Bayesian Evidence Synthesis with Bias Correction

Louisa H. Smith[1,2], Fan Bu[3], Akihiko Nishimura[5], Kristin Kostka[2], Jody-Ann McLeggon[6], Patrick B. Ryan[4], George Hripcsak[6], David Madigan[7], and Marc A. Suchard[3]

[1]Department of Health Sciences, Northeastern University
[2]The OHDSI Center at the Roux Institute, Northeastern University
[3]Department of Biostatistics, University of California, Los Angeles
[4]Janssen Research and Development
[5]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University
[6]Department of Biomedical Informatics, Columbia University
[7]Northeastern University

Background

Evidence synthesis, or meta analysis, is a topic of great interests in health science research. In settings where studies are performed on a federated network of observational health databases, we need to combine and summarize study results without directly extracting patient-level information. Importantly, it is necessary to correct for estimation bias due to residual systematic error in observational data [1, 2, 3]. Conventional meta analysis approaches based on simple mixed effects models prove to be insufficient in these aspects [4]. We introduce a novel, likelihood-based approach for evidence synthesis by jointly learning the meta-analytic effect of interest and correcting for estimation bias through a Bayesian hierarchical modeling framework that admits heterogeneity across data sources. This framework achieves bias correction by analyzing a large set of negative control outcomes, following previous methodological work at OHDSI [3, 5, 6].

Methods

Statistical methodology

We illustrate our statistical methods in the setting of estimating the effect of an exposure on an outcome of interest, though our framework is applicable to any generic meta analysis with likelihood functions available.

Here the true effect size is quantified by the log rate (or risk) ratio, denoted by θ_{i0} where $i = 1, 2, \ldots, M$ indexes each data source. We assume that, due to the presence of residual systematic error, the true effect size θ_{i0} is biased by an additive bias term β_{i0}, leading to a biased estimand $\tilde{\theta}_{i0} = \theta_{i0} + \beta_{i0}$.

We empirically characterize the bias by analyzing a set of N_i negative control outcomes within each data source i, indexed by $j = 1, 2, \ldots, N_i$. For each negative control outcome j, we assume the same additive bias relationship where the biased effect estimand $\tilde{\theta}_{ij} = \theta_{ij} + \beta_{ij}$ with θ_{ij} and β_{ij} denoting the true effect size and estimation bias, respectively. By definition, a negative control outcome is an outcome that has no association with the exposure, and therefore $\theta_{ij} = 0$, leading to $\tilde{\theta}_{ij} = \beta_{ij}$. That is, the uncorrected effect size estimate for the exposure on negative control j within data source i is an estimate of the bias term β_{ij}. Importantly, we further assume that the bias terms β_{i0} and β_{ij}'s are exchangeable within each data source.

We adopt the following generative and prior distributions to build a Bayesian hierarchical model for jointly learning θ_{i0}'s and β_{i0}'s across data sources:

$$\begin{aligned}
\theta_{i0} &\sim \text{Normal}(\mu, \tau^2), \\
\beta_{ij} &\sim \text{Normal}(\delta_i, \gamma_i^2), \\
\delta_i &\sim \text{Normal}(\lambda, \eta^2).
\end{aligned}$$

And further,

$$\begin{aligned}\lambda &\sim \text{Normal}(0, sd^2),\\ \tau, \gamma_i, \eta &\sim \text{halfNormal}(0, 100),\\ \lambda &\sim \text{Normal}(0, 100).\end{aligned}$$

Here, μ denotes the bias-corrected, meta-analytic effect, whereas its prior standard deviation sd is a tunable hyper-parameter. δ_i denotes the average estimation bias within data source i.

Instead of using only the point estimates for $\tilde{\theta}_{ij}$'s, we exploit the likelihood function of the statistical analysis for each exposure-outcome-database triplet. This enables us to employ a fully Bayesian approach and perform statistical inference via Markov chain Monte Carlo (MCMC) sampling. We implement the MCMC via the computational engine `stan` and provide an open-source `R` package at `https://github.com/roux-ohdsi/BBAMA`.

Empirical validation

We empirically validate our methods using results from the EUMAEUS study [7] that evaluated the performance of various epidemiological designs for detecting vaccine adverse events using four insurance 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 one electronic health records database ((Optum Electronic Health Records (**OptumEHR**))). Profile likelihood functions were retained and made publicly available for the historical comparator design and self-controlled case series design, which allow us to implement our Bayesian framework without re-executing the study on the data sources.

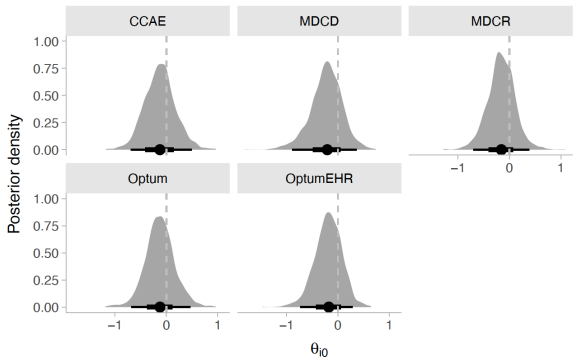
We analyze effects of the seasonal flu vaccine during the 2017-2018 flu season on 93 negative control outcomes and 279 positive control outcomes synthesized from the negative controls with different known effect sizes [1].

Results

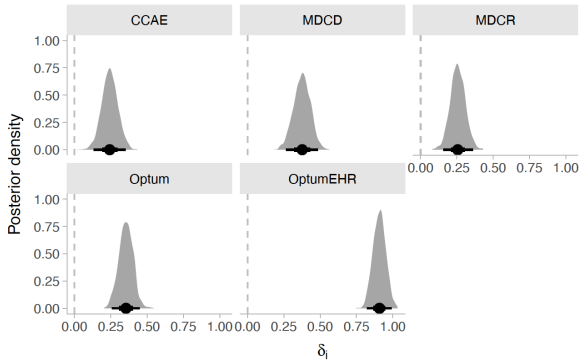
We first showcase the flexibility of the proposed Bayesian evidence synthesis approach by examining the occurrence of animal bite wounds after receiving seasonal flu vaccination using the historical comparator design where we set the prior standard deviation $sd = 5$. Figure 1 presents the posterior distributions of the bias-corrected effect sizes (left) and average empirical bias (right) across the five databases. Our framework is able to identify heterogeneity across databases on bias δ_i (note the different density curves in plot (b)) while recognizing the homogeneity in effect sizes across databases after performing bias correction (note the similarity between posterior density curves for θ_{i0} in plot (a)). This is a much desired feature thanks to the shrinkage property of Bayesian hierarchical models — our framework allows differing effects and bias across different data sources, while encompassing subset models that assume shared effect sizes or bias.

We then validate and compare results obtained by the proposed approach against frequentist estimates with empirical calibration [3]. Table 1 summarizes the effect size estimates and 95% credible (confidence) intervals generated by the two approaches across 93 negative control outcomes with true rate ratio RR = 1 and 93 positive control outcomes with true rate ratio RR = 1.5. Here we apply a leave-one-out procedure, where the control outcome to be estimated for is left out from the bias correction component of the model. We present the median RR estimates across control outcomes, coverage rates and widths of the credible (confidence) intervals. Across the five databases the Bayesian and frequentist effect estimates are similarly accurate, with comparable coverage rates. However, the Bayesian credible intervals are noticeably narrower compared to the confidence intervals generated by empirical calibration. This indicates that our framework can provide more information precision while maintaining the quality of uncertainty quantification.

[1] Information on the control outcomes is provided in the EUMAEUS study protocol at: `https://ohdsi-studies.github.io/Eumaeus/Protocol.html`.



(a) Posterior distributions of bias corrected effect sizes (θ_{i0}) across five databases.



(b) Posterior distributions of average estimation bias (δ_i) across five databases.

Figure 1: Effects of seasonal flu vaccine on animal bite wounds across five databases under the historical comparator design. **Left**: learned effect sizes with bias correction. **Right**: learned estimation bias. Posterior medians and 95% credible intervals are annotated on each posterior density curve.

	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

Table 1: Effect size (rate ratio, RR) estimates obtained using the proposed Bayesian approach and frequentist approach with empirical calibration across 93 negative control outcomes (true RR = 1) and positive control outcomes with true RR = 1.5, with coverage rates and widths of 95% credible/confidence intervals.

Conclusion

We introduce a novel Bayesian evidence synthesis approach to perform meta analysis of observational studies executed on distributed data sources. Through a Bayesian hierarchical modeling framework that admits data source heterogeneity, we can jointly learn the meta-analytic effect size while performing bias correction by analyzing a large set of negative control outcomes. Using results from the EUMAEUS study on vaccine safety, we demonstrate the flexibility of the Bayesian hierarchical framework, and that our approach can generate accurate effect estimates comparable to frequentist estimates with empirical calibration while providing more precise uncertainty quantification.

References

[1] Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)*, 21(3):383, 2010.

[2] Benjamin F Arnold, Ayse Ercumen, Jade Benjamin-Chung, and John M Colford Jr. Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies. *Epidemiology (Cambridge, Mass.)*, 27(5):637, 2016.

[3] Martijn J Schuemie, George Hripcsak, Patrick B Ryan, David Madigan, and Marc A Suchard. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences*, 115(11):2571–2577, 2018.

[4] Martijn J Schuemie, Yong Chen, David Madigan, and Marc A Suchard. Combining cox regressions across a heterogeneous distributed research network facing small and zero counts. *Statistical methods in medical research*, 31(3):438–450, 2022.

[5] Jami J Mulgrave, David Madigan, and George Hripcsak. Bayesian posterior interval calibration to improve the interpretability of observational studies. *arXiv preprint arXiv:2003.06002*, 2020.

[6] Fan Bu, Martijn J Schuemie, Akihiko Nishimura, Louisa H Smith, Kristin Kostka, Thomas Falconer, Jody-Ann McLeggon, Patrick B Ryan, George Hripcsak, and Marc A Suchard. Bayesian safety surveillance with adaptive bias correction. *arXiv preprint arXiv:2305.12034*, 2023.

[7] Martijn J Schuemie, Faaizah Arshad, Nicole Pratt, Fredrik Nyberg, Thamir M Alshammari, George Hripcsak, Patrick Ryan, Daniel Prieto-Alhambra, Lana YH Lai, Xintong Li, et al. Vaccine safety surveillance using routinely collected healthcare data—an empirical evaluation of epidemiological designs. *Frontiers in Pharmacology*, page 2532, 2022.