Causally sound priors for binary experiments

Nicholas J. Irons and Carlos Cinelli
University of Washington
Department of Statistics

Motivating example

Does aspirin reduce the risk of fatal heart attack?

(Frequentist Version)

Does aspirin reduce the risk of fatal heart attack?
(Frequentist Version)

Physicians' Health Study (1989)

Does aspirin reduce the risk of fatal heart attack?
(Frequentist Version)

Physicians' Health Study (1989)

- Large-scale, double-blind, randomized controlled trial, where 22,000 healthy male physicians were randomly assigned to receive 325 mg aspirin or placebo every other day.

Does aspirin reduce the risk of fatal heart attack?
(Frequentist Version)

Physicians' Health Study (1989)

- Large-scale, double-blind, randomized controlled trial, where 22,000 healthy male physicians were randomly assigned to receive 325 mg aspirin or placebo every other day.
- Treatment group: 10 out of 11,037 subjects experienced fatal heart attack.

Does aspirin reduce the risk of fatal heart attack?
(Frequentist Version)

Physicians' Health Study (1989)

- Large-scale, double-blind, randomized controlled trial, where 22,000 healthy male physicians were randomly assigned to receive 325 mg aspirin or placebo every other day.
- Treatment group: 10 out of 11,037 subjects experienced fatal heart attack.
- Control group: 26 out of 11,034 subjects. Is this strong enough evidence to claim that aspirin works?

Does aspirin reduce the risk of fatal heart attack?
(Frequentist Version)

Physicians' Health Study (1989)

- Large-scale, double-blind, randomized controlled trial, where 22,000 healthy male physicians were randomly assigned to receive 325 mg aspirin or placebo every other day.
- Treatment group: 10 out of 11,037 subjects experienced fatal heart attack.
- Control group: 26 out of 11,034 subjects. Is this strong enough evidence to claim that aspirin works?
- Estimated risk ratio (θ_1/θ_0): 0.39 [0.19, 0.80], rejects the null hypothesis of zero effect with a p-value of ~0.008.

Does aspirin reduce the risk of fatal heart attack?
(Frequentist Version)

Physicians' Health Study (1989)

- Large-scale, double-blind, randomized controlled trial, where 22,000 healthy male physicians were randomly assigned to receive 325 mg aspirin or placebo every other day.
- Treatment group: 10 out of 11,037 subjects experienced fatal heart attack.
- Control group: 26 out of 11,034 subjects. Is this strong enough evidence to claim that aspirin works?
- Estimated risk ratio (θ_1/θ_0): 0.39 [0.19, 0.80], rejects the null hypothesis of zero effect with a p-value of ~0.008.

Frequentist: yes!

Does aspirin reduce the risk of fatal heart attack?

(Bayes Version)

Does aspirin reduce the risk of fatal heart attack?
(Bayes Version)

Bayes factors using current main default approaches:

Does aspirin reduce the risk of fatal heart attack?
(Bayes Version)

Bayes factors using current main default approaches:

- Independent beta (IB) approach using beta(1,1) priors:
  - Bayes factor BF_{01} = 20.27
  - ***Strong*** evidence in favor of the null (H_0: $\theta_0 = \theta_1$)

Does aspirin reduce the risk of fatal heart attack?
(Bayes Version)

Bayes factors using current main default approaches:

- Independent beta (IB) approach using beta(1,1) priors:
  - Bayes factor $BF_{01} = 20.27$
  - ***Strong*** evidence in favor of the null (H_0: $\theta_0 = \theta_1$)

- Logit transform (LT) approach using N(0,1) priors:
  - Bayes factor $BF_{10} = 5.24$
  - ***Moderate*** evidence in favor of the alternative (H_1: $\theta_0 \neq \theta_1$)

Does aspirin reduce the risk of fatal heart attack?
(Bayes Version)

Bayes factors using current main default approaches:

- Independent beta (IB) approach using beta(1,1) priors:
  - Bayes factor $BF_{01} = 20.27$
  - ***Strong*** evidence in favor of the null (H_0: $\theta_0 = \theta_1$)

- Logit transform (LT) approach using N(0,1) priors:
  - Bayes factor $BF_{10} = 5.24$
  - ***Moderate*** evidence in favor of the alternative (H_1: $\theta_0 \neq \theta_1$)

Bayesian:?

How can we make sense of these results?

How can we make sense of these results?

- Bayesian inference (in particular Bayes factors) can sometimes be highly sensitive to the prior.

How can we make sense of these results?

- Bayesian inference (in particular Bayes factors) can sometimes be highly sensitive to the prior.
- The IB and LT approaches encode implicit assumptions that do not align with the causal structure of an RCT and are unlikely to reflect reasonable prior beliefs about the joint distribution of risks.

How can we make sense of these results?

- Bayesian inference (in particular Bayes factors) can sometimes be highly sensitive to the prior.
- The IB and LT approaches encode implicit assumptions that do not align with the causal structure of an RCT and are unlikely to reflect reasonable prior beliefs about the joint distribution of risks.
- Taking a causal inference perspective, we frame the analysis of binary experiments in terms of clinically meaningful parameters: the **b**aseline **r**isk, **e**fficacy, and **a**dverse **s**ide **e**ffects (**BREASE**) of treatment.

How can we make sense of these results?

- Bayesian inference (in particular Bayes factors) can sometimes be highly sensitive to the prior.
- The IB and LT approaches encode implicit assumptions that do not align with the causal structure of an RCT and are unlikely to reflect reasonable prior beliefs about the joint distribution of risks.
- Taking a causal inference perspective, we frame the analysis of binary experiments in terms of clinically meaningful parameters: the **b**aseline **r**isk, **e**fficacy, and **a**dverse **s**ide **e**ffects (**BREASE**) of treatment.
- This allows us to naturally induce priors on the joint distribution of risks in the treatment and control groups that respect the underlying causal structure of the problem.

How can we make sense of these results?

- Bayesian inference (in particular Bayes factors) can sometimes be highly sensitive to the prior.
- The IB and LT approaches encode implicit assumptions that do not align with the causal structure of an RCT and are unlikely to reflect reasonable prior beliefs about the joint distribution of risks.
- Taking a causal inference perspective, we frame the analysis of binary experiments in terms of clinically meaningful parameters: the **b**aseline **r**isk, **e**fficacy, and **a**dverse **s**ide **e**ffects (**BREASE**) of treatment.
- This allows us to naturally induce priors on the joint distribution of risks in the treatment and control groups that respect the underlying causal structure of the problem.
- Our proposal can:
    - help analysts better distinguish robust from fragile findings;
    - clarify what one needs to believe in order to claim that a treatment is effective;
    - reconcile disparate results obtained from different methods.

Preliminaries

Notation and assumptions

Notation and assumptions

- **Definitions**:
  - Let $N = N_1 + N_0$ denote the total number of subjects in the study;
  - Z_i is a binary treatment (Z=1, treat.), Y_i is a binary outcome (Y=1, death);
  - $Y_i(z)$ is the potential outcome of subject *i* under experimental condition *z*

Notation and assumptions

- **Definitions**:
  - Let $N = N_1 + N_0$ denote the total number of subjects in the study;
  - Z_i is a binary treatment (Z=1, treat.), Y_i is a binary outcome (Y=1, death);
  - $Y_i(z)$ is the potential outcome of subject *i* under experimental condition ***z***

- **Assumption 1 (*Consistency*):** $Y_i = Y_i(Z_i)$.

Notation and assumptions

- **Definitions**:
  - Let $N = N_1 + N_0$ denote the total number of subjects in the study;
  - Z_i is a binary treatment (Z=1, treat.), Y_i is a binary outcome (Y=1, death);
  - $Y_i(z)$ is the potential outcome of subject *i* under experimental condition ***z***

- **Assumption 1 (*Consistency*):** $Y_i = Y_i(Z_i)$.

- **Assumption 2 (*Super-population*):** IID draws from a common population.

Notation and assumptions

- **Definitions**:
  - Let $N = N_1 + N_0$ denote the total number of subjects in the study;
  - Z_i is a binary treatment (Z=1, treat.), Y_i is a binary outcome (Y=1, death);
  - $Y_i(z)$ is the potential outcome of subject *i* under experimental condition ***z***

- **Assumption 1 (*Consistency*):** $Y_i = Y_i(Z_i)$.

- **Assumption 2 (*Super-population*):** IID draws from a common population.

- **Assumption 3 (*Ignorability*):** $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp Z_i$ (Holds in RCTs).

Likelihood

- Under A1-3, observed counts follow independent binomial distributions:

$$y_0 = \sum_{i:Z_i=0} Y_i \sim \text{Binomial}(N_0, \theta_0) \quad \perp\!\!\!\perp \quad y_1 = \sum_{i:Z_i=1} Y_i \sim \text{Binomial}(N_1, \theta_1)$$

Likelihood

- Under A1-3, observed counts follow independent binomial distributions:

$$y_0 = \sum_{i:Z_i=0} Y_i \sim \text{Binomial}(N_0, \theta_0) \quad \perp\!\!\!\perp \quad y_1 = \sum_{i:Z_i=1} Y_i \sim \text{Binomial}(N_1, \theta_1)$$

where we define the ***baseline risk*** and ***risk of treatment***, respectively,

$$\theta_0 = \mathbb{P}(Y_i(0) = 1), \quad \theta_1 = \mathbb{P}(Y_i(1) = 1)$$

Likelihood

- Under A1-3, observed counts follow independent binomial distributions:

$$y_0 = \sum_{i:Z_i=0} Y_i \sim \text{Binomial}(N_0, \theta_0) \quad \perp\!\!\!\perp \quad y_1 = \sum_{i:Z_i=1} Y_i \sim \text{Binomial}(N_1, \theta_1)$$

  where we define the ***baseline risk*** and ***risk of treatment***, respectively,

$$\theta_0 = \mathbb{P}(Y_i(0) = 1), \quad \theta_1 = \mathbb{P}(Y_i(1) = 1)$$

- Bayesian inference is carried out using the joint posterior distribution of the risks (θ_0,θ_1), which requires specification of a joint prior on (θ_0,θ_1).

Likelihood

- Under A1-3, observed counts follow independent binomial distributions:

$$y_0 = \sum_{i:Z_i=0} Y_i \sim \text{Binomial}(N_0, \theta_0) \quad \perp\!\!\!\perp \quad y_1 = \sum_{i:Z_i=1} Y_i \sim \text{Binomial}(N_1, \theta_1)$$

  where we define the ***baseline risk*** and ***risk of treatment***, respectively,

$$\theta_0 = \mathbb{P}(Y_i(0) = 1), \quad \theta_1 = \mathbb{P}(Y_i(1) = 1)$$

- Bayesian inference is carried out using the joint posterior distribution of the risks (θ_0,θ_1), which requires specification of a joint prior on (θ_0,θ_1).
- Two main approaches: (i) Independent Beta and (ii) Logistic transformation.

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

- Main advantage: simplicity

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

- Main advantage: simplicity
  - Estimation and posterior simulation can be carried out exactly (no MCMC required)

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

- Main advantage: simplicity
  - Estimation and posterior simulation can be carried out exactly (no MCMC required)
  - Marginal likelihood and Bayes factors available analytically

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

- Main advantage: simplicity
  - Estimation and posterior simulation can be carried out exactly (no MCMC required)
  - Marginal likelihood and Bayes factors available analytically

- Significant Drawback:

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

- Main advantage: simplicity
  - Estimation and posterior simulation can be carried out exactly (no MCMC required)
  - Marginal likelihood and Bayes factors available analytically

- Significant Drawback:
  - We would expect knowledge about risks in the treatment and control groups to be dependent, e.g., disease prevalence in placebo and vaccine arms.

Independent beta (IB) approach

- Under the IB approach, we assign the prior

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

  - A common default choice is the Beta(1,1), a flat uniform prior.

- Main advantage: simplicity
  - Estimation and posterior simulation can be carried out exactly (no MCMC required)
  - Marginal likelihood and Bayes factors available analytically

- Significant Drawback:
  - We would expect knowledge about risks in the treatment and control groups to be dependent, e.g., disease prevalence in placebo and vaccine arms.
  - The IB prior fails to accommodate this dependence, (θ_0,θ_1) are independent *a priori and a posteriori*!

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

  - Common default choice is N(0,1), the standard normal prior. β is the grand log odds and ψ is the log odds ratio. Reparameterization of a logistic regression.

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

  - Common default choice is N(0,1), the standard normal prior. β is the grand log odds and ψ is the log odds ratio. Reparameterization of a logistic regression.

- Main advantage: induces dependence across treatment arms

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

  - Common default choice is N(0,1), the standard normal prior. β is the grand log odds and ψ is the log odds ratio. Reparameterization of a logistic regression.

- Main advantage: induces dependence across treatment arms
  - Encodes correlation in (θ_0,θ_1) through their shared dependence on (β,ψ).

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

  - Common default choice is N(0,1), the standard normal prior. β is the grand log odds and ψ is the log odds ratio. Reparameterization of a logistic regression.

- Main advantage: induces dependence across treatment arms
  - Encodes correlation in (θ_0,θ_1) through their shared dependence on (β,ψ).
- Significant drawbacks:

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

  - Common default choice is N(0,1), the standard normal prior. β is the grand log odds and ψ is the log odds ratio. Reparameterization of a logistic regression.

- Main advantage: induces dependence across treatment arms
  - Encodes correlation in (θ_0,θ_1) through their shared dependence on (β,ψ).
- Significant drawbacks:
  - Less intuitive parametrization. Odds ratios are notoriously difficult to reason about.

Logit transform (LT) approach (logistic regression)

- Assign independent normal priors to the parameters (β,ψ) satisfying

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

  - Common default choice is N(0,1), the standard normal prior. β is the grand log odds and ψ is the log odds ratio. Reparameterization of a logistic regression.

- Main advantage: induces dependence across treatment arms
  - Encodes correlation in (θ_0,θ_1) through their shared dependence on (β,ψ).
- Significant drawbacks:
  - Less intuitive parametrization. Odds ratios are notoriously difficult to reason about.
  - Not analytically tractable. Posterior sampling and marginal likelihood calculation must be carried out approximately (e.g., using MCMC or Laplace approximation).

The BREASE framework

Baseline Risk, Efficacy, and Adverse Side Effects

- 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:

$$\theta_1 = \theta_0(1 - \eta_e) + (1 - \theta_0)\eta_s$$

Baseline Risk, Efficacy, and Adverse Side Effects

- 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:

$$\theta_1 = \theta_0(1 - \eta_e) + (1 - \theta_0)\eta_s$$

- Here the *efficacy* of treatment is the probability that treatment is sufficient to save or cure a patient: $\eta_e = \mathbb{P}(Y_i(1) = 0 | Y_i(0) = 1)$

Baseline Risk, Efficacy, and Adverse Side Effects

- 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:

$$\theta_1 = \theta_0(1 - \eta_e) + (1 - \theta_0)\eta_s$$

- Here the *efficacy* of treatment is the probability that treatment is sufficient to save or cure a patient: $\eta_e = \mathbb{P}(Y_i(1) = 0 | Y_i(0) = 1)$

- Similarly, the *risk of adverse side effects* is the probability that treatment causes death of an otherwise healthy patient: $\eta_s = \mathbb{P}(Y_i(1) = 1 | Y_i(0) = 0)$

Baseline Risk, Efficacy, and Adverse Side Effects

- 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:

$$\theta_1 = \theta_0(1 - \eta_e) + (1 - \theta_0)\eta_s$$

- Here the *efficacy* of treatment is the probability that treatment is sufficient to save or cure a patient: $\eta_e = \mathbb{P}(Y_i(1) = 0 | Y_i(0) = 1)$
- Similarly, the *risk of adverse side effects* is the probability that treatment causes death of an otherwise healthy patient: $\eta_s = \mathbb{P}(Y_i(1) = 1 | Y_i(0) = 0)$
- Parameterization highlights the natural dependence between θ_0 and θ_1 that is easy to miss without framing the problem in the language of potential outcomes

Partial identification

- The counterfactual parameters η_{e} and η_{s} are only partially identified by the observed data,

Partial identification

- The counterfactual parameters η_e and η_s are only partially identified by the observed data,

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

Partial identification

- The counterfactual parameters η_e and η_s are only partially identified by the observed data,

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

- However, partial identification of (η_e, η_s) does not affect that of (θ_0,θ_1).

Partial identification

- The counterfactual parameters η_e and η_s are only partially identified by the observed data,

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

- However, partial identification of (η_e, η_s) does not affect that of (θ_0,θ_1).
- We can think of placing priors on (η_e, η_s) as a way to induce sensible priors on (θ_0,θ_1), respecting the causal nature of the data generating process.

Partial identification

- The counterfactual parameters η_e and η_s are only partially identified by the observed data,

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

- However, partial identification of (η_e, η_s) does not affect that of (θ_0,θ_1).
- We can think of placing priors on (η_e, η_s) as a way to induce sensible priors on (θ_0,θ_1), respecting the causal nature of the data generating process.
- If $\eta_s = 0$, then is η_e is indeed point identified and corresponds to what is commonly known as "efficacy" in the clinical trials literature.

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

- Here, the Beta*(μ,n) is parametrized to have mean μ, prior “sample size” n.

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

- Here, the Beta*(μ,n) is parametrized to have mean μ, prior “sample size” n.
- We refer to it as the BREASE(μ,n) prior, with $\mu = (\mu_0,\mu_e,\mu_s)$ and $n = (n_0,n_e,n_s)$.

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

- Here, the Beta*(μ,n) is parametrized to have mean μ, prior “sample size” n.
- We refer to it as the BREASE(μ,n) prior, with μ = (μ_0,μ_e,μ_s) and n = (n_0,n_e,n_s).
- Features of our approach:

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

- Here, the Beta*(μ,n) is parametrized to have mean μ, prior “sample size” n.
- We refer to it as the BREASE(μ,n) prior, with $\mu = (\mu_0,\mu_e,\mu_s)$ and $n = (n_0,n_e,n_s)$.
- Features of our approach:
  - Naturally induces dependence in (θ_0,θ_1) *a priori*.

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

- Here, the Beta*(μ,n) is parametrized to have mean μ, prior “sample size” n.
- We refer to it as the BREASE(μ,n) prior, with μ = (μ_0,μ_e,μ_s) and n = (n_0,n_e,n_s).
- Features of our approach:
  - Naturally induces dependence in (θ_0,θ_1) *a priori*.
  - Tractable: analytical Bayes factor formulae, exact posterior sampling, and accurate data-augmentation posterior sampling algorithm.

Prior specification

- We propose placing independent beta priors on the BREASE parameters:

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

- Here, the Beta*(μ,n) is parametrized to have mean μ, prior “sample size” n.
- We refer to it as the BREASE(μ,n) prior, with $\mu = (\mu_0,\mu_e,\mu_s)$ and $n = (n_0,n_e,n_s)$.
- Features of our approach:
  - Naturally induces dependence in (θ_0,θ_1) *a priori*.
  - Tractable: analytical Bayes factor formulae, exact posterior sampling, and accurate data-augmentation posterior sampling algorithm.
  - Simplifies prior elicitation and sensitivity analysis.

Induced distributions

- We show that the BREASE prior induces a *Generalized* Dirichlet prior on the response types (Dickey, 1983).
  - The regular Dirichlet prior is a common choice in the Bayesian analysis of RCTs with non-compliance and is a special case of the BREASE prior.

Induced distributions

- We show that the BREASE prior induces a *Generalized* Dirichlet prior on the response types (Dickey, 1983).
  - The regular Dirichlet prior is a common choice in the Bayesian analysis of RCTs with non-compliance and is a special case of the BREASE prior.
- We derive the induced distribution on (θ_0, θ_1) and several of its properties:

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

When prior expected effect of treatment is small, risks in the treatment and control groups are positively correlated *a priori*, which often aligns with clinical expectations.

Default prior

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

- The induced prior on (θ_0,θ_1) has desirable properties that capture the best features of the default IB and LT priors:

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

- The induced prior on (θ_0,θ_1) has desirable properties that capture the best features of the default IB and LT priors:
  - puts flat uniform marginals on θ_0 and θ_1 (as with the IB approach)

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

- The induced prior on (θ_0,θ_1) has desirable properties that capture the best features of the default IB and LT priors:
  - puts flat uniform marginals on θ_0 and θ_1 (as with the IB approach)
  - induces prior correlation between parameters (as with the LT approach)

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

- The induced prior on (θ_0,θ_1) has desirable properties that capture the best features of the default IB and LT priors:
  - puts flat uniform marginals on θ_0 and θ_1 (as with the IB approach)
  - induces prior correlation between parameters (as with the LT approach)
  - assumes no effect on average (as with the IB and LT approaches)

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

- The induced prior on (θ_0,θ_1) has desirable properties that capture the best features of the default IB and LT priors:
  - puts flat uniform marginals on θ_0 and θ_1 (as with the IB approach)
  - induces prior correlation between parameters (as with the LT approach)
  - assumes no effect on average (as with the IB and LT approaches)
  - depends on a single parameter μ denoting the expected benefits or harm of treatment.

Default prior

- We propose the following as a default: BREASE(½, μ, μ; 2, 1, 1)

$$\theta_0 \sim \text{Beta}^*(1/2, 2), \quad \eta_e \sim \text{Beta}^*(\mu, 1), \quad \eta_s \sim \text{Beta}^*(\mu, 1)$$

- The induced prior on (θ_0,θ_1) has desirable properties that capture the best features of the default IB and LT priors:
  - puts flat uniform marginals on θ_0 and θ_1 (as with the IB approach)
  - induces prior correlation between parameters (as with the LT approach)
  - assumes no effect on average (as with the IB and LT approaches)
  - depends on a single parameter μ denoting the expected benefits or harm of treatment.
  - We recommend μ = 0.3: expects moderate effects, positive correlation of risks, and concentrates mass on the diagonal $\theta_0 = \theta_1$.

Posterior sampling

Posterior sampling

The BREASE posterior is given by a mixture of independent betas

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

Posterior sampling

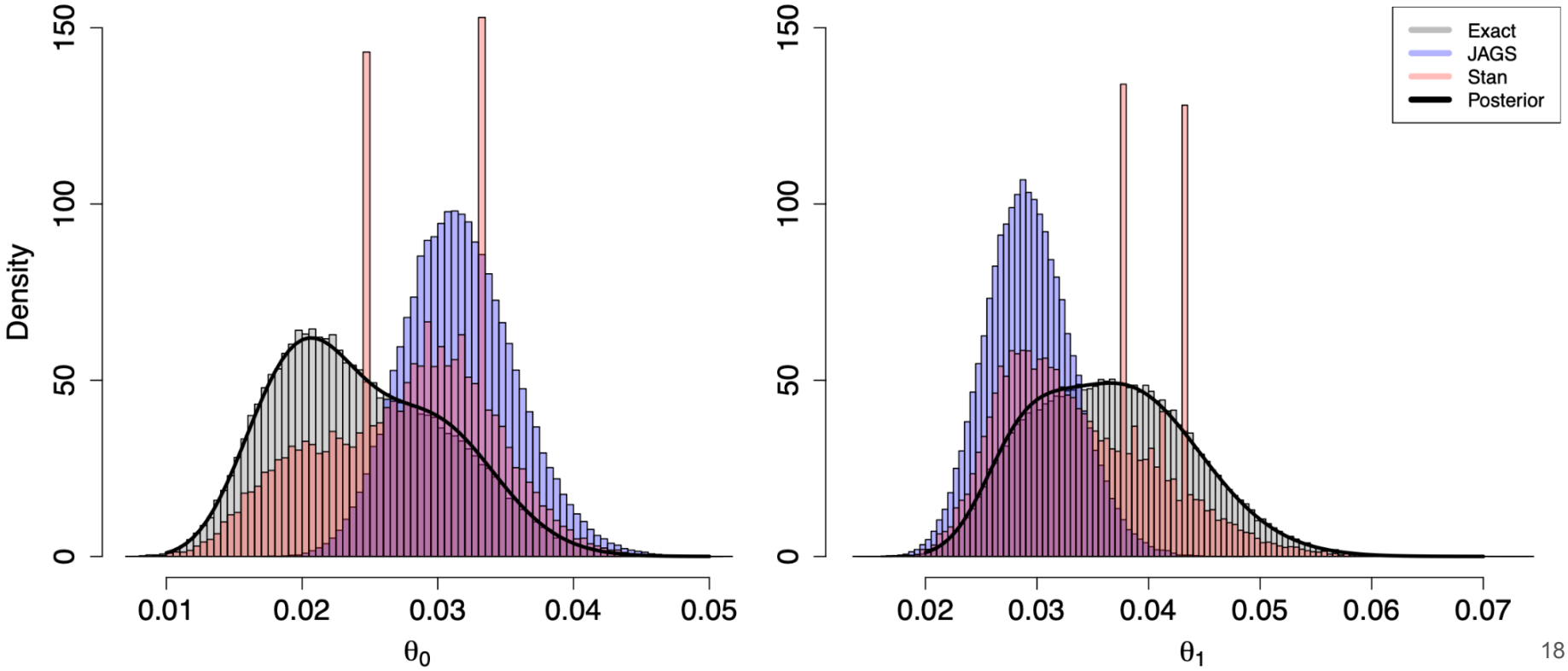
The BREASE posterior is given by a mixture of independent betas

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

We can thus sample exactly via simulation, by first sampling the mixture weights, and then sampling from the independent beta distributions.

We also derive an efficient and accurate data-augmentation algorithm.

Posterior sampling



Marginal likelihoods and Bayes factors

- We also provide analytical formulae for marginal likelihoods and Bayes factors

Theorem 3. *The marginal likelihood of the data under M_0 is given by a simple beta-binomial distribution. Under M_1, it is given by the following weighted sum of beta functions*[8]

$$
\begin{aligned}
L_1(\mathcal{D}) = &\binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{j=0}^{y_1}\sum_{k=0}^{N_1-y_1}\binom{y_1}{j}\binom{N_1-y_1}{k} \\
&\times \frac{B(y_0+j+k+\mu_0 n_0, N-(y_0+j+k)+(1-\mu_0)n_0)}{B(\mu_0 n_0,(1-\mu_0)n_0)} \\
&\times \frac{B(k+\mu_e n_e, j+(1-\mu_e)n_e)}{B(\mu_e n_e,(1-\mu_e)n_e)} \\
&\times \frac{B(y_1-j+\mu_s n_s, N_1-y_1-k+(1-\mu_s)n_s)}{B(\mu_s n_s,(1-\mu_s)n_s)}. \qquad (3.17)
\end{aligned}
$$

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

$$
L_-(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1<\theta_0|\mathcal{D})}{\pi_1(\theta_1<\theta_0)}, \qquad L_+(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1>\theta_0|\mathcal{D})}{\pi_1(\theta_1>\theta_0)}. \qquad (3.18)
$$

Marginal likelihoods and Bayes factors

- We also provide analytical formulae for marginal likelihoods and Bayes factors
- Hypothesis testing done automatically

Theorem 3. *The marginal likelihood of the data under M_0 is given by a simple beta-binomial distribution. Under M_1, it is given by the following weighted sum of beta functions*[8]

$$
\begin{aligned}
L_1(\mathcal{D}) = &\binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{j=0}^{y_1}\sum_{k=0}^{N_1-y_1}\binom{y_1}{j}\binom{N_1-y_1}{k} \\
&\times \frac{B(y_0+j+k+\mu_0 n_0, N-(y_0+j+k)+(1-\mu_0)n_0)}{B(\mu_0 n_0,(1-\mu_0)n_0)} \\
&\times \frac{B(k+\mu_e n_e, j+(1-\mu_e)n_e)}{B(\mu_e n_e,(1-\mu_e)n_e)} \\
&\times \frac{B(y_1-j+\mu_s n_s, N_1-y_1-k+(1-\mu_s)n_s)}{B(\mu_s n_s,(1-\mu_s)n_s)}. \qquad (3.17)
\end{aligned}
$$

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

$$
L_-(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 < \theta_0|\mathcal{D})}{\pi_1(\theta_1 < \theta_0)}, \qquad L_+(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 > \theta_0|\mathcal{D})}{\pi_1(\theta_1 > \theta_0)}. \qquad (3.18)
$$

Marginal likelihoods and Bayes factors

- We also provide analytical formulae for marginal likelihoods and Bayes factors
- Hypothesis testing done automatically
  - No MCMC required (see last slide)

Theorem 3. *The marginal likelihood of the data under M_0 is given by a simple beta-binomial distribution. Under M_1, it is given by the following weighted sum of beta functions*[8]

$$
\begin{aligned}
L_1(\mathcal{D}) = &\binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{j=0}^{y_1}\sum_{k=0}^{N_1-y_1}\binom{y_1}{j}\binom{N_1-y_1}{k} \\
&\times \frac{B(y_0+j+k+\mu_0 n_0, N-(y_0+j+k)+(1-\mu_0)n_0)}{B(\mu_0 n_0,(1-\mu_0)n_0)} \\
&\times \frac{B(k+\mu_e n_e, j+(1-\mu_e)n_e)}{B(\mu_e n_e,(1-\mu_e)n_e)} \\
&\times \frac{B(y_1-j+\mu_s n_s, N_1-y_1-k+(1-\mu_s)n_s)}{B(\mu_s n_s,(1-\mu_s)n_s)}. \qquad (3.17)
\end{aligned}
$$

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

$$
L_-(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 < \theta_0|\mathcal{D})}{\pi_1(\theta_1 < \theta_0)}, \qquad L_+(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 > \theta_0|\mathcal{D})}{\pi_1(\theta_1 > \theta_0)}. \qquad (3.18)
$$

Marginal likelihoods and Bayes factors

- We also provide analytical formulae for marginal likelihoods and Bayes factors
- Hypothesis testing done automatically
  - No MCMC required (see last slide)
- Simple and fast sensitivity analysis.

Theorem 3. *The marginal likelihood of the data under M_0 is given by a simple beta-binomial distribution. Under M_1, it is given by the following weighted sum of beta functions*[8]

$$
\begin{aligned}
L_1(\mathcal{D}) = &\binom{N_0}{y_0}\binom{N_1}{y_1}\sum_{j=0}^{y_1}\sum_{k=0}^{N_1-y_1}\binom{y_1}{j}\binom{N_1-y_1}{k} \\
&\times \frac{B(y_0+j+k+\mu_0 n_0, N-(y_0+j+k)+(1-\mu_0)n_0)}{B(\mu_0 n_0,(1-\mu_0)n_0)} \\
&\times \frac{B(k+\mu_e n_e, j+(1-\mu_e)n_e)}{B(\mu_e n_e,(1-\mu_e)n_e)} \\
&\times \frac{B(y_1-j+\mu_s n_s, N_1-y_1-k+(1-\mu_s)n_s)}{B(\mu_s n_s,(1-\mu_s)n_s)}. \qquad (3.17)
\end{aligned}
$$

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

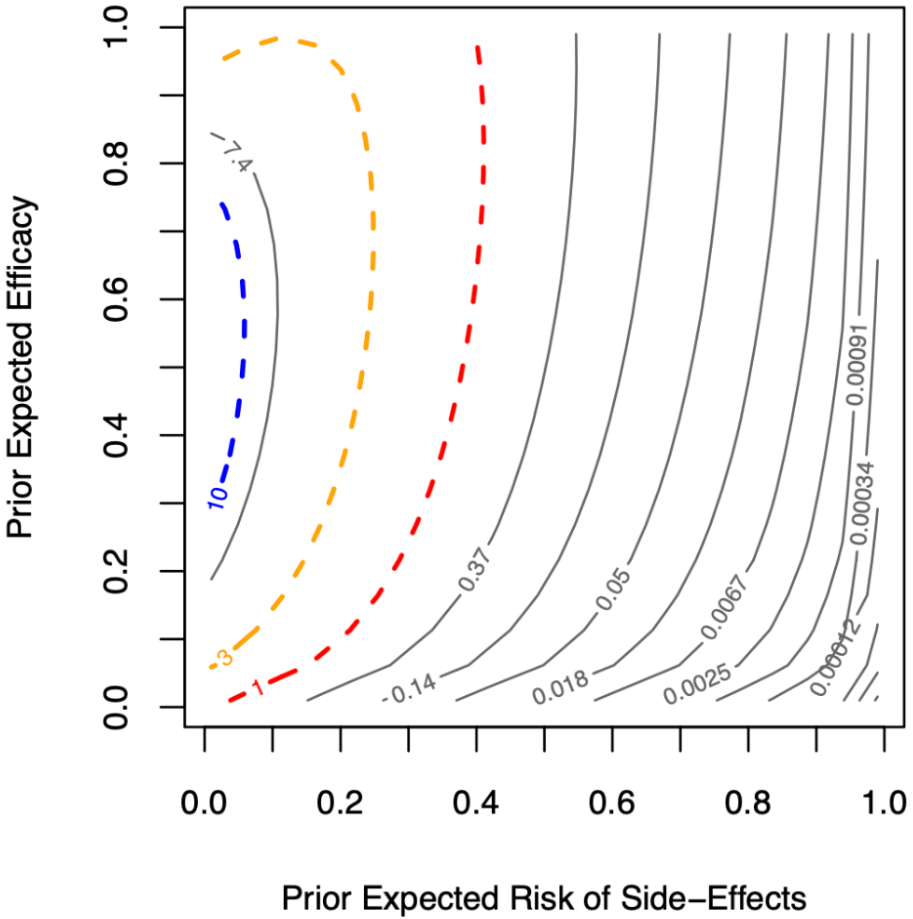
$$
L_-(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 < \theta_0|\mathcal{D})}{\pi_1(\theta_1 < \theta_0)}, \qquad L_+(\mathcal{D}) = L_1(\mathcal{D}) \times \frac{\pi_1(\theta_1 > \theta_0|\mathcal{D})}{\pi_1(\theta_1 > \theta_0)}. \qquad (3.18)
$$

Back to our example

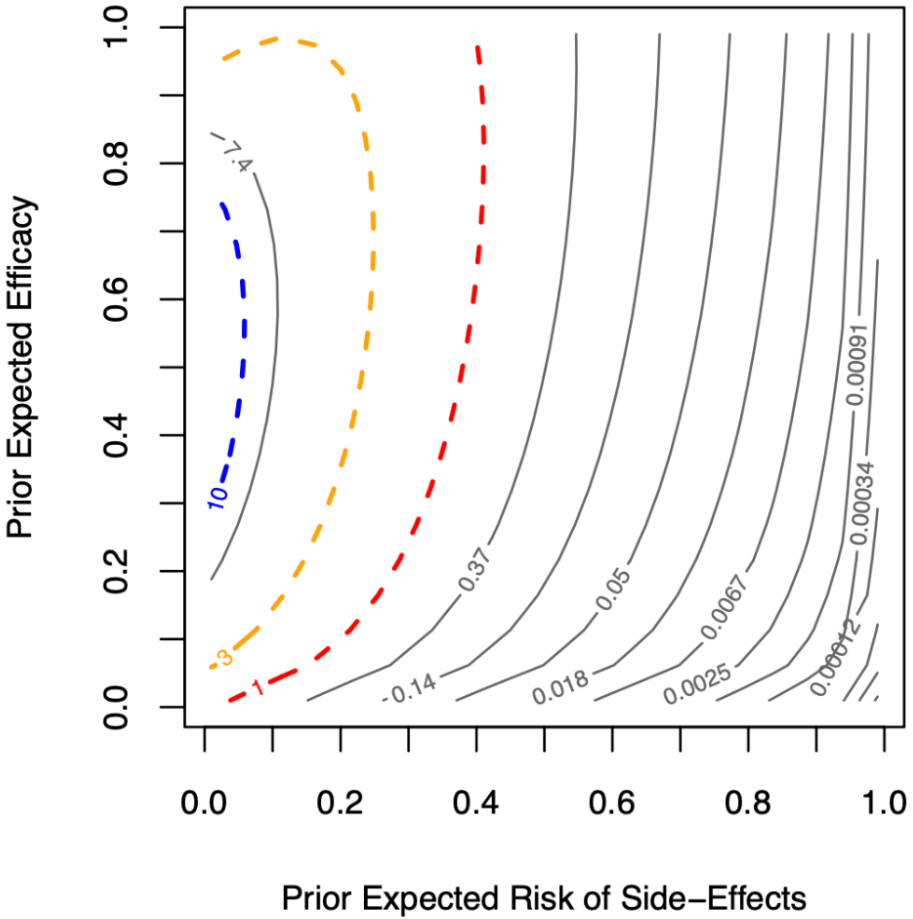
A catalog of results

- Frequentist approach: P = 0.008
  - **Reject the null** (H_0: $\theta_0 = \theta_1$)
- Default IB: $BF_{01} = 20.27$
  - ***Strong*** evidence in favor of the null
- Default LT: $BF_{10} = 5.24$
  - ***Moderate*** evidence in favor of the alternative (H_1: $\theta_0 \neq \theta_1$)
- Default BREASE: $BF_{10} = 1.2$
  - ***Equivocal*** evidence between H_0 and H_1

Prior sensitivity of the BF (Aspirin study)



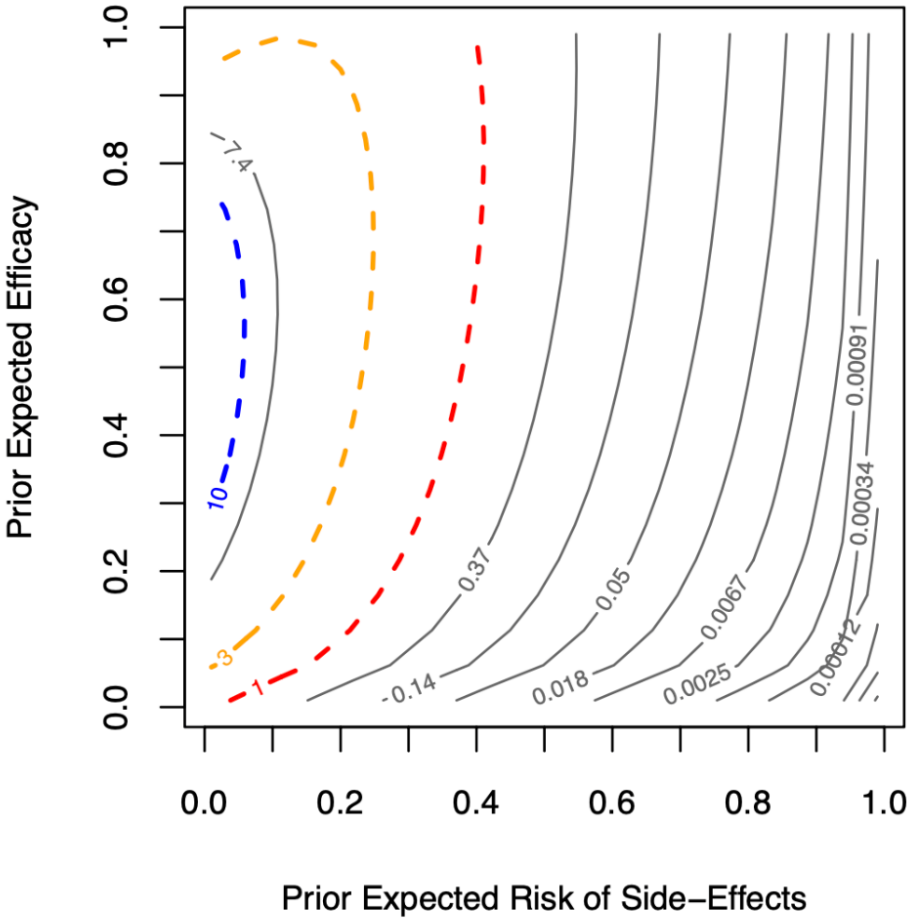
Prior sensitivity of the BF (Aspirin study)



Strong evidence against the null only when:

- adverse side effects are expected to be small (< 1%), and
- efficacy is expected to be relatively large (between 30% and 70%).

Prior sensitivity of the BF (Aspirin study)

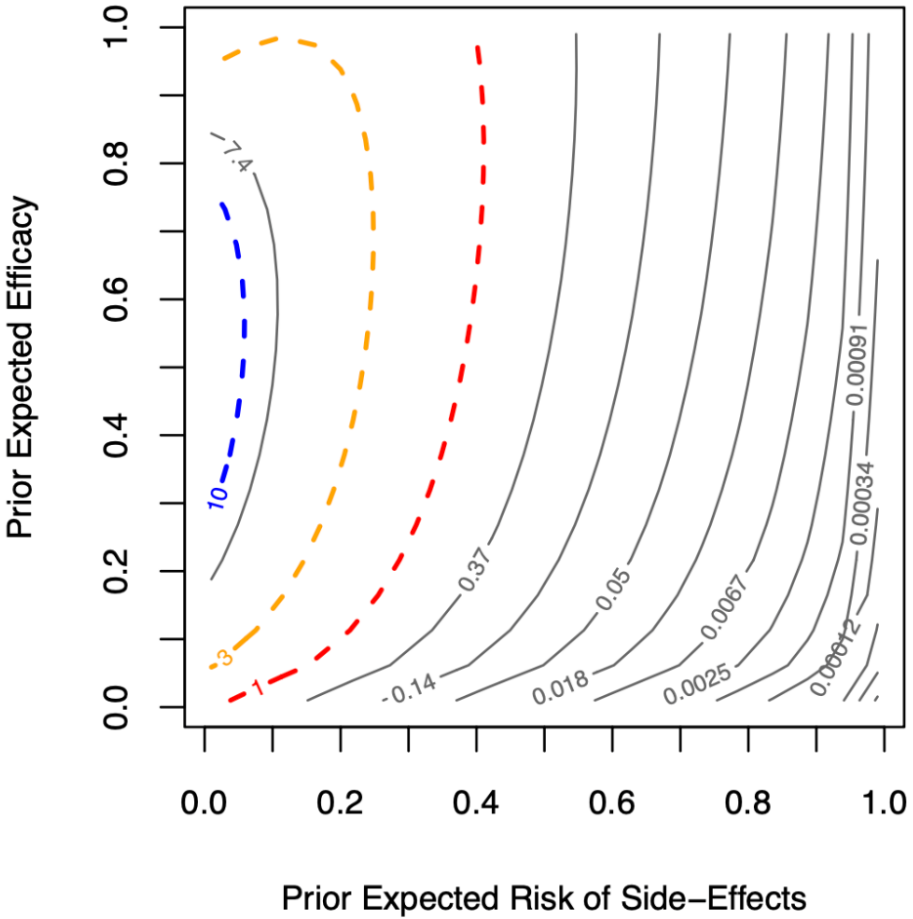


Strong evidence against the null only when:

- adverse side effects are expected to be small (< 1%), and
- efficacy is expected to be relatively large (between 30% and 70%).

For other hyperparameters, the evidence is moderate, weak, or favors the null.

Prior sensitivity of the BF (Aspirin study)



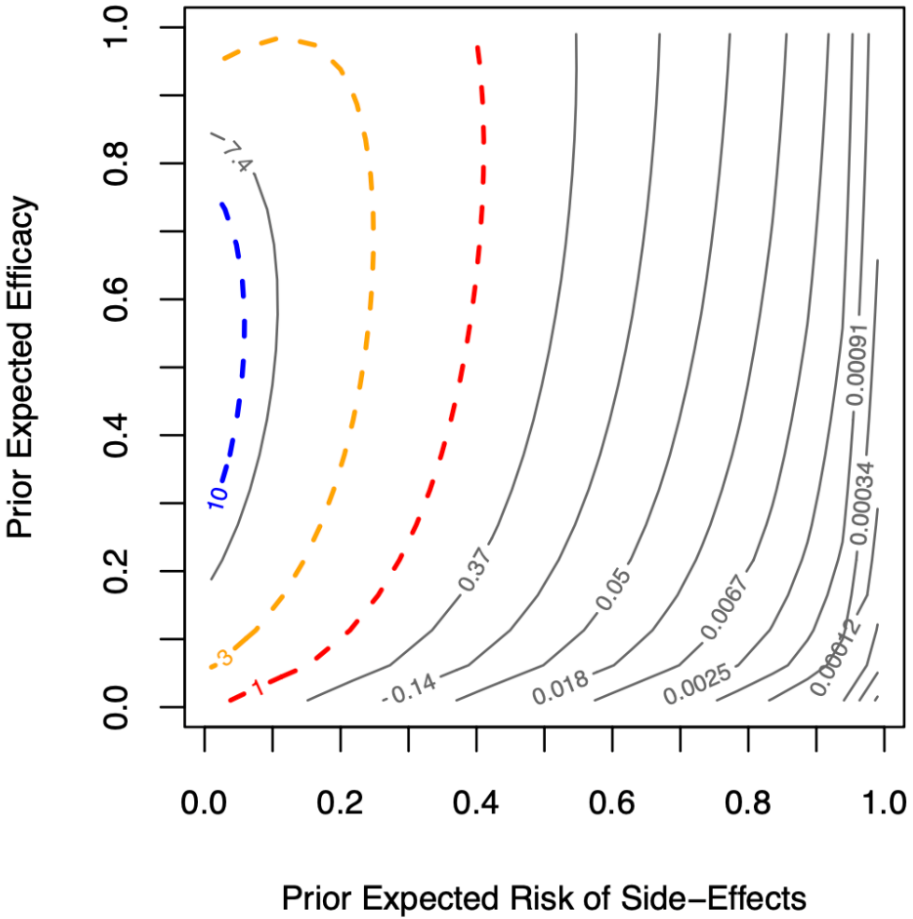
Strong evidence against the null only when:

- adverse side effects are expected to be small (< 1%), and
- efficacy is expected to be relatively large (between 30% and 70%).

For other hyperparameters, the evidence is moderate, weak, or favors the null.

The results of the trial are ambiguous: the conclusion that aspirin prevents fatal heart attack strongly depends on prior expected efficacy.

Prior sensitivity of the BF (Aspirin study)



Strong evidence against the null only when:

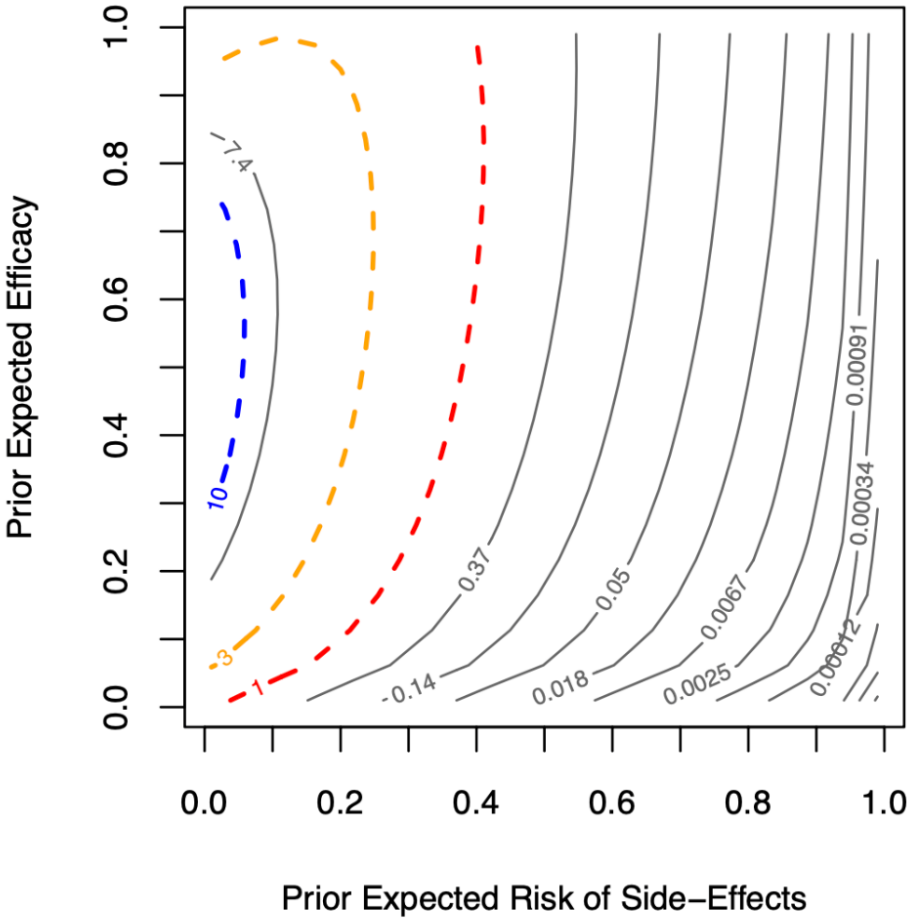
- adverse side effects are expected to be small (< 1%), and
- efficacy is expected to be relatively large (between 30% and 70%).

For other hyperparameters, the evidence is moderate, weak, or favors the null.

The results of the trial are ambiguous: the conclusion that aspirin prevents fatal heart attack strongly depends on prior expected efficacy.

Note: this need not always be the case!

Prior sensitivity of the BF (Aspirin study)



Strong evidence against the null only when:

- adverse side effects are expected to be small (< 1%), and
- efficacy is expected to be relatively large (between 30% and 70%).

For other hyperparameters, the evidence is moderate, weak, or favors the null.

The results of the trial are ambiguous: the conclusion that aspirin prevents fatal heart attack strongly depends on prior expected efficacy.

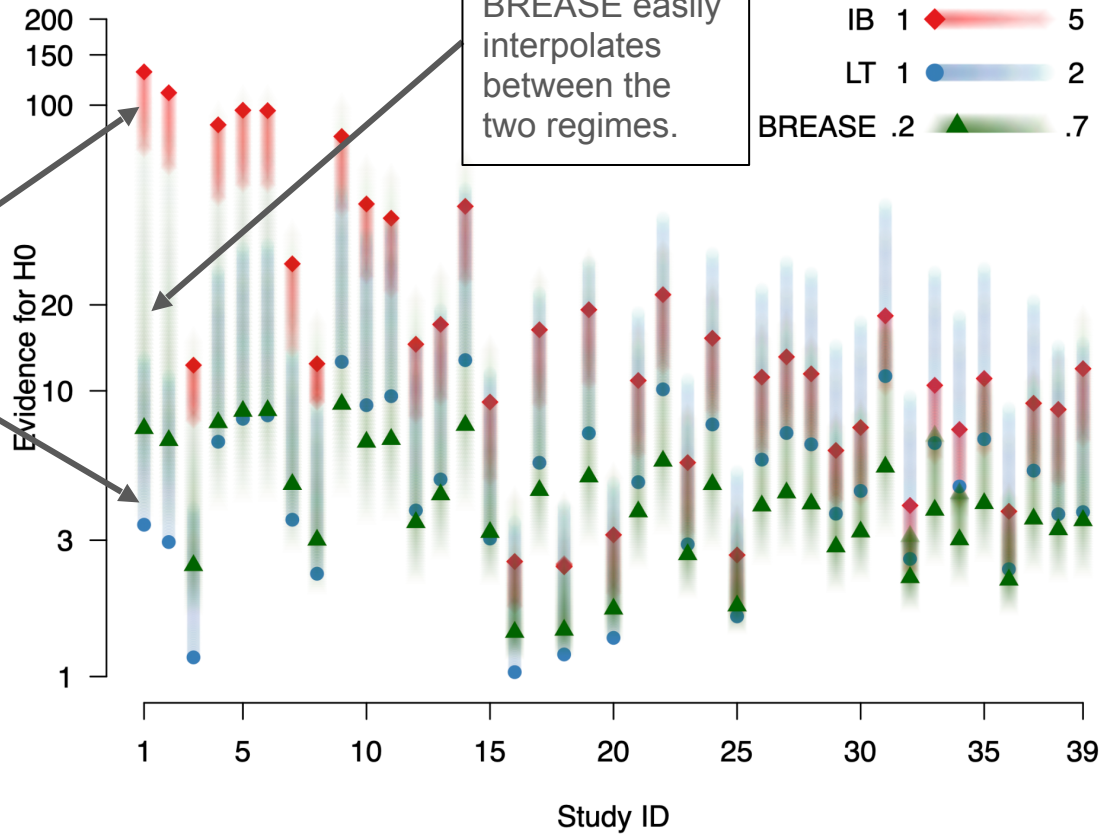
Note: this need not always be the case!

Eg. Reanalysis of the Pfizer COVID-19 trial shows a very robust effect.

BF prior sensitivity across studies in *NEJM*

Large gap in BFs even when reasonably varying IB and LT parameters.

Varying μ, BREASE easily interpolates between the two regimes.



Conclusion

We introduce the BREASE framework for the analysis of binary experiments. Our approach has a number of desirable characteristics when compared to current mainstream alternatives:

- Induces prior dependence between risks in the treatment and control groups in a causally principled way;
- Admits analytical formulae for the Bayes factor and tractable posterior sampling;
- Facilitates elicitation of prior knowledge and sensitivity analysis;
- R code available on Github, and R package coming soon!

arxiv.org/abs/2308.13713