Bayesian meta-analysis with negative control calibration

Contents

Introduction

We want to estimate the effect of a drug or vaccine on some adverse outcome. For example, we want to know if this year's flu vaccine increases the risk of Guillain-Barré Syndrome. We quantify this additional risk via a parameter θ_0 (e.g., a log rate ratio), such that $\theta_0 = 0$ if the vaccine does not have a causal effect on the outcome. We have attempt to estimate the parameter of interest with an estimand we denote $\tilde{\theta}_0$. However, it is not the case that $\tilde{\theta}_0 = \theta_0$; that is, the estimand is biased due to factors such as confounding and measurement error.

Furthermore, we collect (biased) estimates from multiple data sources. Heterogeneity across data sources (due to differences in population, measurement, etc.) means that not only is there variability in the true effect θ_0 within each data source, but also variability in the extent of the bias. Our goal is to account for bias in these data sources while combining evidence across the data.

We use a Bayesian framework to do so and to allow us to answer various questions using posterior probabilities: how likely is it that the vaccine increases risk of the outcome in any of the data sources, overall, for an as-yet-unseen data source, etc.

Setup

Assumptions and notation

Let M denote the number of data sources indexed by $i = 1, ..., M$. Using data y_{i0}, we estimate $\tilde{\theta}_{i0}$, which is biased for the causal parameter θ_{i0} in a given data source.

We will assume that we can decompose this biased estimand into the sum of the true causal effect and the bias, which we will denote β_{i0}, such that $\tilde{\theta}_{i0} = \theta_{i0} + \beta_{i0}$.

We also assume that the bias belongs to some (data-source-specific) distribution, e.g., $\beta_{i0} \sim \text{Normal}(\delta_i, \gamma_i^2)$. We can estimate the parameters of the bias distribution using effect estimates for outcomes with a known true effect size; in this case, negative controls with a truly null causal effect.

That is, within a data source i we estimate for N_i negative control outcomes indexed by j, $j = 1, ..., N_i$, the same measure of effect $\tilde{\theta}_{ij}$, using data y_{ij}. We assume that each is subject to bias from the same distribution as the outcome of interest; as with that outcome, $\tilde{\theta}_{ij} = \theta_{ij} + \beta_{ij}$, where $\beta_{ij} \sim \text{Normal}(\delta_i, \gamma_i^2)$.

However, we know that $\theta_{ij} = 0$ for each of the negative control outcomes: the true effect is null. This means that $\tilde{\theta}_{ij} = \beta_{ij}$; that is, the effect estimate for a negative control will also be an estimate of a draw from the bias distribution. We can therefore use these estimates of $\tilde{\theta}_{ij} = \beta_{ij}$ to estimate the parameters of the bias distribution for a given data source, δ_i and γ_i.

Finally, we assume random data source-specific effects for both the effect and the bias distributions; specifically, we assume that $\theta_{i0} \sim \text{Normal}(\mu, \tau^2)$ and $\delta_i \sim \text{Normal}(\lambda, \eta^2)$. That is, the true effect within each data source is drawn from a distribution with mean effect size μ, and τ^2 describes the heterogeneity in true effects across data sources. Similarly, the mean of the bias distribution is drawn from a shared distribution across datasets. These distributions may of course be specified differently depending on how θ_{i0} is defined and/or any plausible assumptions.

Figure 1 describes the relationships between these parameters and the data graphically.

Bayesian framework

Under this scenario, we can estimate the posterior probability of the parameters given the data on the outcomes of interest as well as the negative controls using Bayes' rule (with bolded parameters indicating the vectors of data source-/negative control-specific parameters):

$$P(\mu, \tau, \lambda, \eta, \theta_0, \beta, \delta, \gamma \mid \mathbf{y}) \propto P(\mathbf{y} \mid \mu, \tau, \lambda, \eta, \theta_0, \beta, \delta, \gamma) P(\mu, \tau, \lambda, \eta, \theta_0, \beta, \delta, \gamma)$$

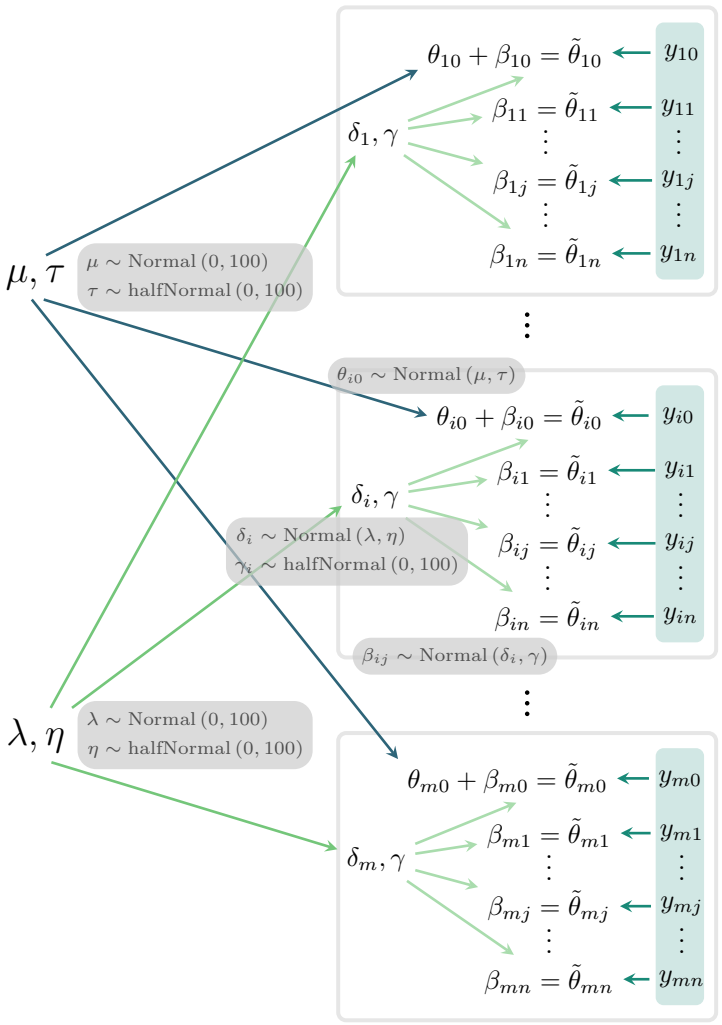

FIGURE 1: The hierarchical model incorporating bias adjustment. Priors are highlighted in grey, and data is highlighted in green.

Given the assumptions about the hierarchical nature of the true effects and the shared bias distributions (and additional assumptions about the independence of e.g., the bias and effect sizes, which could be relaxed), we can write this as:

$$P(\mu, \tau, \lambda, \eta, \theta_0, \beta, \delta, \gamma \mid \mathbf{y}) \propto$$
$$P(\mu, \tau, \lambda, \eta) \prod_{i=1}^{m} P(y_{i0} \mid \theta_{i0}, \beta_{i0}) P(\theta_{i0} \mid \mu, \tau) P(\beta_{i0} \mid \delta_i, \gamma_i) P(\delta_i \mid \lambda, \eta) P(\gamma_i)$$
$$\times \left\{ \prod_{j=1}^{n_i} P(y_{ij} \mid \beta_{ij}) P(\beta_{ij} \mid \delta_i, \gamma_i) \right\}$$

Along with the distributional assumptions about the overall effect size and the bias distribution, we also assign priors to each of μ, τ, λ, η, and γ_i. Figure 1 gives some examples of priors.

In addition, for the $P(y_{i0} \mid \theta_{i0}, \beta_{i0})$ terms, we need the likelihoods for the biased effects of interest ($\tilde{\theta}_{i0}$) given the data y_{i0}. In the absence of directly collecting y_{i0} (e.g., if data security rules preclude combining the full data likelihoods at a central site), we can use estimates of the likelihoods (e.g., the profile likelihoods evaluated across a grid of values). For the negative control data y_{ij}, we might use point estimates and standard errors from a previously estimated model with a normal likelihood.

Implementation

Eumaeus study data

OHDSI's Eumaeus study evaluated a number of methods for vaccine safety research across 5 data sources. The data sources used were 4 administrative claims databases (IBM MarketScan Commercial Claims and Encounters (CCAE), IBM MarketScan Medicare Supplemental Database (MDCR), IBM MarketScan Multi-State Medicaid Database (MDCD), Optum Clinformatics Data Mart (Optum)), and an electronic health records database (Optum Electronic Health Records (OptumEHR)).

I reanalyzed some of the Eumaeus data for seasonal flu vaccine, using the historical comparator and the self-controlled case series designs. A set of 93 negative control outcomes with no plausible causal link to the vaccine were chosen (see protocol). Positive control outcomes were generated using the negative controls. These had known rate ratios of 1.5, 2, and 4.

Data on post-vaccination outcomes was collected from 2017-01-09 to 2018-31-05. For the historical comparator design, outcome rates between 2016-01-09 and 2017-31-05 were computed within each database as a comparison. For the self-controlled case series design, all pre-vaccination person-time, aside from the 30 days immediately preceding vaccination, was used as a comparison. Data was analyzed after each month of the 2017-18 flu season to mimic post-approval surveillance.

Example: animal bite wounds after flu vaccination, historical comparator design

As an example, consider the risk of animal bite wounds after flu vaccination, data for which is presented in Table 1. Across all 5 data sources, the rate of such wounds is elevated after vaccination compared to the historical rate, statistically significantly so in 3 of them. However, we believe these estimates are biased and wish to correct them, as well as estimate the probability that the rate of animal bite wounds is indeed truly elevated.

In the framework described above, we assume that the observed counts, y_{i0}, follow a Poisson distribution:

$$y_{i0} \sim \text{Poisson}\left(\exp\left(\tilde{\theta}_{i0}\right) \times y_i^*\right)$$

where y_i^* is the expected number of cases calcuated using the historical comparison rate (which is considered a fixed value in this design).

TABLE 1: Animal bite wounds after flu vaccination during the 2017-18 flu season, compared to overall animal bite rate during a comparable time period the previous year. Rate ratios are presented with 95% confidence intervals.

Data source	Outcomes (y_i)	Expected outcomes (y_i^*)	Rate ratio ($\exp(\tilde{\theta}_{i0})$)
CCAE	1478	1159.0	1.28 (1.21, 1.34)
MDCD	727	711.2	1.02 (0.95, 1.10)
MDCR	98	91.8	1.07 (0.87, 1.30)
Optum	1280	991.8	1.29 (1.22, 1.36)
OptumEHR	962	501.4	1.92 (1.80, 2.04)

For the bias adjustment we can use data on the other 92 negative controls included in the Eumaeus study. Data for each outcome was not available from each data source, but each had at least 79 additional negative controls (Table 2).

For the main results, I set normal (half-normal for the standard deviation parameters) priors with mean 0 and standard deviation of 5. The `stan` program used is given in the final section.

TABLE 2: Number of negative control outcomes used in the analysis of animal wound bites.

Data source	Number of negative controls (N_i)
CCAE	83
MDCD	80
MDCR	79
Optum	83
OptumEHR	92

Example results

Figure 2 shows the posterior distributions of the true log rate ratio (θ_{i0}) within each data source. We can see that the null value of 0 is well within the 95% highest density interval for each (as well as the 66% highest density interval). Table 3 contains point and interval estimates for comparison with the biased estimates in Table 1.

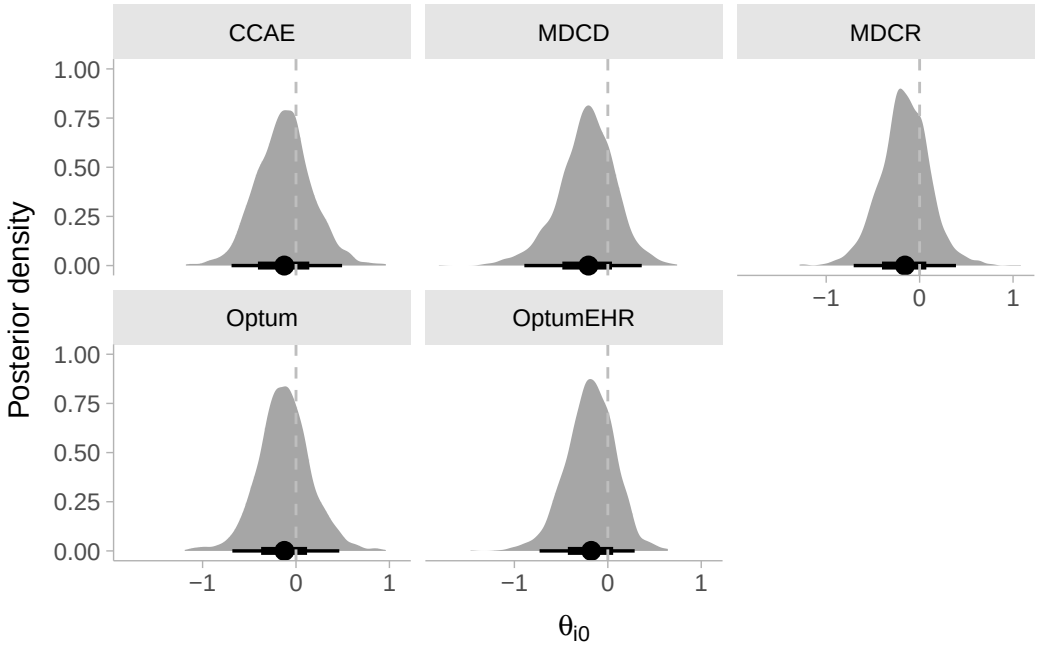

FIGURE 2: Posterior distributions of log rate ratio of flu vaccination on animal wound bites within each data source (θ_{i0}). Medians, 66%, and 95% intervals are indicated with the horizonal dots/bars. The null value of 0 is indicated with a dashed vertical line.

TABLE 3: Medians and 95% credible intervals for the true rate ratio (θ_{i0}) within each data source.

Data source	Median rate ratio (95% credible interval)
CCAE	0.88 (0.50, 1.64)
MDCD	0.81 (0.41, 1.44)
MDCR	0.85 (0.49, 1.48)
Optum	0.88 (0.51, 1.59)
OptumEHR	0.84 (0.48, 1.33)

In addition, we can examine the posterior distributions of the other parameters. For example, Figure 3 shows that the probability that the mean of the bias distribution (δ_i) is greater than 0 is essentially 1 for each data source.

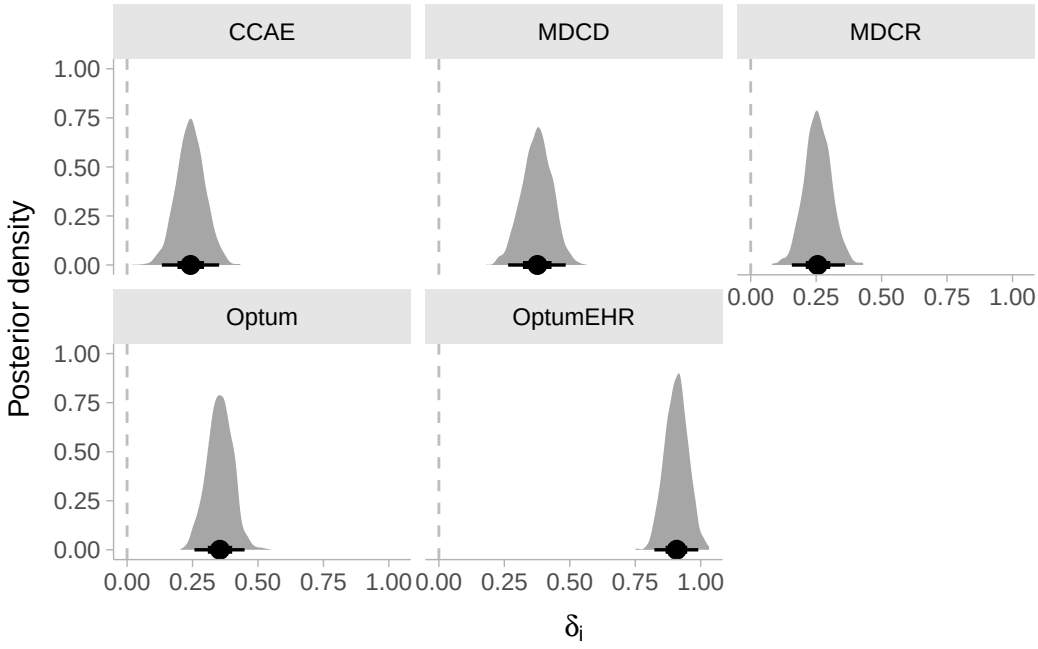

FIGURE 3: Posterior distributions of the mean bias within each data source (δ_i). While each data source has on average positive bias, some sources are apparently much more biased than other.

Figure 4 shows the posterior distributions of the parameters of the overall effect distribution. From these posteriors we can say that there's a probability of 0.25 that the mean of the log rate ratio distribution across data sources is above log(1) and a probability of 0.03 that it is above log(1.5). We can also say that the probability that standard deviation of the effect sizes across data sources is less than 0.41 (the difference between a rate ratio of 1.5 and one of 1) is 0.74. That is, there is little heterogeneity in the magnitude of this effect (which is expected, given that the effect should be null in all data sources).

Comparison of priors

I also explored how the choice of prior distribution – in particular, regarding the certainty about the mean of the overall distribution, μ – affected inference from the posterior. I compared priors distributions for μ of Normal$(0, sd^2)$, using values of 10, 5, 1, and 0.1 for the standard deviation of the prior. In this example only a standard deviation of 0.1 meaningfully affected inference, as can be seen from the point and interval estimates in Table 4. In Figure 5, the posteriors are almost identical for $sd \geq 1$, but are noticeably narrower for $sd = 0.1$. In experiments with other outcomes, I found differences for $sd = 1$, but consistently little difference between $sd = 5$ and $sd = 10$.

I did not compare choice of distribution for the priors, or parameter values for the priors of

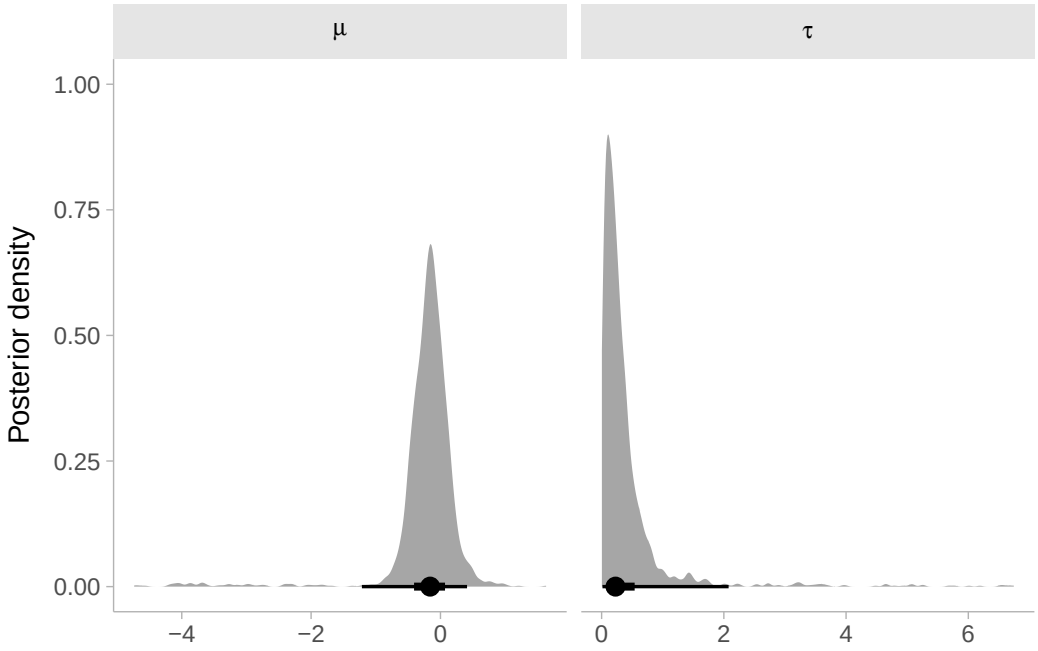

FIGURE 4: Posterior distributions of the mean (μ) and standard deviation (τ) of the overall true log rate ratio distribution.

other parameters. It would also be interesting to explore the prior on τ, which determines the heterogeneity across data sources.

TABLE 4: Point and interval estimates compared across prior standard deviations of the prior distribution of μ.

Data source	sd = 10	sd = 5	sd = 1	sd = 0.1
CCAE	0.86 (0.51, 1.60)	0.86 (0.49, 1.60)	0.86 (0.51, 1.67)	0.97 (0.60, 1.64)
MDCD	0.81 (0.44, 1.44)	0.81 (0.42, 1.41)	0.82 (0.44, 1.39)	0.93 (0.51, 1.41)
MDCR	0.83 (0.52, 1.44)	0.84 (0.49, 1.39)	0.83 (0.51, 1.41)	0.93 (0.58, 1.39)
Optum	0.85 (0.47, 1.52)	0.85 (0.49, 1.48)	0.85 (0.50, 1.44)	0.95 (0.57, 1.44)
OptumEHR	0.82 (0.49, 1.33)	0.81 (0.48, 1.36)	0.82 (0.49, 1.33)	0.92 (0.55, 1.31)

Self-controlled case series

I also fit the model to data from the self-controlled case series design. This design is notably less biased than the historical comparator design. In Table 5 I contrast the crude estimate (i.e., of $\tilde{\theta}_{i0}$) and the empirically calibrated estimate from this design with the median of the posterior. As in the previous section, I compared across values of the standard deviation of the prior for μ. In this example, we see that there wasn't a lot of bias (the crude and empirical calibration estimates are similar), but the data-source-specific estimates are brought closer together in the Bayesian approach due to the random effects.

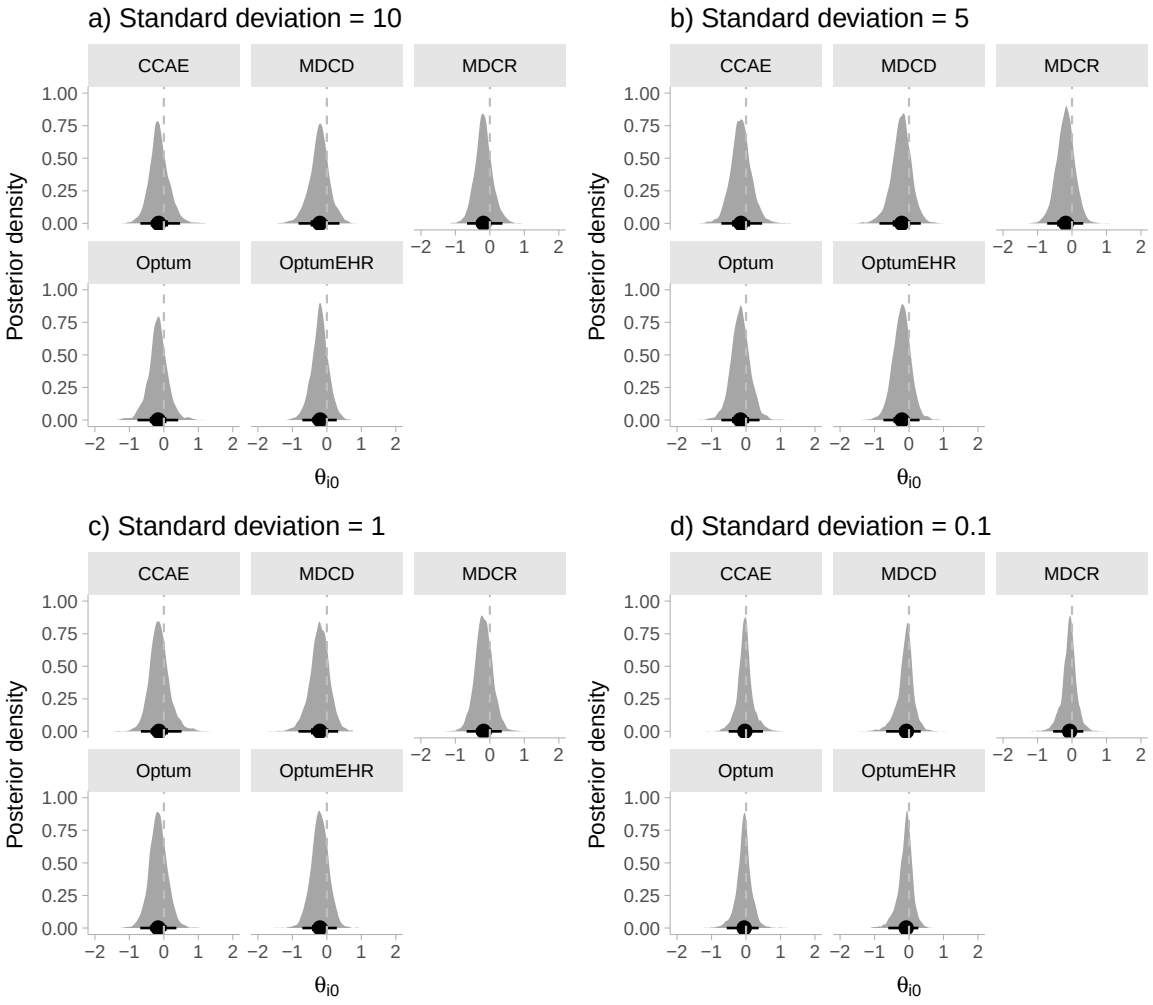

FIGURE 5: Posterior distributions of data source-specific effects when assuming a prior distribution of those effects with a range of standard deviations.

Table 5: Comparison of results between the frequentist empirical calibration approach and the Bayesian approach for the self-controlled case series design across a range of priors.

Database	Crude estimate	Empirical calibration	Prior standard deviation	Bayesian approach
CCAE	0.97 (0.92, 1.03)	0.92 (0.71, 1.19)	10	0.93 (0.71, 1.32)
			5	0.94 (0.72, 1.34)
			1	0.93 (0.71, 1.35)
			0.1	0.97 (0.76, 1.34)
MDCD	0.81 (0.75, 0.87)	0.82 (0.62, 1.08)	10	0.89 (0.66, 1.20)
			5	0.89 (0.65, 1.21)
			1	0.89 (0.66, 1.20)
			0.1	0.93 (0.69, 1.20)
MDCR	0.96 (0.77, 1.18)	0.93 (0.63, 1.35)	10	0.84 (0.64, 1.07)
			5	0.84 (0.63, 1.06)
			1	0.84 (0.63, 1.06)
			0.1	0.89 (0.67, 1.08)
Optum	1.04 (0.98, 1.10)	1.06 (0.78, 1.46)	10	0.91 (0.69, 1.30)
			5	0.92 (0.70, 1.32)
			1	0.91 (0.69, 1.29)
			0.1	0.96 (0.74, 1.30)
OptumEHR	0.86 (0.80, 0.92)	0.76 (0.50, 1.15)	10	0.89 (0.65, 1.23)
			5	0.89 (0.64, 1.25)
			1	0.89 (0.64, 1.19)
			0.1	0.94 (0.68, 1.24)
Overall			10	0.89 (0.68, 1.20)
			5	0.90 (0.69, 1.20)
			1	0.89 (0.69, 1.18)
			0.1	0.95 (0.82, 1.11)

Overall historical comparator performance

To investigate overall performance across many outcomes, I fit models for each of the negative control outcomes and the positive controls of magnitude 1.5, after each (cumulative) month of data (Table 6). I used a leave-one-out structure, in which the negative control being tested, or used to generate the positive control of interest, was not included. Again I used standard deviations of 5 for all the prior distributions.

Table 6: Total number of control outcomes evaluated.

True RR	Months	Outcomes
1.0	1-9	93
1.5	1-9	141

I compared the results to those from a frequentist empirical calibration setup to confirm that they were similar (as expected). Overall, the point estimates (means/medians of the posterior distributions) were comparable to those in the frequentist empirical calibration (Table 7). The 95% credible intervals tended to be narrower than the 95% confidence intervals.

TABLE 7: Comparisons between results from the Bayesian procedure and from the empirical calibration procedure.

	Median RR		Interval coverage		Interval width (log scale)	
Database	Bayesian	Empirical	Bayesian	Empirical	Bayesian	Empirical
True RR = 1						
CCAE	0.95	0.98	96.40%	96.40%	1.27	1.88
MDCD	0.96	0.91	96.30%	97.50%	1.30	2.08
MDCR	0.97	0.93	96.20%	97.50%	1.20	1.73
Optum	0.97	1.00	97.60%	96.40%	1.17	1.65
OptumEHR	0.99	1.01	97.80%	98.90%	1.14	1.49
True RR = 1.5						
CCAE	1.53	1.48	96.90%	96.90%	1.76	1.85
MDCD	1.42	1.38	96.20%	96.20%	1.61	2.04
MDCR	1.42	1.36	96.80%	96.80%	1.30	1.52
Optum	1.53	1.52	97.30%	97.30%	1.53	1.63
OptumEHR	1.58	1.53	98.60%	98.60%	1.42	1.47

Of course, there are many more interesting things we could say about the Bayesian approach using the estimated posterior distributions. For example, we might want to look at the probability over time that an adverse event is more likely among the vaccinated; that is, the posterior probability of a rate ratio > 1.

For negative controls, we would expect that probability not to be too extreme. Figure 6 shows the posterior probability that the mean of the overall effect distribution (μ) is greater than 0 after each of the nine months of the analysis.

Similarly, we can look at the data-source-specific probabilities. Figure 7 shows the posterior probability over time that the causal rate rate is greater than 1 for positive control outcomes whose true rate ratio is 1.5.

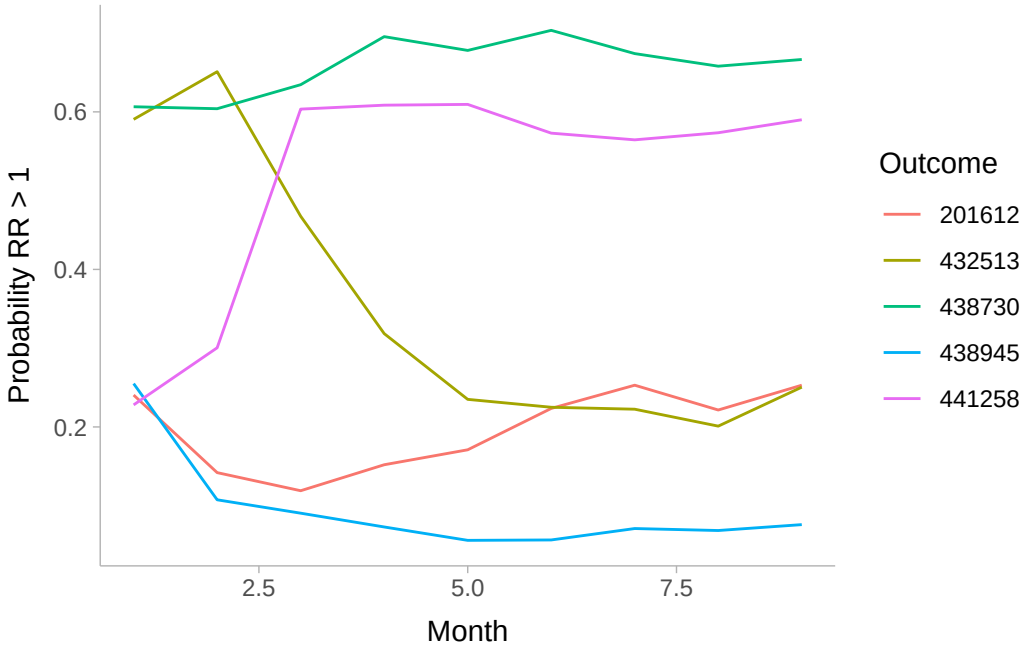

FIGURE 6: The posterior probability that the mean of the overall distribution of log rate ratios is greater than 0, across the analysis period. All outcomes are negative controls, so the true log rare ratio within each data source should be 0.

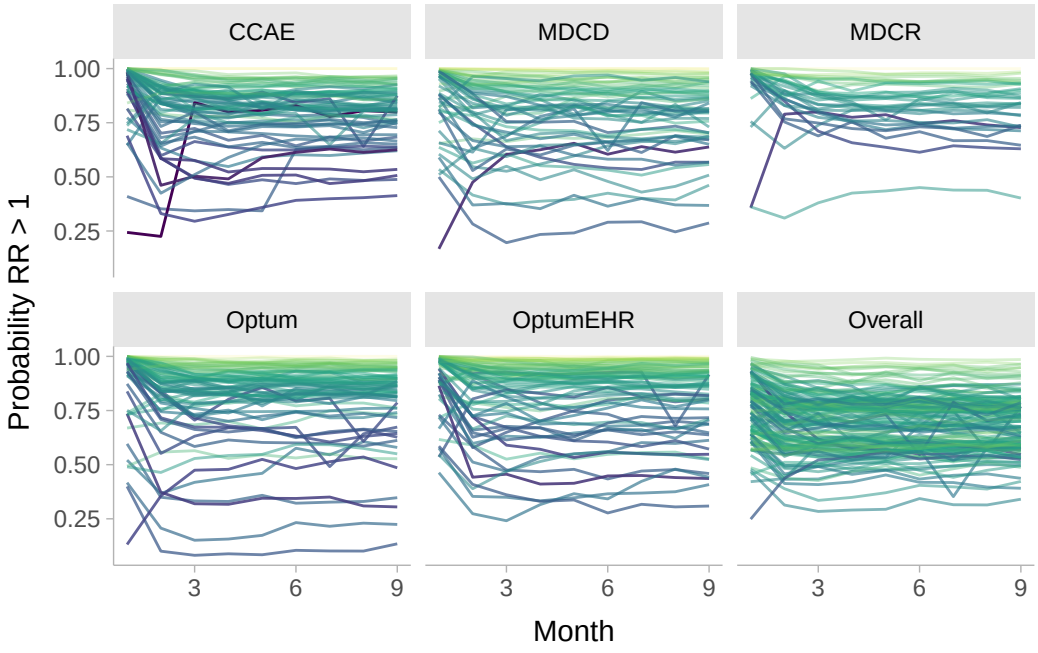

FIGURE 7: The posterior probability that the data-source-specific rate ratio is greater than 1, for all the positive controls (true rate ratio = 1.5). Darker colors indicate more change in the posterior probability over the analysis period. As expected, each data source/outcome combination generally ends with $> 50\%$ of the posterior density above 1. The final panel shows the mean of the overall effect distribution.

Stan code

The following program implements the meta-analysis for the historical controls setting. It is easily adjusted to allow for, e.g., normal (or approximately normal) likelihoods. In the case in which a grid of values at which the likelihood has been evaluated is more appropriate, the code is slightly more complex and takes longer to run.

```
functions {
  // Poisson likelihood for the number of outcomes in terms of the log rate ratio theta
  // and the bias beta given a fixed expected number of outcomes in the comparison group
  real poisson_RR_lpmf(int[] y, real[] theta, real[] beta, real[] y_star) {
    int N = num_elements(theta);
    real theta_tilde[N];
    real lambda[N];
    for (i in 1:N) {
      theta_tilde[i] = theta[i] + beta[i];
      lambda[i] = exp(theta_tilde[i]) * y_star[i];
    }
    return poisson_lpmf(y | lambda);
  }
}

data {
  int<lower=1> M; // number of sites
  int y_0[M]; // y_{0i}
  real<lower=0> y_o_star[M]; // y_{0i}^*

  int<lower=1> N; // number of negative controls (overall, i.e., sum(N_i))
  int<lower=1,upper=N> site[N]; // which site a negative control belongs to
  int y_j[N]; // y_ij
  real<lower=0> y_j_star[N];
  real zeros[N]; //literally just 0s

  // prior distributions
  real mu_mean;
  real<lower=0> mu_sd;
  real tau_mean;
  real<lower=0> tau_sd;
  real lambda_mean;
  real<lower=0> lambda_sd;
  real eta_mean;
  real<lower=0> eta_sd;
```

```
  // can have separate priors for the sd
  // of the site-specific bias if you want
  real gamma_mean[M];
  real<lower=0> gamma_sd[M];
}

parameters {
  // parameters for the overall effect distribution
  real mu;
  real<lower=0> tau;

  // parameters for the overall bias distribution
  real lambda;
  real<lower=0> eta;

  //data source-specific parameters for the bias distribution
  real delta[M];
  real<lower=0> gamma[M];

  // true data source-specific effect
  real theta_0[M];

  // true biases for the effects of interest and for negative controls
  real beta_0[M];
  real beta_j[N];
}

model {
  // priors for overall effect
  mu ~ normal(mu_mean, mu_sd);
  tau ~ normal(tau_mean, tau_sd);

  // priors for bias distribution
  lambda ~ normal(lambda_mean, lambda_sd);
  eta ~ normal(eta_mean, eta_sd);

  // data-source specific bias distribution
  delta ~ normal(lambda, eta);
  gamma ~ normal(gamma_mean, gamma_sd);

  // negative controls
```

```
  for (j in 1:N){
    beta_j[j] ~ normal(delta[site[j]], gamma[site[j]]);
  }
  y_j ~ poisson_RR(zeros, beta_j, y_j_star);

  // effect of interest
  theta_0 ~ normal(mu, tau);
  beta_0 ~ normal(delta, gamma);
  y_0 ~ poisson_RR(theta_0, beta_0, y_0_star);
}
```