$$
\begin{aligned}
&\pi(\theta_0, \eta_e, \eta_s | y_1(0), x_1(1), \mathcal{D}) \\
&\quad \propto \pi(y_1(0), x_1(1), \mathcal{D} | \theta_0, \eta_e, \eta_s) \times \pi(\theta_0, \eta_e, \eta_s) \\
&\quad = \pi(y_1(0), x_1(1) | \mathcal{D}, \theta_0, \eta_e, \eta_s) \times \pi(\mathcal{D} | \theta_0, \eta_e, \eta_s) \times \pi(\theta_0, \eta_e, \eta_s) \\
&\quad = \pi(y_1(0) | y_1, \theta_0, \eta_e, \eta_s) \times \pi(x_1(1) | y_1, N_1, \theta_0, \eta_e, \eta_s) \\
&\qquad \times \pi(\mathcal{D} | \theta_0, \eta_e, \eta_s) \times \pi(\theta_0, \eta_e, \eta_s) \\
&\quad = \text{Binomial}\left(y_1(0); y_1, \frac{(1-\theta_0)\eta_s}{\theta_1}\right) \times \text{Binomial}\left(x_1(1); N_1 - y_1, \frac{\theta_0 \eta_e}{1-\theta_1}\right) \\
&\qquad \times \text{Binomial}(y_0; N_0, \theta_0) \times \text{Binomial}(y_1; N_1, \theta_1) \\
&\qquad \times \text{Beta}(\theta_0; \mu_0 n_0, (1-\mu_0) n_0) \times \text{Beta}(\eta_e; \mu_e n_e, (1-\mu_e) n_e) \times \text{Beta}(\eta_s; \mu_s n_s, (1-\mu_s) n_s) \\
&\quad \propto \theta_0^{y_0 + y_1 - y_1(0) + x_1(1) + \mu_0 n_0 - 1} (1-\theta_0)^{N - (y_0 + y_1 - y_1(0) + x_1(1)) + (1-\mu_0) n_0 - 1} \\
&\qquad \times \eta_e^{x_1(1) + \mu_e n_e - 1} (1-\eta_e)^{y_1 - y_1(0) + (1-\mu_e) n_e - 1} \\
&\qquad \times \eta_s^{y_1(0) + \mu_s n_s - 1} (1-\eta_s)^{N_1 - y_1 - x_1(1) + (1-\mu_s) n_s - 1}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\pi(\theta_0, \eta_e, \eta_s | y_1(0), x_1(1), \mathcal{D}) \\
&\quad = \text{Beta}(\theta_0; y_0 + y_1 - y_1(0) + x_1(1) + \mu_0 n_0, N - (y_0 + y_1 - y_1(0) + x_1(1)) + (1-\mu_0) n_0) \\
&\qquad \times \text{Beta}(\eta_e; x_1(1) + \mu_e n_e, y_1 - y_1(0) + (1-\mu_e) n_e) \\
&\qquad \times \text{Beta}(\eta_s; y_1(0) + \mu_s n_s, N_1 - y_1 - x_1(1) + (1-\mu_s) n_s). \qquad \text{(C.2)}
\end{aligned}
$$

Similarly, for the mixture weights we have

$$
\begin{aligned}
&\pi(y_1(0), x_1(1) | \mathcal{D}) = \int \pi(y_1(0), x_1(1), \theta_0, \eta_e, \eta_s | \mathcal{D}) d\theta_0 d\eta_e d\eta_s \\
&\quad = \int \pi(y_1(0), x_1(1) | \theta_0, \eta_e, \eta_s, \mathcal{D}) \pi(\theta_0, \eta_e, \eta_s | \mathcal{D}) d\theta_0 d\eta_e d\eta_s \\
&\quad \propto \binom{y_1}{y_1(0)} \binom{N_1 - y_1}{x_1(1)} \text{B}(x_1(1) + \mu_e n_e, y_1 - y_1(0) + (1-\mu_e) n_e) \\
&\qquad \times \text{B}(y_0 + y_1 - y_1(0) + x_1(1) + \mu_0 n_0, N - (y_0 + y_1 - y_1(0) + x_1(1)) + (1-\mu_0) n_0) \\
&\qquad \times \text{B}(y_1(0) + \mu_s n_s, N_1 - y_1 - x_1(1) + (1-\mu_s) n_s). \qquad \text{(C.3)}
\end{aligned}
$$

Hence, we can sample from the mixture distribution (3.10) as follows:

(i) Sample the unobserved count $x_1(1) \in \{0, \ldots, N_1 - y_1\}$ conditional on $\mathcal{D}$ with proba-

bility

$$\pi(x_1(1)|\mathcal{D}) = \sum_{y_1(0)=0}^{y_1} \pi(y_1(0), x_1(1)|\mathcal{D}).$$

according to (3.11)

(ii) Sample $y_1(0) \in \{0, \ldots, y_1\}$ conditional on $(x_1(1), \mathcal{D})$ with probability

$$\pi(y_1(0)|x_1(1), \mathcal{D}) \propto \pi(y_1(0), x_1(1)|\mathcal{D}).$$

according to (3.11),

(iii) Sample $(\theta_0, \eta_e, \eta_s)$ conditional on $(y_1(0), x_1(1), \mathcal{D})$ from the independent beta distribution (3.12).

□

C.2 Sampling under monotonicity: no harm

Here we derive the BREASE posterior sampling algorithm under the "no harm" ($\eta_s = 0$) monotonicity model $M_-^{'}$ (E.1).

Theorem 4. *Let (θ_0, η_e) be random variables drawn according to Algorithm 2. Then (θ_0, η_e) are distributed according to the posterior of model $M_-^{'}$ (E.1).*

Proof. In this case, we make use of the posterior mixture representation

$$\pi(\theta_0, \eta_e|\mathcal{D}) = \sum_{x_1(1)=0}^{N_1-y_1} \pi(\theta_0, \eta_e|x_1(1), \mathcal{D}) \times \pi(x_1(1)|\mathcal{D}). \tag{C.4}$$

As discussed in Section C.1, we have

$$x_1(1)|(y_1, N_1, \theta_0, \eta_e) \sim \text{Binomial}\left(N_1 - y_1, \frac{\theta_0\eta_e}{1-\theta_1}\right).$$

Note that $\theta_1 = (1-\eta_e)\theta_0$ by hypothesis. Consequently, we have

$$
\begin{aligned}
&\pi(\theta_0, \eta_e | x_1(1), \mathcal{D}) \\
&\quad \propto \pi(x_1(1), \mathcal{D} | \theta_0, \eta_e) \times \pi(\theta_0, \eta_e) \\
&\quad = \pi(x_1(1) | \mathcal{D}, \theta_0, \eta_e) \times \pi(\mathcal{D} | \theta_0, \eta_e) \times \pi(\theta_0, \eta_e) \\
&\quad = \pi(x_1(1) | y_1, N_1, \theta_0, \eta_e) \times \pi(\mathcal{D} | \theta_0, \eta_e) \times \pi(\theta_0, \eta_e) \\
&\quad = \text{Binomial}\left(x_1(1); N_1 - y_1, \frac{\theta_0 \eta_e}{1-\theta_1}\right) \times \text{Binomial}(y_0; N_0, \theta_0) \times \text{Binomial}(y_1; N_1, \theta_1) \\
&\qquad \times \text{Beta}(\theta_0; \mu_0 n_0, (1-\mu_0) n_0) \times \text{Beta}(\eta_e; \mu_e n_e, (1-\mu_e) n_e) \\
&\quad \propto \theta_0^{y_0 + y_1 + x_1(1) + \mu_0 n_0 - 1} (1-\theta_0)^{N - (y_0 + y_1 + x_1(1)) + (1-\mu_0) n_0 - 1} \\
&\qquad \times \eta_e^{x_1(1) + \mu_e n_e - 1} (1-\eta_e)^{y_1 + (1-\mu_e) n_e - 1}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\pi(\theta_0, \eta_e | x_1(1), \mathcal{D}) \\
&\qquad = \text{Beta}(\theta_0; y_0 + y_1 + x_1(1) + \mu_0 n_0, N - (y_0 + y_1 + x_1(1)) + (1-\mu_0) n_0) \\
&\qquad\quad \times \text{Beta}(\eta_e; x_1(1) + \mu_e n_e, y_1 + (1-\mu_e) n_e).
\end{aligned} \tag{C.5}
$$

Similarly, for the mixture weights we have

$$
\begin{aligned}
\pi(x_1(1) | \mathcal{D}) &= \int \pi(x_1(1), \theta_0, \eta_e | \mathcal{D}) d\theta_0 d\eta_e \\
&= \int \pi(x_1(1) | \theta_0, \eta_e, \mathcal{D}) \pi(\theta_0, \eta_e | \mathcal{D}) d\theta_0 d\eta_e \\
&\propto \binom{N_1 - y_1}{x_1(1)} \text{B}(x_1(1) + \mu_e n_e, y_1 + (1-\mu_e) n_e) \\
&\qquad \times \text{B}(y_0 + y_1 + x_1(1) + \mu_0 n_0, N - (y_0 + y_1 + x_1(1)) + (1-\mu_0) n_0).
\end{aligned} \tag{C.6}
$$

Algorithm 2 defines the procedure to sample from the distribution C.4 based on these calculations. □

C.3 Sampling under monotonicity: no benefit

Here we derive the BREASE posterior sampling algorithm under the "no benefit" ($\eta_e = 0$) monotonicity model M_+' (E.2).

Theorem 5. *Let (θ_0, η_s) be random variables drawn according to Algorithm 3. Then (θ_0, η_s) are distributed according to the posterior of model M_+' (E.2).*

Algorithm 2 "No harm" ($\eta_s = 0$) posterior sampling algorithm

Input: Data $\mathcal{D} = (y_0, y_1, N_0, N_1)$, hyperparameters (μ_0, μ_e, n_0, n_e), and desired number of posterior samples T.

Iterate: For sample $t \in \{1, \ldots, T\}$,

(i) Sample $x_1(1) \in \{0, \ldots, N_1 - y_1\}$ conditional on $\mathcal{D}$ with probability $\pi(x_1(1)|\mathcal{D})$ given by (C.6).

(ii) Sample (θ_0, η_e) conditional on $(x_1(1), \mathcal{D})$ from the independent beta distribution (C.5).

Output: Posterior samples $\{(\theta_0^{(t)}, \eta_e^{(t)})\}_{t\in\{1,\ldots,T\}}$.

Proof. In this case, we make use of the posterior mixture representation

$$\pi(\theta_0, \eta_s|\mathcal{D}) = \sum_{y_1(0)=0}^{y_1} \pi(\theta_0, \eta_s|y_1(0), \mathcal{D}) \times \pi(y_1(0)|\mathcal{D}). \tag{C.7}$$

As discussed in Section C.1, we have

$$y_1(0)|(y_1, \theta_0, \eta_s) \sim \text{Binomial}\left(y_1, \frac{(1-\theta_0)\eta_s}{\theta_1}\right).$$

Note that $\theta_1 = \theta_0 + (1-\theta_0)\eta_s$ by hypothesis. Consequently, we have

$$\begin{aligned}
&\pi(\theta_0, \eta_s|y_1(0), \mathcal{D}) \\
&\quad\propto \pi(y_1(0), \mathcal{D}|\theta_0, \eta_s) \times \pi(\theta_0, \eta_s) \\
&\quad= \pi(y_1(0)|\mathcal{D}, \theta_0, \eta_s) \times \pi(\mathcal{D}|\theta_0, \eta_s) \times \pi(\theta_0, \eta_s) \\
&\quad= \pi(y_1(0)|y_1, \theta_0, \eta_s) \times \pi(\mathcal{D}|\theta_0, \eta_s) \times \pi(\theta_0, \eta_s) \\
&\quad= \text{Binomial}\left(y_1(0); y_1, \frac{(1-\theta_0)\eta_s}{\theta_1}\right) \times \text{Binomial}(y_0; N_0, \theta_0) \times \text{Binomial}(y_1; N_1, \theta_1) \\
&\qquad\times \text{Beta}(\theta_0; \mu_0 n_0, (1-\mu_0)n_0) \times \text{Beta}(\eta_s; \mu_s n_s, (1-\mu_s)n_s) \\
&\quad\propto \theta_0^{y_0+y_1-y_1(0)+\mu_0 n_0 - 1}(1-\theta_0)^{N-(y_0+y_1-y_1(0))+(1-\mu_0)n_0-1} \\
&\qquad\times \eta_s^{y_1(0)+\mu_s n_s - 1}(1-\eta_s)^{N_1-y_1+(1-\mu_s)n_s-1}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\pi(\theta_0, \eta_s|y_1(0), \mathcal{D}) \\
&\qquad= \text{Beta}(\theta_0; y_0 + y_1 - y_1(0) + \mu_0 n_0, N - (y_0 + y_1 - y_1(0)) + (1-\mu_0)n_0) \\
&\qquad\quad\times \text{Beta}(\eta_s; y_1(0) + \mu_s n_s, N_1 - y_1 + (1-\mu_s)n_s).
\end{aligned} \tag{C.8}$$

Similarly, for the mixture weights we have

$$\begin{aligned}\pi(y_1(0)|\mathcal{D}) &= \int \pi(y_1(0), \theta_0, \eta_s|\mathcal{D}) d\theta_0 d\eta_s \\ &= \int \pi(y_1(0)|\theta_0, \eta_s, \mathcal{D}) \pi(\theta_0, \eta_s|\mathcal{D}) d\theta_0 d\eta_s \\ &\propto \binom{y_1}{y_1(0)} \mathrm{B}(y_0 + y_1 - y_1(0) + \mu_0 n_0, N - (y_0 + y_1 - y_1(0)) + (1-\mu_0) n_0) \\ &\quad \times \mathrm{B}(y_1(0) + \mu_s n_s, N_1 - y_1 + (1-\mu_s) n_s). \end{aligned} \tag{C.9}$$

Algorithm 3 defines the procedure to sample from the distribution C.7 based on these calculations. □

Algorithm 3 "No benefit" ($\eta_e = 0$) posterior sampling algorithm

Input: Data $\mathcal{D} = (y_0, y_1, N_0, N_1)$, hyperparameters (μ_0, μ_s, n_0, n_s), and desired number of posterior samples T.

Iterate: For sample $t \in \{1, \ldots, T\}$,

(i) Sample $y_1(0) \in \{0, \ldots, y_1\}$ conditional on $\mathcal{D}$ with probability $\pi(y_1(0)|\mathcal{D})$ given by (C.9).

(ii) Sample (θ_0, η_s) conditional on $(y_1(0), \mathcal{D})$ from the independent beta distribution (C.8).

Output: Posterior samples $\{(\theta_0^{(t)}, \eta_s^{(t)})\}_{t \in \{1,\ldots,T\}}$.

C.4 Sampling with an alternate prior under $H_0 : \theta_0 = \theta_1$

We now derive a sampling algorithm for the aggregated Dirichlet prior under H_0 introduced in Section E.1.1:

$$\mathbf{p}^* = (p_{00}, p_{10}^*, p_{11}) \sim \mathrm{Dirichlet}((1-\mu_s)n_s, \mu_e n_e + \mu_s n_s, (1-\mu_e)n_e), \qquad p_{10}^* = p_{10} + p_{01}.$$

The algorithm is based on the posterior decomposition

$$\pi(\mathbf{p}^*|\mathcal{D}) = \sum_{w(0)=0}^{y_0+y_1} \sum_{w(1)=0}^{N_0+N_1-y_0-y_1} \pi(\mathbf{p}^*|w(0), w(1), \mathcal{D}) \times \pi(w(0), w(1)|\mathcal{D}), \tag{C.10}$$

where

$$w(0) = y_0(0) + y_1(0), \qquad w(1) = x_0(1) + x_1(1).$$

We have

$$(y_0(0), y_0(1), x_0(0), x_0(1))|(\mathbf{p}^*, N_0) \sim \text{Multinomial}_{N_0}(p^*_{10}/2, p_{11}, p_{00}, p^*_{10}/2),$$
$$(y_1(0), y_1(1), x_1(0), x_1(1))|(\mathbf{p}^*, N_1) \sim \text{Multinomial}_{N_1}(p^*_{10}/2, p_{11}, p_{00}, p^*_{10}/2),$$

and the two distributions are independent. It follows that

$$\begin{aligned}
y_0(0)|(y_0, \mathbf{p}^*) &\sim \text{Binomial}\left(y_0, \frac{p^*_{10}}{p^*_{10} + 2p_{11}}\right), \\
x_0(1)|(y_0, N_0, \mathbf{p}^*) &\sim \text{Binomial}\left(N_0 - y_0, \frac{p^*_{10}}{p^*_{10} + 2p_{00}}\right), \\
y_1(0)|(y_1, \mathbf{p}^*) &\sim \text{Binomial}\left(y_1, \frac{p^*_{10}}{p^*_{10} + 2p_{11}}\right), \\
x_1(1)|(y_1, N_1, \mathbf{p}^*) &\sim \text{Binomial}\left(N_1 - y_1, \frac{p^*_{10}}{p^*_{10} + 2p_{00}}\right),
\end{aligned}$$

independently. Hence, $w(0)$ and $w(1)$ are distributed independently as

$$\begin{aligned}
w(0)|(y_0, y_1, \mathbf{p}^*) &\sim \text{Binomial}\left(y_0 + y_1, \frac{p^*_{10}}{p^*_{10} + 2p_{11}}\right), \\
w(1)|(\mathcal{D}, \mathbf{p}^*) &\sim \text{Binomial}\left(N_0 + N_1 - y_0 - y_1, \frac{p^*_{10}}{p^*_{10} + 2p_{00}}\right),
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&\pi(\mathbf{p}^*|w(0),w(1), \mathcal{D}) \\
&\quad\propto \pi(w(0), w(1), \mathcal{D}|\mathbf{p}^*) \times \pi(\mathbf{p}^*) \\
&\quad= \pi(w(0), w(1)|\mathcal{D}, \mathbf{p}^*) \times \pi(\mathcal{D}|\mathbf{p}^*) \times \pi(\mathbf{p}^*) \\
&\quad= \pi(w(0)|y_0, y_1, \mathbf{p}^*) \times \pi(w(1)|\mathcal{D}, \mathbf{p}^*) \\
&\qquad\times \pi(\mathcal{D}|\mathbf{p}^*) \times \pi(\mathbf{p}^*) \\
&\quad= \text{Binomial}\left(w(0); y_0 + y_1, \frac{p^*_{10}}{p^*_{10} + 2p_{11}}\right) \\
&\qquad\times \text{Binomial}\left(w(1); N_0 + N_1 - y_0 - y_1, \frac{p^*_{10}}{p^*_{10} + 2p_{00}}\right) \\
&\qquad\times \text{Binomial}(y_0; N_0, p^*_{10}/2 + p_{11}) \times \text{Binomial}(y_1; N_1, p^*_{10}/2 + p_{11}) \\
&\qquad\times (p^*_{10})^{\mu_e n_e + \mu_s n_s - 1} p_{11}^{(1-\mu_e)n_e - 1} p_{00}^{(1-\mu_s)n_s - 1} \\
&\quad\propto (p^*_{10})^{w(0) + w(1) + \mu_e n_e + \mu_s n_s - 1} p_{11}^{y_0 + y_1 - w(0) + (1-\mu_e)n_e - 1} p_{00}^{N_0 + N_1 - y_0 - y_1 - w(1) + (1-\mu_s)n_s - 1}
\end{aligned}$$

It follows that

$$\mathbf{p}^*|(w(0), w(1), \mathcal{D}) \sim \text{Dirichlet}(a_{00}, a_{10}, a_{11}), \tag{C.11}$$

where

$$\begin{aligned}
a_{00} &= N_0 + N_1 - y_0 - y_1 - w(1) + (1 - \mu_s)n_s, \\
a_{10} &= w(0) + w(1) + \mu_e n_e + \mu_s n_s, \\
a_{11} &= y_0 + y_1 - w(0) + (1 - \mu_e)n_e.
\end{aligned}$$

Consequently, for the mixture weights we have

$$\begin{aligned}
\pi(w(0), w(1)|\mathcal{D}) &= \int \pi(w(0), w(1), \mathbf{p}^*|\mathcal{D}) d\mathbf{p}^* \\
&= \int \pi(w(0), w(1)|\mathbf{p}^*, \mathcal{D}) \pi(\mathbf{p}^*|\mathcal{D}) d\mathbf{p}^* \\
&\propto \binom{y_0 + y_1}{w(0)} \binom{N_0 + N_1 - y_0 - y_1}{w(1)} \\
&\quad \times \int (p_{10}^*/2)^{w(0)+w(1)+\mu_e n_e + \mu_s n_s - 1} p_{11}^{y_0 + y_1 - w(0) + (1-\mu_e)n_e - 1} p_{00}^{N_0 + N_1 - y_0 - y_1 - w(1) + (1-\mu_s)n_s - 1} d\mathbf{p}^* \\
&\propto 2^{-(w(0)+w(1))} \binom{y_0 + y_1}{w(0)} \binom{N_0 + N_1 - y_0 - y_1}{w(1)} \text{B}(a_{00}, a_{10}, a_{11}).
\end{aligned} \tag{C.12}$$

Algorithm 4 defines the procedure to sample from the distribution C.10 based on these calculations.

D Posterior quantities of interest

In addition to marginal likelihoods, we can derive analytical expressions for certain relevant functionals of the BREASE posterior distribution $\pi(\theta_0, \eta_e, \eta_s|\mathcal{D})$. While posterior quantities can generally be easily estimated using simple Monte Carlo approximation with samples obtained from Algorithm 1, analytical formulae may be of value, e.g., for conducting prior sensitivity analysis of treatment effect estimands without needing to sample the posterior for every choice of the hyperparameters (μ, n).

The risk difference $\theta_1 - \theta_0$ and risk ratio θ_1/θ_0 are of particular interest in practice, with expectations of their posterior distributions often reported. We first note that, since the posterior $\pi(\theta_0, \eta_e, \eta_s|\mathcal{D})$ is a mixture of independent beta distributions, conditional and marginal expectations and percentiles can be easily computed by first calculating expectations or percentiles of the beta summands and averaging these quantities across the mixture weights. For example, using the mixture representation (3.10) of the posterior, we

Algorithm 4 Alternate $H_0 : \theta_0 = \theta_1$ posterior sampling algorithm

Input: Data (y_0, y_1, N_0, N_1), hyperparameters (μ_e, μ_s, n_e, n_s), and posterior samples T.
Iterate: For sample $t \in \{1, \dots, T\}$,

(i) Sample $w(1) \in \{0, \dots, N_0 + N_1 - y_0 - y_1\}$ conditional on (y_0, y_1, N_0, N_1) as

$$\pi(w(1)|y_0, y_1, N_0, N_1) = \sum_{w(0)=0}^{y_0+y_1} \pi(w(0), w(1)|y_0, y_1, N_0, N_1).$$

(ii) Sample $w(0) \in \{0, \dots, y_0 + y_1\}$ conditional on $(w(1), y_0, y_1, N_0, N_1)$ with probability

$$\pi(w(0)|w(1), y_0, y_1, N_0, N_1) \propto \pi(w(0), w(1)|y_0, y_1, N_0, N_1).$$

(iii) Sample $\mathbf{p}^* = (p_{00}, p_{10}^*, p_{11})$ conditional on $(w(0), w(1), y_0, y_1, N_0, N_1)$ from the Dirichlet distribution (C.11).

(iv) Transform $\mathbf{p}^*$ to obtain samples of $(\theta_0, \theta_1, \eta_e, \eta_s)$ via

$$\theta_0 = p_{10}^*/2 + p_{11} = \theta_1, \qquad \eta_e = \frac{p_{10}^*}{p_{10}^* + 2p_{11}}, \qquad \eta_s = \frac{p_{10}^*}{p_{10}^* + 2p_{00}}.$$

Output: Posterior samples $\{((\mathbf{p}^*)^{(t)}, \theta_0^{(t)}, \theta_1^{(t)}, \eta_e^{(t)}, \eta_s^{(t)})\}_{t \in \{1,\dots,T\}}$.

have

$$\begin{aligned}\mathbb{E}[\theta_0|\mathcal{D}] &= \int \theta_0 \cdot \pi(\theta_0, \eta_e, \eta_s|\mathcal{D}) d\theta_0 d\eta_e d\eta_s \\ &= \sum_{y_1(0)=0}^{y_1} \sum_{x_1(1)=0}^{N_1-y_1} \pi(y_1(0), x_1(1)|\mathcal{D}) \int \theta_0 \cdot \pi(\theta_0, \eta_e, \eta_s|y_1(0), x_1(1), \mathcal{D}) d\theta_0 d\eta_e d\eta_s.\end{aligned}$$

Applying equations (3.11) and (3.12) then yields an expression for $\mathbb{E}[\theta_0|\mathcal{D}]$ in terms of the data $\mathcal{D}$ and hyperparameters (μ, n), which we omit for brevity. In a similar fashion, by exploiting the mixture-of-betas representation of the posterior, we can easily calculate posterior expectations of polynomials $\sum_{(\alpha_0,\alpha_e,\alpha_s)} a_{(\alpha_0,\alpha_e,\alpha_s)} \theta_0^{\alpha_0} \eta_e^{\alpha_e} \eta_s^{\alpha_s}$, including those with negative exponents, assuming $\mathcal{D}$ and (μ, n) are such that the integrals converge.

In particular, assuming treatment is not harmful ($\eta_s = 0$), the efficacy can be written in terms of the risk ratio as $\eta_e = 1 - \theta_1/\theta_0$. The formulae derived in Appendix C.2 can then be applied to calculate $\mathbb{E}[\theta_1/\theta_0|\mathcal{D}] = 1 - \mathbb{E}[\eta_e|\mathcal{D}]$ using the posterior $\pi(\theta_0, \eta_e|\mathcal{D})$ under the monotonicity assumption. More generally, we have

$$\begin{aligned}\mathbb{E}[\theta_1/\theta_0|\mathcal{D}] &= \mathbb{E}\left[\frac{\theta_0(1 - \eta_e - \eta_s) + \eta_s}{\theta_0}\middle|\mathcal{D}\right] \\ &= 1 - \mathbb{E}[\eta_e|\mathcal{D}] - \mathbb{E}[\eta_s|\mathcal{D}] + \mathbb{E}[\theta_0^{-1}\eta_s|\mathcal{D}].\end{aligned}$$

Similarly, the expected posterior risk difference can be obtained as

$$\mathbb{E}[\theta_1 - \theta_0|\mathcal{D}] = \mathbb{E}[\eta_s|\mathcal{D}] - \mathbb{E}[\theta_0\eta_e|\mathcal{D}] - \mathbb{E}[\theta_0\eta_s|\mathcal{D}].$$

In Section 4 we demonstrate how to conduct sensitivity analysis with the BREASE prior for Bayes factors using the marginal likelihoods derived in Section 3.4. The discussion therein applies just as well to treatment effects and other posterior quantities.

E Alternative models and priors

E.1 Other priors for H_0

Recalling that $\theta_0 = p_{10} + p_{11}$ and $\theta_1 = p_{01} + p_{11}$, we see that $\theta_0 = \theta_1$ if and only if $p_{10} = p_{01}$. In this light, we discuss some alternate priors that conform to these constraints. While instantiating H_0 using the beta-binomial model M_0 (3.14) should be preferable in most applications, the prior we discuss here may apply in cases where one has stronger prior information concerning the efficacy and side effects of treatment (η_e, η_s) rather than the baseline risk θ_0 itself.

E.1.1 Aggregated Dirichlet

With a Dirichlet$^*(\mu_0, \mu_e, \mu_s; n_0)$ prior on $\mathbf{p}$, we have by the aggregation property of the Dirichlet distribution (Ng, Tian, and Tang, 2011)

$$(p_{00}, p_{10} + p_{01}, p_{11}) \sim \text{Dirichlet}((1-\mu_s)n_s, \mu_e n_e + \mu_s n_s, (1-\mu_e)n_e),$$

where $n_e = \mu_0 n_0$ and $n_s = (1-\mu_0)n_0$. Assuming H_0 holds, and defining $p^*_{10} = p_{10} + p_{01} = 2p_{10}$, we obtain the Dirichlet prior density on the aggregated cell probabilities

$$\pi(p_{00}, p^*_{10}) = \text{B}((1-\mu_s)n_s, \mu_e n_e + \mu_s n_s, (1-\mu_e)n_e)^{-1} p_{00}^{(1-\mu_s)n_s - 1} (p^*_{10})^{\mu_e n_e + \mu_s n_s - 1} p_{11}^{(1-\mu_e)n_e - 1},$$

where $p_{11} = 1 - p_{00} - p^*_{10}$ and $\text{B}(a_{00}, a_{10}, a_{11})$ is the multivariate beta function:

$$\text{B}(a_{00}, a_{10}, a_{11}) = \frac{\Gamma(a_{00})\Gamma(a_{10})\Gamma(a_{11})}{\Gamma(a_{00} + a_{10} + a_{11})}.$$

This prior allows for exact posterior sampling and marginal likelihood calculation in cases where we may have stronger prior information concerning the efficacy and side effects of treatment (η_e, η_s) than the baseline risk θ_0. Indeed, note that the prior is fully specified by the hyperparameters (μ_e, μ_s, n_e, n_s). Recalling that the Dirichlet* prior is obtained from

the generalized Dirichlet by setting $n_e = \mu_0 n_0$ and $n_s = (1-\mu_0)n_0$, we see that this prior assumes that we have as much prior knowledge on θ_0 as we do on (η_e, η_s).

With this parametrization, the likelihood under H_0 is given by

$$L(\mathcal{D}|p) = \binom{N_0}{y_0}\binom{N_1}{y_1}(p^*_{10}/2 + p_{11})^{y_0+y_1}(p_{00} + p^*_{10}/2)^{N_0+N_1-y_0-y_1}.$$

The posterior is then

$$\begin{aligned}\pi(p_{00}, p^*_{10}|\mathcal{D}) \propto \binom{N_0}{y_0}\binom{N_1}{y_1}&\mathrm{B}((1-\mu_s)n_s, \mu_e n_e + \mu_s n_s, (1-\mu_e)n_e)^{-1} \\ &\times (p^*_{10}/2 + p_{11})^{y_0+y_1}(p_{00} + p^*_{10}/2)^{N_0+N_1-y_0-y_1} \\ &\times (p^*_{10})^{\mu_e n_e + \mu_s n_s - 1} p_{11}^{(1-\mu_e)n_e - 1} p_{00}^{(1-\mu_s)n_s - 1}.\end{aligned}$$

From here we can apply the binomial theorem twice to quickly see that the posterior is a mixture of Dirichlet densities on the probability vector $\mathbf{p}^* = (p_{00}, p^*_{10}, p_{11})$. This yields the marginal likelihood formula

$$\begin{aligned}L(\mathcal{D}) = &\binom{N_0}{y_0}\binom{N_1}{y_1}\mathrm{B}((1-\mu_s)n_s, \mu_e n_e + \mu_s n_s, (1-\mu_e)n_e)^{-1} \\ &\times \sum_{j=0}^{y_0+y_1}\sum_{k=0}^{N_0+N_1-y_0-y_1} 2^{-(j+k)}\binom{y_0+y_1}{j}\binom{N_0+N_1-y_0-y_1}{k}\mathrm{B}(a_{00}(j,k), a_{10}(j,k), a_{11}(j,k)),\end{aligned}$$

where we define

$$\begin{aligned}a_{00}(j,k) &= N_0 + N_1 - y_0 - y_1 + (1-\mu_s)n_s - k, \\ a_{10}(j,k) &= j + k + \mu_e n_e + \mu_s n_s, \\ a_{11}(j,k) &= y_0 + y_1 + (1-\mu_e)n_e - j.\end{aligned}$$

In Section C.4, we derive an algorithm for exact posterior sampling using the aggregated Dirichlet prior on $(p_{00}, p^*_{10}, p_{11})$.

E.2 Other priors for H_- and H_+

Another approach for specifying models for H_- and H_+, which is both natural and computationally convenient, is to impose a monotonicity assumption on M_1, and set $\eta_s = 0$ or $\eta_e = 0$ respectively. This results in the following models,

$$M_-^{'}: \ (\theta_0, \eta_e) \sim \mathrm{Beta}^*(\mu_0, n_0) \times \mathrm{Beta}^*(\mu_e, n_e), \qquad \theta_1 = (1-\eta_e)\theta_0 \tag{E.1}$$

$$M_+^{'}: \ (\theta_0, \eta_s) \sim \mathrm{Beta}^*(\mu_0, n_0) \times \mathrm{Beta}^*(\mu_s, n_s), \qquad \theta_1 = \theta_0 + \eta_s(1-\theta_0), \tag{E.2}$$

with marginal likelihoods given by

$$L_{-}^{'}(\mathcal{D}) = \binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{k=0}^{N_1-y_1}\binom{N_1-y_1}{k} \times \frac{\mathrm{B}(y_0+y_1+k+\mu_0 n_0, N-(y_0+y_1+k)+(1-\mu_0)n_0)}{\mathrm{B}(\mu_0 n_0,(1-\mu_0)n_0)} \times \frac{\mathrm{B}(k+\mu_e n_e, y_1+(1-\mu_e)n_e)}{\mathrm{B}(\mu_e n_e,(1-\mu_e)n_e)},$$

and

$$L_{+}^{'}(\mathcal{D}) = \binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{j=0}^{y_1}\binom{y_1}{j} \times \frac{\mathrm{B}(y_0+j+\mu_0 n_0, N-(y_0+j)+(1-\mu_0)n_0)}{\mathrm{B}(\mu_0 n_0,(1-\mu_0)n_0)} \times \frac{\mathrm{B}(y_1-j+\mu_s n_s, N_1-y_1+(1-\mu_s)n_s)}{\mathrm{B}(\mu_s n_s,(1-\mu_s)n_s)}.$$

Here we interpret the constraint $\eta_s = 0$ (or $\eta_e = 0$) simply as a causally principled way to derive a prior compatible with the desired constraint $H_- : \theta_1 < \theta_0$ (or $H_+ : \theta_1 > \theta_0$), and not as testing the former constraint in lieu of the latter.[21] One interesting characteristic of models $M_-^{'}$ and $M_+^{'}$ is that they do not put θ_0 and θ_1 on equal footing, even when choosing beta priors compatible with the BREASE$(1/2, \mu, \mu; 2, 1, 1)$ distribution, which places flat marginals on θ_0 and θ_1. This is usually desirable, e.g., when the control condition indeed denotes a well understood baseline, such as a standard of care. Symmetry of θ_0 and θ_1, however, can also be easily restored by switching the roles of the "treatment" and "control" conditions, as discussed in Appendix E.2. Algorithms to sample exactly from the posterior under $M_-^{'}$ and $M_+^{'}$ are provided in Appendix C.

Returning to the model $M_-^{'}$ $(E.1)$, some natural values for the prior hyperparameters are

$$\mu_0 = \mu_e = 1/2, \qquad n_0 = n_e = 2,$$

which define a flat Uniform$(0,1)^2$ prior on (θ_0, η_e). The resulting conditional prior on θ_1 is

$$\theta_1 | \theta_0 \sim \text{Uniform}(0, \theta_0),$$

which presents an intuitive representation of the hypothesis $H_- : \theta_1 < \theta_0$. Note, however,

[21] In general, the data cannot differentiate the stronger constraint, such as $\eta_s = 0$ (no one is hurt by the treatment), from the weaker constraint $\theta_1 < \theta_0$ (the treatment is beneficial on average), since the likelihood depends only on θ_1 and θ_0. Thus, in this case, differences in using M_- or $M_-^{'}$ amounts to differences only in the induced priors satisfying the same testable constraint $\theta_1 < \theta_0$, such as one placing more (or less) mass on smaller (or larger) effects than the other.

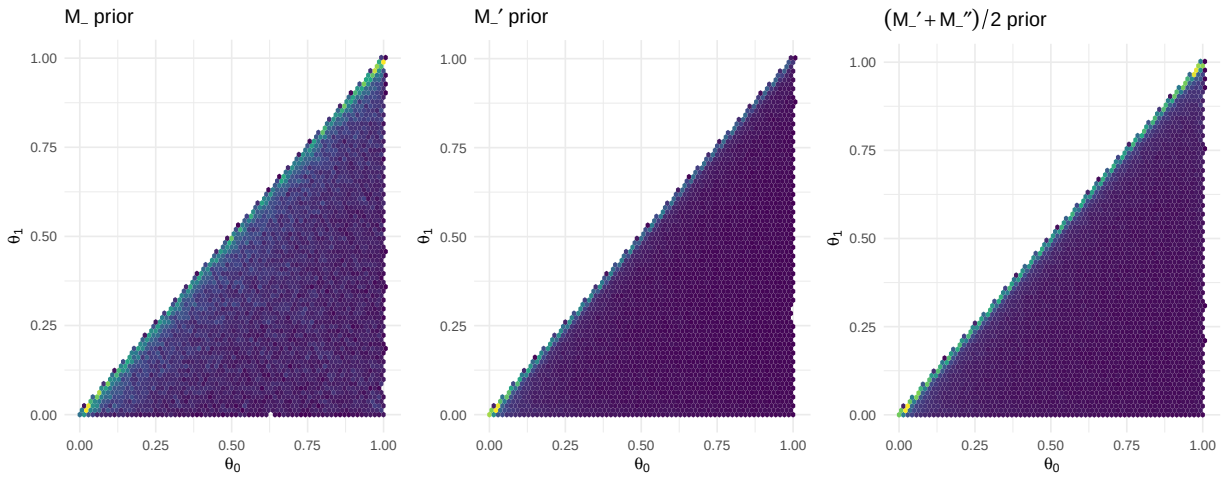


Figure 7: Left: heatmap of joint prior on (θ_0, θ_1) implied by the M_- prior (3.15) with $\mu_0 = 1/2, \mu_e = \mu_s = 0.3, n_0 = 2, n_e = n_s = 1$. Center: prior on (θ_0, θ_1) under $M_-^{'}$ with the same values of (μ_0, μ_e, n_0, n_e). Right: prior on (θ_0, θ_1) under the mixture model $(M_-^{'} + M_-^{''})/2$ with $\mu_1 = 1/2, \mu_s^{'} = 0.3, n_1 = 2, n_s^{'} = 1$ and the same values of (μ_0, μ_e, n_0, n_e).

that this specification of the model handles θ_0 as the baseline quantity. We can also go in the other direction, specifying priors on θ_1 and the "side effects of placebo" $\eta_s^{'}$ and defining

$$\theta_0 = \theta_1 + (1 - \theta_1)\eta_s^{'},$$

which also instantiates $H_- : \theta_1 < \theta_0$. We denote by $M_-^{''}$ the model

$$\begin{aligned}(\theta_1, \eta_s^{'}) &\sim \text{Beta}^*(\mu_1, n_1) \times \text{Beta}^*(\mu_s^{'}, n_s^{'}),\\ \theta_0 &= \theta_1 + (1 - \theta_1)\eta_s^{'}.\end{aligned}$$

This asymmetry in our handling of θ_0 and θ_1 is reflected in the joint priors of (θ_0, θ_1) under $M_-^{'}$ and $M_-^{''}$. As the central panel of Figure 7 exhibits, the $M_-^{'}$ joint prior tends to favor small proportions (whereas $M_-^{''}$, not plotted, favors large proportions). On the other hand, sampling $(\theta_0, \eta_e, \eta_s)$ from the BREASE prior truncated to the set $\{(\theta_0, \eta_e, \eta_s) : \theta_1 < \theta_0\}$ (i.e., the M_- prior (3.15)) yields a symmetric joint density on (θ_0, θ_1) (left panel of Figure 7). To assuage this asymmetry, we can put θ_0 and θ_1 on equal footing when testing the one-sided hypothesis H_- (and, similarly, H_+) by using a prior that averages those under $M_-^{'}$ and $M_-^{''}$, as in the right panel of Figure 7. In practice, we can decompose H_- into the submodels $M_-^{'}$ and $M_-^{''}$ and report the marginal likelihood of H_- as the average of the submodel marginal likelihoods. As the marginal likelihood under $M_-^{''}$ is also available analytically, this procedure comes with negligible added computational cost.

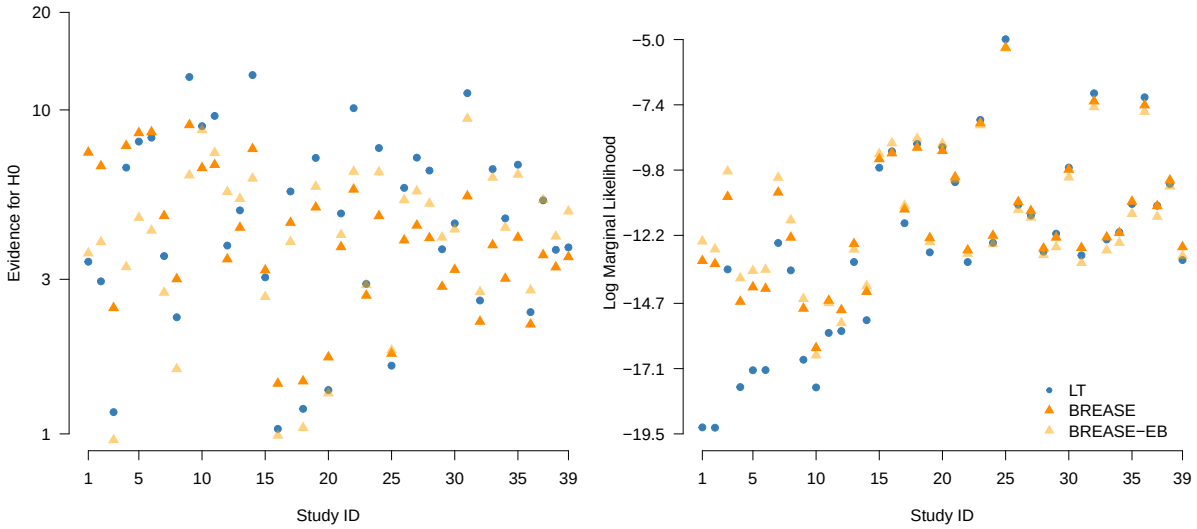


Figure 8: Comparison of Bayes factors (BF_{01}) and log marginal likelihoods under model M_1 (3.13) of the default LT, default BREASE, and empirical Bayes BREASE priors across the 39 *NEJM* studies.

E.3 An empirical Bayes prior

As η_e and η_s are counterfactual probabilities, they are not generally point-identified from data. However, since θ_0 and θ_1 are identifiable, we can derive robust bounds on their range of possible values based on the observed data (Tian and Pearl, 2000). Equation (3.3) implies the following algebraic constraints on η_e and η_s:

$$\max\left\{0, \frac{\theta_0 - \theta_1}{\theta_0}\right\} \leq \eta_e \leq \min\left\{1, \frac{1-\theta_1}{\theta_0}\right\}, \tag{E.3}$$

$$\max\left\{0, \frac{\theta_1 - \theta_0}{1-\theta_0}\right\} \leq \eta_s \leq \min\left\{1, \frac{\theta_1}{1-\theta_0}\right\}. \tag{E.4}$$

The inequalities (E.3) and (E.4) define the (marginal) partially identified regions of η_e and η_s, respectively. Denote these intervals by $I_e(\theta_0, \theta_1) = [\ell_e(\theta_0, \theta_1), u_e(\theta_0, \theta_1)]$ and $I_s(\theta_0, \theta_1) = [\ell_s(\theta_0, \theta_1), u_s(\theta_0, \theta_1)]$. In the limit of infinite data, the posterior mass of η_e and η_s will concentrate within $I_e(\theta_0^*, \theta_1^*)$ and $I_s(\theta_0^*, \theta_1^*)$, respectively, assuming θ_0^*, θ_1^* are the true values.

When conducting a Bayesian hypothesis test, a main concern is the sensitivity of Bayes factors to the prior. As demonstrated in Section 4, a prior that places unreasonable assumptions on the treatment effects can lead to questionable conclusions. In this light, it may be desired to take a data-driven approach to prior specification that concentrates prior mass near the partially identified intervals of η_e and η_s. For example, we can set the prior

means μ_e and μ_s to equal their midpoints:

$$\hat{\mu}_e = \frac{1}{2}\left(\ell_e(\hat{\theta}_0, \hat{\theta}_1) + u_e(\hat{\theta}_0, \hat{\theta}_1)\right),$$
$$\hat{\mu}_s = \frac{1}{2}\left(\ell_s(\hat{\theta}_0, \hat{\theta}_1) + u_s(\hat{\theta}_0, \hat{\theta}_1)\right),$$

where we use point estimates of the population proportions:

$$\hat{\theta}_0 = \frac{y_0 + 1}{N_0 + 2}, \qquad \hat{\theta}_1 = \frac{y_1 + 1}{N_1 + 2}.$$

As $\hat{\theta}_0$ shrinks the sample proportion toward 1/2, it avoids division by zero in (E.3) and (E.4). Hence, we might consider priors of the form BREASE$(1/2, \hat{\mu}_e, \hat{\mu}_s; 2, n, n)$ with $n \geq 0$. As this prior is estimated from the observed data, it can be thought of as an empirical Bayes approach (Robbins, 1992). As such, we denote it by BREASE-EB(n).

Note that when $n = 1$ and $\hat{\theta}_0 = \hat{\theta}_1 = 1/2$ (e.g., in the absence of data or when the sample proportions are 1/2), we obtain a vague Jeffreys marginal prior Beta$(1/2, 1/2)$ on η_e and η_s. The choice of prior sample size $n = 1$ yields something resembling a unit information prior (Kass and Wasserman, 1995), wherein the prior mean is estimated from data and its spread is chosen so that the information content of the prior matches that of a single observation.

Figure 8 compares Bayes factors (BF_{01}) and log marginal likelihoods under model M_1 (3.13) of the default LT$(0, 0; 1, 1)$, BREASE$(1/2, 0.3, 0.3; 2, 1, 1)$, and BREASE-EB(1) priors across the 39 *NEJM* studies reporting null results. The BREASE and BREASE-EB priors tend to provide the most equivocal Bayes factors on average, with mean BF_{01} equal to 4.41, 4.42, and 5.38 for the BREASE-EB, BREASE, and LT priors, respectively. However, BREASE-EB Bayes factors tend to be closer to those of the LT approach than the default BREASE prior, with mean absolute percentage differences from the LT BF_{01} of 19% for the former and 32% for the latter.

Comparing log marginal likelihoods, which quantify the predictive performance of a model, we see that the BREASE-EB and default BREASE priors perform similarly, and generally better than the default LT prior, although the default BREASE performs slightly better overall. Indeed, the default BREASE log marginal likelihood exceeds the LT in 74% of the studies compared to 59% for the BREASE-EB prior. Furthermore, the default BREASE outperforms BREASE-EB in 62% of the studies.

F Bayes factors with the IB and LT approaches

Following Dablander et al. (2022), we calculate the Bayes factor BF_{10} for the IB approach using the Savage-Dickey density ratio method applied to the difference of proportions $\eta = \theta_0 - \theta_1$ (Wagenmakers et al., 2010). A formula for the prior density of η at the null $H_0 : \eta = 0$ can be found in Appendix A of (Dablander et al., 2022). The Bayes factor using the $\text{IB}((a, a), (a, a))$ prior under H_1 as described in Section 2.2.1 is then

$$\text{BF}_{10} = \frac{\text{B}(2a - 1, 2a - 1)\text{B}(a + y_0, a + N_0 - y_0)\text{B}(a + y_1, a + N_1 - y_1)}{\text{B}(2a + y_0 + y_1 - 1, 2a + N_0 - y_0 + N_1 - y_1 - 1)\text{B}(a, a)^2}.$$

Posterior estimates and credible intervals under H_1 are calculated using exact samples from the independent beta posterior.

Bayes factors BF_{10} for the LT approach are calculating using the **abtest** package in R (Gronau, Raj, and Wagenmakers, 2021). The package uses a Laplace approximation to calculate BF_{10}, which is shown to have good performance. The LT prior under H_1 is as described in Section 2.2.2. Under $H_0 : \psi = 0$, the prior is $\beta \sim \text{Normal}(\mu_\beta, \sigma_\beta)$ with default values $(\mu_\beta, \sigma_\beta) = (0, 1)$. Posterior estimates and credible intervals under H_1 are calculated using posterior samples output by **abtest**. As **abtest** only reports marginal likelihoods up to a multiplicative constant, we used RJAGS (Plummer, 2023) to generate MCMC samples from the LT posterior and THAMES (Metodiev et al., 2023) to estimate the LT marginal likelihood for Figure 5b using the samples.