Supporting Information for "Bayesian data fusion: probabilistic sensitivity analysis for unmeasured confounding using informative priors based on secondary data"

by Leah Comment, Corwin Zigler, Brent A. Coull, and Linda Valeri

A Assumptions for the Bayesian g-formula

Assumption 1 (Positivity). *Let Q be any node in the causal graph G, and q any value in the support of Q. Then for any regime $g_0 \in \{g, g'\}$, $p(q|\text{pa}_{g_0}(Q) = \tilde{pa}) > 0$, where $\tilde{pa}$ is any value taken by $\text{pa}_{g_0}(Q)$, the parent nodes of Q under g_0. Furthermore, it must hold for all q and $\tilde{pa}$ that $p(q|\text{pa}(Q) = \tilde{pa}) > 0$, where $\text{pa}(Q)$ without a subscript indicates the parent nodes of Q in the naturally occurring treatment assignment mechanism.*

Assumption 2 (Consistency). *For any regime $g_0 \in \{g, g'\}$, $Y = Y^{g_0}$ whenever V takes on the values prescribed by g_0. If V is a single binary treatment, this statement simplifies to $Y = VY^1 + (1 - V)Y^0$.*

Assumption 3 (Conditional exchangeability). *For any variable V_0 in the intervention set and every regime $g_0 \in \{g, g'\}$, there exists a set of measured variables $C \subset \{Z, W\}$ such that $Y^{g_0} \perp\!\!\!\perp V_0|C$.*

Assumption 4 (Correct parametric model specification). *For every node $Q \in \{V, W, Y\}$ modeled conditional on variables C with parameters θ_Q, the parametric model $f(Q|C, \theta_Q)$ is correctly specified.*

B Bayesian g-formula algorithms for other causal estimands

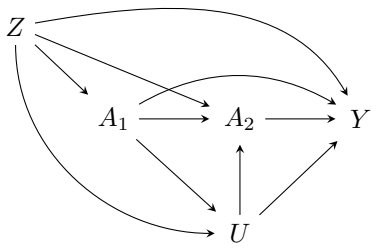
For simplicity of exposition, we assume a single unmeasured confounder U throughout the following algorithms, but U can be vector-valued, or there may be L distinct unmeasured confounders $U_1, U_2, \ldots, U_L$ occupying different positions in the causal graph. Although the algorithms outlined below describe BDF with "forward simulation" (i.e., posterior prediction), Web Appendix D describes how to leverage more computationally efficient closed-form versions (BDF-CF) that are available in certain circumstances (e.g., discrete data, certain outcome models with identity link functions). Throughout, we continue using n_1 to denote the sample size of the main data sources.

B.1 Time-varying confounding of a longitudinal exposure

Suppose that the true causal DAG is as in Supplemental Figure 1, with a discrete exposure-induced unmeasured variable U acting as a confounder of the time-varying exposure A measured at two time points to yield $A = (A_1, A_2)$. Denote the regimes of interest with $g = (A_1 = a_1, A_2 = a_2)$ and $g' = (A_1 = a_1', A_2 = a_2')$, with the causal estimand of interest being the superpopulation average causal effect $\tau_{sp} = \mathbb{E}\left[Y^g - Y^{g'}\right]$.

Adopt parametric generalized linear models indexed by θ_Q for $Q \in \{U, A_2, Y\}$, with $h_Q(\cdot)$ denoting the link function and η_Q the linear predictor term, which is a function only of the parent nodes $\text{pa}(Q)$ and θ_Q. Equation 1 gives a general model form.

$$h_Q(Q_i|\text{pa}(Q_i), \theta_Q) = \eta_{Qi} \tag{1}$$



Supplemental Figure 1: Time-varying causal structure with outcome Y, exposures A_1 and A_2, baseline confounder(s) Z, and time-varying confounder U

The intervention set is $V = (A_1, A_2)$. For $g_0 \in \{g, g'\}$ let $\text{pa}^{g_0}(Q)$ be the intervened parents of Q where any $V \in \{\text{pa}(Q)\}$ has been set (deterministically or stochastically) in accordance with g_0 and the relevant model.

To emphasize that η_{Qi} depends on the parameters θ_Q as well as the values of the parents of Q for observation i, we can also write it as $\eta_{Qi}(\text{pa}(Q), \theta_Q)$.

With models $p(\theta_Q|q, \text{pa}(Q))$ and $\theta = (\theta_U, \theta_{A_2}, \theta_Y)$, the likelihood is

$$\mathcal{L}(\theta) = \prod_{i=1}^{n_1} p(y_i|a_{2i}, u_i, a_{1i}, z_i, \theta_Y) p(a_{2i}|u_i, a_{1i}, z_i, \theta_{A_2}) p(u_i|a_{1i}, z_i, \theta_U)$$

For discrete U, this yields the marginal likelihood of

$$\mathcal{L}^\dagger(\theta) = \prod_{i=1}^{n_1} \left[\sum_u p(y_i|a_{2i}, u_i = u, a_{1i}, z_i, \theta_Y) p(a_{2i}|u_i = u, a_{1i}, z_i, \theta_{A_2}) p(u_i = u|a_{1i}, z_i, \theta_U) \right]$$

1. Fit maximum likelihood models in the external data to obtain the prior $\pi(\theta)$ as detailed in Section **??**.

2. Use NUTS with target probability distribution proportional to $\mathcal{L}^\dagger(\theta) \times \pi(\theta)$ in order to obtain posterior samples of the regression parameter vector θ. For some large B, let $\theta^{(1)}, \ldots, \theta^{(B)}$ denote the B posterior samples remaining after discarding warmup iterations.

3. For MCMC iteration $b = 1, \ldots, B$, sample a length-n_1 weight vector $(d_1^{(b)}, \ldots, d_n^{(b)})$ from a Dirichlet$(1, \ldots, 1)$. For $i = 1, \ldots, n_1$:

   a) For $g_0 \in \{g, g'\}$, set $\tilde{a_1}_i^{g_0(b)}$ deterministically or stochastically in accordance with g_0. For example, if g is the static, deterministic regime setting A_1 to level a_1, $\tilde{a_1}_i^{g(b)} = a_1$ for all i and b.

   b) For $g_0 \in \{g, g'\}$, sample $\tilde{u}_i^{g_0(b)}$ in accordance with

   $$h_Q^{-1}\left(\eta_Q\left(\text{pa}^{g_0}(\tilde{x}_i^{g_0}), \theta_Q^{(b)}\right)\right)$$

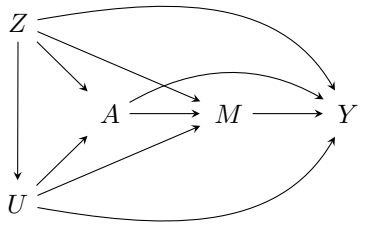
   Concretely, for $g = (A_1 = a_1, A_2 = a_2)$ and a logistic model $\text{logit}\left(P(U_i = 1|Z_i, A_{1i})\right) = \gamma_0 + \gamma_{A_1} A_{1i} + \gamma_Z' Z_i$, sampling $\tilde{u}_i^{g(b)}$ requires drawing from a Bernoulli with success probability

   $$\text{logit}^{-1}\left(\gamma_0^{(b)} + \gamma_{A_1} a_1 + \gamma_Z^{(b)\prime} \tilde{z}_i^{(b)}\right)$$

   c) For $g_0 \in \{g, g'\}$, set $\tilde{a_2}_i^{g_0(b)}$ deterministically or stochastically in accordance with g_0. Concretely, if g is the static, deterministic regime setting A_2 to level a_2, $\tilde{a_2}_i^{g(b)} = a_2$ for all i and b.

   d) For $g_0 \in \{g, g'\}$, draw $\tilde{y}_i^{g_0(b)}$ in accordance with g_0, $\theta_Y^{(b)}$, $\tilde{z}_i^{(b)}$, $\tilde{u}_i^{g_0(b)}$, $\tilde{a_1}_i^{g_0(b)}$, and $\tilde{a_2}_i^{g_0(b)}$. Then calculate the individual-level causal effect

   $$\tilde{\phi}_i^{(b)} = \tilde{y}_i^{g(b)} - \tilde{y}_i^{g'(b)}$$



Supplemental Figure 2: Mediation causal structure with outcome Y, exposures A, mediator M, baseline confounder(s) Z, and unmeasured confounder U

Alternatively, if the conditional mean of Y has a closed form $\mu(\theta_Y, z, u, a_1, a_2)$, define the individual-level causal effect as

$$\tilde{\phi}_i^{(b)} = \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{g(b)}, \tilde{a_1}_i^{g(b)}, \tilde{a_2}_i^{g(b)}\right) - \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{g'(b)}, \tilde{a_1}_i^{g'(b)}, \tilde{a_2}_i^{g'(b)}\right).$$

For example, if Y_i is conditionally normal with mean $\alpha_0 + \alpha_Z' Z_i + \alpha_{A_1} A_{1i} + \alpha_U U_i + \alpha_{A_2} A_{2i} + \alpha_{intx} \times A_{1i} \times A_{2i} \times U_i$ and the contrast of interest compares regimes $g = (A_1 = a_1, A_2 = a_2)$ and $g' = (A_1 = a_1', A_2 = a_2')$,

$$\begin{aligned}\tilde{\phi}_i^{(b)} =& \alpha_{A_1}^{(b)} a_1 + \alpha_U^{(b)} \tilde{u}_i^{g(b)} + \alpha_{A_2}^{(b)} a_2 + \alpha_{intx}^{(b)} a_1 \times a_2 \times \tilde{u}_i^{g(b)} - \\ & \alpha_{A_1}^{(b)} a_1' + \alpha_U^{(b)} \tilde{u}_i^{g'(b)} + \alpha_{A_2}^{(b)} a_2' + \alpha_{intx}^{(b)} a_1' \times a_2' \times \tilde{u}_i^{g'(b)}\end{aligned}$$

4. Calculate population estimate $ACE^{(b)} = \sum_{i=1}^{n_1} (d_i^{(b)} \times \tilde{\phi}_i^{(b)})$.

5. Construct a point estimate for ACE as the posterior mean $\widehat{ACE} = \sum_{b=1}^{B} ACE^{(b)}/B$, and create quantile-based 95% credible intervals as the 2.5^{th} and 97.5^{th} quantiles of $(ACE^{(1)}, \ldots, ACE^{(B)})$.

B.2 Natural direct effects

Suppose the true causal diagram underlying the mediation is shown in Supplemental Figure 2. The unmeasured confounder U can confound the (1) exposure-mediator, (2), exposure-outcome, or (3) mediator-outcome relationships, as well as any combination of (1) - (3).

The population average natural direct effect of changing exposure A to a instead of a^*, while holding the mediator M to its natural value under $A = a^*$, is given by $NDE = \mathbb{E}\left[Y^{aM^{a^*}} - Y^{a^* M^{a^*}}\right]$. This estimand has an intervention set $V = \{A, M\}$ and can be formulated as a contrast in the regimes $g = (A = a, M = M^{a^*})$ and $g' = (A = a^*, M = M^{a^*})$.

1. Fit maximum likelihood models in the external data to obtain the prior $\pi(\theta)$ as detailed in Section 3.3.

2. Use NUTS with target probability distribution proportional to $\mathcal{L}^\dagger(\theta) \times \pi(\theta)$ in order to obtain posterior samples of the regression parameter vector θ. For some large B, let $\theta^{(1)}, \ldots, \theta^{(B)}$ denote the B posterior samples remaining after discarding warmup iterations.

3. For MCMC iteration $b = 1, \ldots, B$, sample a length-n_1 weight vector $(d_1^{(b)}, \ldots, d_n^{(b)})$ from a Dirichlet$(1, \ldots, 1)$. For $i = 1, \ldots, n_1$:

   a) Sample $\tilde{u}_i^{(b)}$ in accordance with $\theta_U^{(b)}$ and z_i using

   $$h_U^{-1}\left(\eta_U\left(z_i, \theta_U^{(b)}\right)\right)$$

   In contrast to previous algorithms, this sampling does not need to be done for each $g_0 \in \{g, g'\}$ because U *cannot* be a descendant of A or M in the causal graph for the natural direct effect to be well defined.

b) Sample mediator $\tilde{m}_i^{a^*(b)}$ according to $\theta_M^{(b)}$, z_i, and $\tilde{u}_i^{(b)}$ using

$$h_M^{-1}\left(\eta_M\left(z_i, \tilde{u}_i^{(b)}, a^*, \theta_M^{(b)}\right)\right)$$

c) For $g_0 \in \{g, g'\}$, draw $\tilde{y}_i^{g_0(b)}$ in accordance with g_0, $\theta_Y^{(b)}$, $\tilde{z}_i^{(b)}$, $\tilde{u}_i^{(b)}$, and $\tilde{m}_i^{a^*(b)}$. For regime $g = (A = a, M = M^{a^*})$ this involves

$$h_Y^{-1}\left(\eta_Y\left(z_i, \tilde{u}_i^{(b)}, \tilde{m}_i^{a^*(b)}, a, \theta_Y^{(b)}\right)\right)$$

while for $g' = (A = a^*, M = M^{a^*})$ the relevant equation will involve

$$h_Y^{-1}\left(\eta_Y\left(z_i, \tilde{u}_i^{(b)}, \tilde{m}_i^{a^*(b)}, a^*, \theta_Y^{(b)}\right)\right)$$

Then calculate the individual-level causal effect

$$\tilde{\phi}_i^{(b)} = \tilde{y}_i^{g(b)} - \tilde{y}_i^{g'(b)}$$

Alternatively, if the conditional mean of Y has a closed form $\mu(\theta_Y, z, u, a, m)$, define the individual-level causal effect as

$$\tilde{\phi}_i^{(b)} = \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{(b)}, a, \tilde{m}_i^{(b)}\right) - \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{(b)}, a^*, \tilde{m}_i^{a^*(b)}\right).$$

4. Calculate population estimate $NDE^{(b)} = \sum_{i=1}^{n_1}(d_i^{(b)} \times \tilde{\phi}_i^{(b)})$.

5. Construct a point estimate for NDE as the posterior mean $\widehat{NDE} = \sum_{b=1}^{B} NDE^{(b)}/B$, and create quantile-based 95% credible intervals as the 2.5^{th} and 97.5^{th} quantiles of $(NDE^{(1)}, \ldots, NDE^{(B)})$.

B.3 Natural indirect effects

Again suppose that the correct causal diagram is as in Supplemental Figure 2, where natural direct effects are well defined.

The population average natural indirect effect is the effect of changing the mediator M from the value it naturally takes under exposure $A = a^*$ to the value it naturally takes under $A = a$, while holding the exposure constant at level a. In potential outcome notation, this quantity is given by $NIE = \mathbb{E}\left[Y^{aM^a} - Y^{aM^{a^*}}\right]$. This estimand has an intervention set $V = \{A, M\}$ and can be formulated as a contrast in the regimes $g = (A = a, M = M^a)$ and $g' = (A = a, M = M^{a^*})$.

The estimation algorithm is the same as in Section B.2 until Step 3, where it continues as follows.

3. For MCMC iteration $b = 1, \ldots, B$, sample a length-n_1 weight vector $(d_1^{(b)}, \ldots, d_n^{(b)})$ from a Dirichlet$(1, \ldots, 1)$. For $i = 1, \ldots, n_1$:

a) Sample $\tilde{u}_i^{(b)}$ in accordance with $\theta_U^{(b)}$ and z_i using

$$h_U^{-1}\left(\eta_U\left(z_i, \theta_U^{(b)}\right)\right)$$

b) For each $a_0 \in \{a, a^*\}$, sample mediator $\tilde{m}_i^{a_0,(b)}$ according to $\theta_M^{(b)}$, z_i, and $\tilde{u}_i^{(b)}$ using

$$h_M^{-1}\left(\eta_M\left(z_i, \tilde{u}_i^{(b)}, a_0, \theta_M^{(b)}\right)\right)$$

c) For $g_0 \in \{g, g'\}$, draw $\tilde{y}_i^{g_0(b)}$ in accordance with g_0, $\theta_Y^{(b)}$, $\tilde{z}_i^{(b)}$, $\tilde{u}_i^{(b)}$, and $\tilde{m}_i^{a_0(b)}$ for the a_0 corresponding to the M counterfactual in g_0. For regime $g = (A = a, M = M^a)$ this involves

$$h_Y^{-1}\left(\eta_Y\left(z_i, \tilde{u}_i^{(b)}, \tilde{m}_i^{a(b)}, a, \theta_Y^{(b)}\right)\right)$$

while for $g' = (A = a, M = M^{a^*})$ the relevant equation will involve

$$h_Y^{-1}\left(\eta_Y\left(z_i, \tilde{u}_i^{(b)}, \tilde{m}_i^{a^*(b)}, a, \theta_Y^{(b)}\right)\right)$$

Then calculate the individual-level causal effect

$$\tilde{\phi}_i^{(b)} = \tilde{y}_i^{g(b)} - \tilde{y}_i^{g'(b)}$$

Alternatively, if the conditional mean of Y has a closed form $\mu(\theta_Y, z, u, a, m)$, define the individual-level causal effect as

$$\tilde{\phi}_i^{(b)} = \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{(b)}, a, \tilde{m}_i^{a(b)}\right) - \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{(b)}, a, \tilde{m}_i^{a^*(b)}\right).$$

4. Calculate population estimate $NIE^{(b)} = \sum_{i=1}^{n_1}(d_i^{(b)} \times \tilde{\phi}_i^{(b)})$.

5. Construct a point estimate for NIE as the posterior mean $\widehat{NIE} = \sum_{b=1}^{B} NIE^{(b)}/B$, and create quantile-based 95% credible intervals as the 2.5^{th} and 97.5^{th} quantiles of $(NIE^{(1)}, \ldots, NIE^{(B)})$.

B.4 Randomized interventional analogs to the natural indirect effect

The estimation algorithm is the same as in Section B.2 until Step 3, where it continues as follows.

3. For MCMC iteration $b = 1, \ldots, B$, sample a length-n_1 weight vector $(d_1^{(b)}, \ldots, d_n^{(b)})$ from a Dirichlet$(1, \ldots, 1)$. For $i = 1, \ldots, n_1$:

   a) For each $g_0 \in \{g, g'\}$ and $a_0 \in \{a, a^*\}$, sample $\tilde{u}_i^{a_0, g_0(b)}$ in accordance with $\theta_U^{(b)}$ and z_i using

   $$h_U^{-1}\left(\eta_U\left(z_i, a_0, \theta_U^{(b)}\right)\right)$$

   b) For $g_0 \in \{g, g'\}$, sample randomized mediator $\tilde{m}_i^{g_0(b)}$ in accordance with $\theta_U^{(b)}$, z_i, and $\tilde{u}_i^{a_0, g_0(b)}$. For regime $g = (A = a, M = H_z(a = a))$, draw $\tilde{m}_i^{g(b)}$ using

   $$h_M^{-1}\left(\eta_M\left(z_i, a, \tilde{u}_i^{a, g(b)}, \theta_M^{(b)}\right)\right)$$

   and for $g' = (A = a, M = H_z(a = a^*))$ draw $\tilde{m}_i^{g'(b)}$ using

   $$h_M^{-1}\left(\eta_M\left(z_i, a^*, \tilde{u}_i^{a^*, g'(b)}, \theta_M^{(b)}\right)\right)$$

   c) Define individual-level causal contrast For $g_0 \in \{g, g'\}$, draw $\tilde{y}_i^{g_0(b)}$ in accordance with g_0, $\theta_Y^{(b)}$, $\tilde{z}_i^{(b)}$, $\tilde{u}_i^{(b)}$, and $\tilde{m}_i^{g_0(b)}$. For regime $g = (A = a, M = H_z(a = a))$, draw $\tilde{y}_i^{g(b)}$ using

   $$h_Y^{-1}\left(\eta_Y\left(z_i, \tilde{u}_i^{a, g(b)}, \tilde{m}_i^{g(b)}, a, \theta_Y^{(b)}\right)\right)$$

   while for $g' = (A = a, M = H_z(a = a^*))$, draw $\tilde{y}_i^{g'(b)}$ using

   $$h_Y^{-1}\left(\eta_Y\left(z_i, \tilde{u}_i^{a, g'(b)}, \tilde{m}_i^{g'(b)}, a, \theta_Y^{(b)}\right)\right)$$

   Then calculate the individual-level causal effect

   $$\tilde{\phi}_i^{(b)} = \tilde{y}_i^{g(b)} - \tilde{y}_i^{g'(b)}$$

   Alternatively, if the conditional mean of Y has a closed form $\mu(\theta_Y, z, u, a, m)$, define the individual-level causal effect as

   $$\tilde{\phi}_i^{(b)} = \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{a, g(b)}, a, \tilde{m}_i^{g(b)}\right) - \mu\left(\theta_Y^{(b)}, \tilde{z}_i^{(b)}, \tilde{u}_i^{a, g'(b)}, a, \tilde{m}_i^{g'(b)}\right).$$

4. Calculate population estimate $rNIE^{(b)} = \sum_{i=1}^{n_1}(d_i^{(b)} \times \tilde{\phi}_i^{(b)})$.

5. Construct a point estimate for $rNIE$ as the posterior mean $\widehat{rNIE} = \sum_{b=1}^{B} rNIE^{(b)}/B$, and create quantile-based 95% credible intervals as the 2.5^{th} and 97.5^{th} quantiles of $(rNIE^{(1)}, \ldots, rNIE^{(B)})$.

B.5 Controlled direct effects

The controlled direct effect is the effect of changing exposure A to level a from a^* while holding the mediator M fixed at level m, i.e., $CDE = \mathbb{E}\left[Y^{am} - Y^{a^*m}\right]$. In general, this requires no unmeasured exposureoutcome confounding and no unmeasured mediatoroutcome confounding. Both of these cases can be addressed by with slight modifications to previously stated versions of the BDF-SIM algorithm in order to obtain B posterior samples $CDE^{(1)}, \ldots, CDE^{(B)}$.

Exposure-outcome confounding or mediator-outcome confounding that not affected by treatment

Suppose the true causal diagram is as in Supplemental Figure 2, where U acts as an exposure-outcome confounder, mediator-outcome confounder, or both. (It may also be an exposure-mediator confounder, but the controlled direct effect is already identified if both the $U \rightarrow A$ and $U \rightarrow Y$ arrows are missing.)

For this causal structure, the controlled direct effect can be estimated using the algorithm from Section B.2, replacing the stochastic assignment of $\tilde{m}_i^{a^*(b)}$ in Step 3b with universal assignment to m for all i and b.

Exposure-induced mediator-outcome confounding

Suppose the true causal diagram is as in Figure 2 in the main text, i.e., where U is an exposure-induced mediator-outcome confounder. Further suppose scientific interest lies in the controlled direct effect of changing A to level a from a^* while holding the mediator M fixed at level m, i.e., $CDE = \mathbb{E}\left[Y^{am} - Y^{a^*m}\right]$.

The controlled direct effect can be estimated using the algorithm from Section B.1, replacing A_1 with A and A_2 with M. The two regimes are $g = (A = a, M = m)$ and $g' = (A = a^*, M = m)$.

C Population vs. sample estimands

As in Section 2 of the main text, consider the problem of estimating the effect of a regime g relative to g', on the difference scale (i.e., $\mathbb{E}[Y^g] - \mathbb{E}[Y^{g\prime}]$).

The counterfactual law of Y under regime g_0 for profile $Z = z_0$ is

$$p(\tilde{y}^{g_0}|z_0, o) \propto \int p(\tilde{y}|g_0, z_0, \theta_Y) d\theta_Y$$

This yields counterfactual posterior mean of:

$$\mathbb{E}[Y^{g_0}|Z = z_0] \propto \int \tilde{y} \times p(\tilde{y}|g_0, z_0, \theta_Y) d\theta_Y$$

The average conditional effect $\tau(z)$ contrasting a regime g to g' evaluated within the $Z = z_0$ group is:

$$\tau(z) = \mathbb{E}[Y^g|Z = z_0] - \mathbb{E}[Y^{g\prime}|Z = z_0] \propto \int \tilde{y} \left[p(\tilde{y}^g|z_0, o) - p(\tilde{y}^{g'}|z_0, o)\right] d\theta_Y$$

In Section 12.4 of their text, Imbens and Rubin (2015) distinguish between estimating what they refer to as a "conditional average treatment effect" (i.e., a finite-sample average effect that conditions on the observed pre-treatment variables in the finite sample) and the average effect in the superpopulation which gave rise to the data set. From a study sample containing n participants, the finite-sample version of this average causal effect is

$$\tau_{cond} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y^g|Z = z_i] - \mathbb{E}[Y^{g\prime}|Z = z_i] \propto \int \tilde{y} \left[p(\tilde{y}^g|z_0, o) - p(\tilde{y}^{g'}|z_0, o)\right] d\theta_Y$$

An estimator of this causal effect is the version implemented in the supplemental code provided by Keil et al. (2015). By contrast, the superpopulation version of the average effect takes into account the additional

uncertainty regarding the distribution of the baseline covariates (i.e., Z). Estimation of this superpopulation quantity is the rationale for modeling Z, either parametrically or using the Bayesian bootstrap.

The posterior for the superpopulation τ_{sp} comparing regime g to g' is given by

$$p(\tau_{sp}|o) \propto \iiint \left[p(\tilde{y}|g_0, \tilde{z}, \theta_Y) - p(\tilde{y}|g', \tilde{z}, \theta_Y)\right] p(\tilde{z}|\theta_Z)\pi(\theta_Z|z)d\theta_Y d\theta_Z d\tilde{z},$$

where $p(\tilde{z}|\theta_Z)\pi(\theta_Z|z)$ can be replaced with $p(\tilde{z}|z)$ via the frequentist or Bayesian bootstrap.

D Potential sources of computational efficiency gains

D.1 Handling non-unique baseline covariate values in the Bayesian bootstrap

Consider our use of the Bayesian bootstrap for modeling $p(z)$, the distribution of baseline covariates. Without loss of generality, assume Z represents n observations of a single covariate. First, let us consider when the Z values are all unique (i.e., there is no $i \neq j$ such that $Z_i = Z_j$ for $i, j \in \{1, \ldots, n\}$). Use of the the Bayesian bootstrap can be seen as a analogous to the posterior distribution after placing an improper Dirichlet$(0, \ldots, 0)$ on the covariate profile probabilities $\theta_Z = (\theta_{Z1}, \ldots, \theta_{Zn}) \in \Delta^n$, where Δ^n is the n-dimensional simplex $\theta_{Zi} > 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^n \theta_{Zi} = 1$.

Now suppose we have duplicated Z values, such that $Z_i = Z_j$ for some $i \neq j$. If there are $K < n$ unique values for Z, then the Bayesian bootstrap on Z amounts to a Dirichlet$(\xi_1, \ldots, \xi_K)$ distribution over a $\theta_Z \in \Delta^k$, where ξ_k is the count of observations with the k^{th} unique covariate profile. We denote this k^{th} unique profile by $z_{(k)}$. Due to the aggregation property of the Dirichlet distribution, this K-dimensional Dirichlet$(\xi_1, \ldots, \xi_K)$ distribution is equivalent to placing on the n-dimensional Dirichlet$(1, \ldots, 1)$ on values for (i, Z_i)– which are unique– and collapsing categories over that first, uniqueness-defining element.

These are equivalent and so we may choose to ignore ties to simplify implementations of the estimator. However the alternative, K-dimensional representation offers an opportunity for computational gains when (1) the estimand of interest has a closed-form expression and/or (2) $K \ll n$. The next subsection describes a BDF algorithm variant that leverages closed-form expressions to speed computations.

D.2 A Bayesian data fusion algorithm for closed-form estimands using the Bayesian bootstrap (BDF-CF)

In certain cases, e.g., when all data are discrete, estimation can be simplified through the use of closed-form expressions for estimators within the MCMC. We now outline a data fusion procedure for the $rNDE$ using the Bayesian g-formula with the lower-dimensional Bayesian bootstrap from Section D.1 for marginalization over baseline covariates. Steps 1, 2, and 5 of the closed-form version are identical to the simulation-based approach in the main text, so we show only steps 3 and 4.

3. Let z_k denote the k^{th} unique covariate profile for $k = 1, \ldots, K$. For $b = 1, \ldots, B$, sample profile weight vector $(d_1^{(b)}, \ldots, d_K^{(b)})$ from a Dirichlet$(\xi_1, \ldots, \xi_K)$ distribution, where ξ_k is the count of observations in the main data set with the k^{th} unique covariate profile. For $k = 1, \ldots, K$, calculate the conditional contrast $\phi_k^{(b)}$ as

$$\begin{aligned}\phi_k^{(b)} =& \text{logit}^{-1}\left(\alpha_0^{(b)} + z_k'\alpha_Z^{(b)} + \alpha_M^{(b)}\tilde{m}_k^{0,g(b)} + \alpha_A^{(b)} + \alpha_{AM}^{(b)}\tilde{m}_k^{0,g(b)} + \alpha_U^{(b)}\tilde{u}_k^{1,g(b)}\right) \\ &- \text{logit}^{-1}\left(\alpha_0^{(b)} + z_k'\alpha_Z^{(b)} + \alpha_M^{(b)}\tilde{m}_k^{0,g'(b)} + \alpha_U^{(b)}\tilde{u}_k^{0,g'(b)}\right)\end{aligned}$$

4. Calculate population estimate $rNDE^{(b)} = \sum_{k=1}^K (d_k^{(b)} \times \phi_k^{(b)})$.

The simulation-based and closed-form Bayesian g-formula approaches are identical with respect to regression parameter estimation, but their differences have implications for extensibility to other causal estimands and scalability to large data sets. In particular, the simulation-based approach is more general because does not require a closed-form expression. However, the closed-form version can take advantage of non-unique values of Z, and does not require computationally intensive forward simulation (i.e., posterior prediction) in addition to the computational burden of parameter MCMC sampling.

D.3 QR decomposition for improved sampler performance

When we applied QR matrix decomposition to the design matrices of our data application, we experienced dramatic improvements in NUTS sampler performance, both with respect to the number of completed MCMC iterations per minute and the number of "effective" parameter draws per MCMC iteration. As such, we strongly recommend implementing this procedure for BDF, particularly with large sample sizes.

However, to calculate the marginalization over U in the BDF likelihood, we needed U to remain on its original (i.e., non-decomposed scale). Setting informative priors on QR-scaled parameters is not straightforward, and selectively QR-scaling the non-U. Below, we outline how to place a $\mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ prior on a regression coefficient vector $\boldsymbol{\beta}_{aug} = (\beta_0, \ldots, \beta_p, \lambda)$ with length $(p+1)$, for a design matrix $\mathbf{X}$ having p columns from N subjects. There is also a single unmeasured confounder U, whose corresponding regression coefficient is λ. The length-N vector of U values is denoted here by $\mathbf{u}$. This decomposition process was used for design matrices $\mathbf{X}_M$ and $\mathbf{X}_Y$; for $\mathbf{X}_U$, the QR decomposition process was more standard because there was no column to keep on the initial scale. Let c_1 and c_2 be scalar tuning parameters.

The purpose of applying the QR decomposition is to perform the MCMC parameter sampling on a scale with more favorable posterior geometry. Specifically, we would prefer to sample $\tilde{\boldsymbol{\beta}}_{aug}$, whose elements are less correlated, with a transformation applied afterwards to recover $\boldsymbol{\beta}_{aug}$.

Apply a thin QR decomposition to obtain $\mathbf{Q}$ and $\mathbf{R}$ such that

$$\mathbf{X}_{N\times p} = \mathbf{Q}_{N\times p}\mathbf{R}_{p\times p}$$

Rescale matrices by scalar c_1:

$$\begin{aligned}\mathbf{Q}^*_{N\times p} &= c_1 \cdot \mathbf{Q} \\ \mathbf{R}^*_{p\times p} &= c_1^{-1} \cdot \mathbf{R}\end{aligned}$$

Define augmented design matrix $\mathbf{X}_{aug}$. For scalar c_2,

$$\begin{aligned}\mathbf{X}_{aug} &= \left(\begin{array}{c|c} \mathbf{X} & \mathbf{u} \end{array}\right)_{N\times(p+1)} \\ &= \left(\begin{array}{c|c} \mathbf{Q}^*\mathbf{R}^* & \mathbf{u} \end{array}\right)_{N\times(p+1)}\end{aligned}$$

Then $\mathbf{X}_{aug} = \mathbf{Q}^*_{aug}\mathbf{R}^*_{aug}$ for scaled augmented matrices $\mathbf{Q}^*_{aug}$ and $\mathbf{R}^*_{aug}$:

$$\begin{aligned}\mathbf{Q}^*_{aug} &= \left(\begin{array}{c|c} \mathbf{Q}^* & c_2 \cdot \mathbf{u} \end{array}\right)_{N\times(p+1)} \\ \mathbf{R}^*_{aug} &= \left(\begin{array}{c|c} \mathbf{R} & \mathbf{0} \\ \hline \mathbf{0} & c_2^{-1} \end{array}\right)_{(p+1)\times(p+1)}\end{aligned}$$

In practice, we found that $c_1 = \sqrt{N-1}$ resulted in better performance than $c_1 = N$. We exclusively investigated and were able to obtain reasonable convergence with $c_2 = 1$.

Sampling can now be performed on the scale of $\tilde{\boldsymbol{\beta}}_{aug}$. To recover $\boldsymbol{\beta}_{aug}$ from $\tilde{\boldsymbol{\beta}}_{aug}$, premultiply $\tilde{\boldsymbol{\beta}}_{aug}$ by the inverse of $\mathbf{R}^*_{aug}$:

$$\boldsymbol{\beta}_{aug} = (\mathbf{R}^*_{aug})^{-1}\tilde{\boldsymbol{\beta}}_{aug}$$

This can be seen from the relationship through the linear predictor $\boldsymbol{\eta}$:

$$\mathbf{Q}^*_{aug}\tilde{\boldsymbol{\beta}}_{aug} = \boldsymbol{\eta} = \mathbf{X}_{aug}\boldsymbol{\beta}_{aug} = (\mathbf{Q}^*_{aug}\mathbf{R}^*_{aug})\boldsymbol{\beta}_{aug}$$

The final question is how to translate our informative prior for $\boldsymbol{\beta}_{aug}$ into an informative prior for $\tilde{\boldsymbol{\beta}}_{aug}$. Here, we use the fact that BDF-derived priors are all multivariate normal. Therefore, the following two priors are equivalent:

$$\boldsymbol{\beta}_{aug} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\tilde{\boldsymbol{\beta}}_{aug} \sim \mathcal{MVN}(\mathbf{R}^*_{aug}\boldsymbol{\mu}, \mathbf{R}^*_{aug}\boldsymbol{\Sigma}{\mathbf{R}^*_{aug}}')$$

E Connection to the power prior

Using prior information derived from a secondary data source is not unique to Bayesian data fusion. One related approach to leveraging external data is the power prior (Ibrahim and Chen, 2000), which in some instances can be related to hierarchical models (Chen and Ibrahim, 2006).

For a simple relationship that is identifiable in historical data D_0, a power prior for a generalized linear regression parameters θ is a prior taking the form

$$\pi(\theta, a_0|D_0) \propto \mathcal{L}(\theta|D_0)^{a_0} \times \pi_0(\theta|c_0)$$

where $0 \leq a_0 \leq 1$ is a scalar prior parameter weighting the historical data relative to the data in the study at hand and $\pi_0(\theta|c_0)$ is an initial prior before seeing D_0. Note that a_0 is not the same as the treatment variable A appearing in the main paper. In the language of the main text, D_0 would be the "secondary" or "external" data.

Due to the consistency and asymptotic normality of maximum likelihood estimators for parameters (θ_Q, Σ_Q) in the regression model for Q, we can approximate the likelihood of θ_Q after conditioning on the external data with

$$\mathcal{L}(\theta_Q|D_0) \approx \mathcal{MVN}\left(\hat{\theta}_{Q,MLE}, \hat{\Sigma}_{Q,MLE}\right)$$

For mediation estimands, this reasoning applies to all $Q \in \{U, M, Y\}$. Then the standard BDF (i.e., with no variance inflation) approximates a power prior in the special case of $a_0 = 1$ and $\pi_0(\theta_Q|c_0) \propto 1$.

$$\pi(\theta_Q, a_0|D_0) \propto \underbrace{\mathcal{L}(\theta_Q|D_0)^{a_0}}_{\approx \mathcal{MVN}\left(\hat{\theta}_{Q,MLE}, \hat{\Sigma}_{Q,MLE}\right)} \times \underbrace{\pi_0(\theta_Q|c_0)}_{1} \tag{2}$$

However, BDF offers a distinct advantage over the analogous power prior: no sharing of individual-level data is required. Although power priors can be implemented based on sufficient statistics in select instances, this is not true for GLMs more generally. BDF priors can always be constructed *solely* on the basis of summary statistics $\hat{\theta}_{Q,MLE}$ and $\hat{\Sigma}_{Q,MLE}$, the sharing of which creates no additional risk of reidentification for study participants.

We attempted to directly compare our BDF approach to power priors using the same simulated external data, but we were unable to achieve satisfactory MCMC performance or convergence when transportability was violated. Despite changing the target Metropolis-Hastings acceptance probability to 0.95 and increasing the maximum treedepth, our implementation of the power prior in Stan did not achieve $\hat{R} < 3$. Run times were also extremely slow even when transportability held (> 24 hours for 500 MCMC iterations and $n_1 = 1000$).

F Frequentist bias corrections

F.1 Overview

The first comparator method, referred to as the delta-gamma (DG) correction, is a classical bias correction method (VanderWeele, 2015). A version for controlled direct effects can be used for the $rNDE$ when the two coincide, i.e., if (1) U is not exposure-induced, and (2) there is no exposure-mediator interaction in the outcome model. Note that for estimands on the risk difference scale, (2) does not hold for logistic models of Y even if the Z-M interaction coefficient is zero. This approach also requires that the effect of U should be the same across all levels of Z (i.e., $\mathbb{E}[Y|a, z, m, U = 1] - \mathbb{E}[Y|a, z, m, U = 0]$ does not depend on z), which cannot hold in a logistic model unless the Z coefficient is zero. Confidence intervals were obtained by bootstrapping the main data 200 times.

A second frequentist correction, referred to as the interaction correction (IX), can accommodate exposure-mediator interaction in the outcome model (VanderWeele, 2010). Originally derived as a bias correction for the NDE, it is more generally applicable than the DG correction, but it similarly requires that U not be exposure-induced. Again, we fit maximum likelihood models to the external data source to derive bias-corrected estimates within each covariate pattern, and confidence intervals were obtained via the bootstrap.

F.2 δ-γ correction ("DG")

For each level of z, the bias due to U for the $rNDE$ on the difference scale, assuming A does not cause U is

$$\begin{aligned} B^{CDE}_{dg,add}(m=0|z) =& \delta_{m=0}(z)\gamma_{m=0}(z) \text{ with} \\ \delta_{m=0}(z) =& P(U=1|z,m=0,a=1) - P(U=1|z,m=0,a=0) \\ \gamma_{m=0}(z) =& \mathbb{E}\left[Y=1|z,a,m=0,u=1\right] - \mathbb{E}\left[Y=1|z,a,m=0,u=1\right] \end{aligned} \tag{3}$$

The DG-corrected population estimate of the $rNDE$ is then given by

$$\widehat{rNDE}_{dg} = \sum_z \left(\widehat{rNDE}_{uc}(z) - B^{CDE}_{dg,add}(m=0|z)\right) p(z). \tag{4}$$

F.3 Interaction correction ("IX")

The bias in the additive NDE from by a mediator-outcome confounder U which is *not* exposure-induced is given by:

$$\begin{aligned} B^{NDE}_{ix,add}(z) =& \sum_{m,u} \left(\left[\mathbb{E}\left[Y|a,m,z,u\right] P(u|a,m,z) - \mathbb{E}\left[Y|a^*,m,z,u\right] P(u|a^*,m,z) \right] P(m|a^*,x) \right) \\ & - \sum_{m,u} \left(\left[\mathbb{E}\left[Y|a,m,z,u\right] - \mathbb{E}\left[Y|a^*,m,z,u\right] \right] P(m|a^*,z,u) P(u|z) \right) \end{aligned}$$

In our context, $a=1$ and $a^*=0$. To eliminate the possibility of model misspecification in the estimation of the bias correction term $B^{NDE}_{ix,add}(z)$, saturated parametric logistic regression models were adopted for U and M when possible. (For the data application, sparseness in the covariates made this impossible, and parametric models with many interaction terms were fit to reduce, but not eliminate, misspecification.) The model for Y used to obtain $\mathbb{E}\left[Y|a=1,m,z,u\right]$ was the same logistic model used to obtain the naive $rNDE$, except with U as an additional term.

The corrected estimate of the population $rNDE$ was calculated by

$$\widehat{rNDE}_{ix} = \sum_z \left(\widehat{rNDE}_{uc}(z) - B^{NDE}_{ix,add}(z)\right) p(z). \tag{5}$$

G Data generation procedure for simulations

The "no-interaction" case corresponding to no statistical interaction has $\Delta_{Y,AM} = 0$.

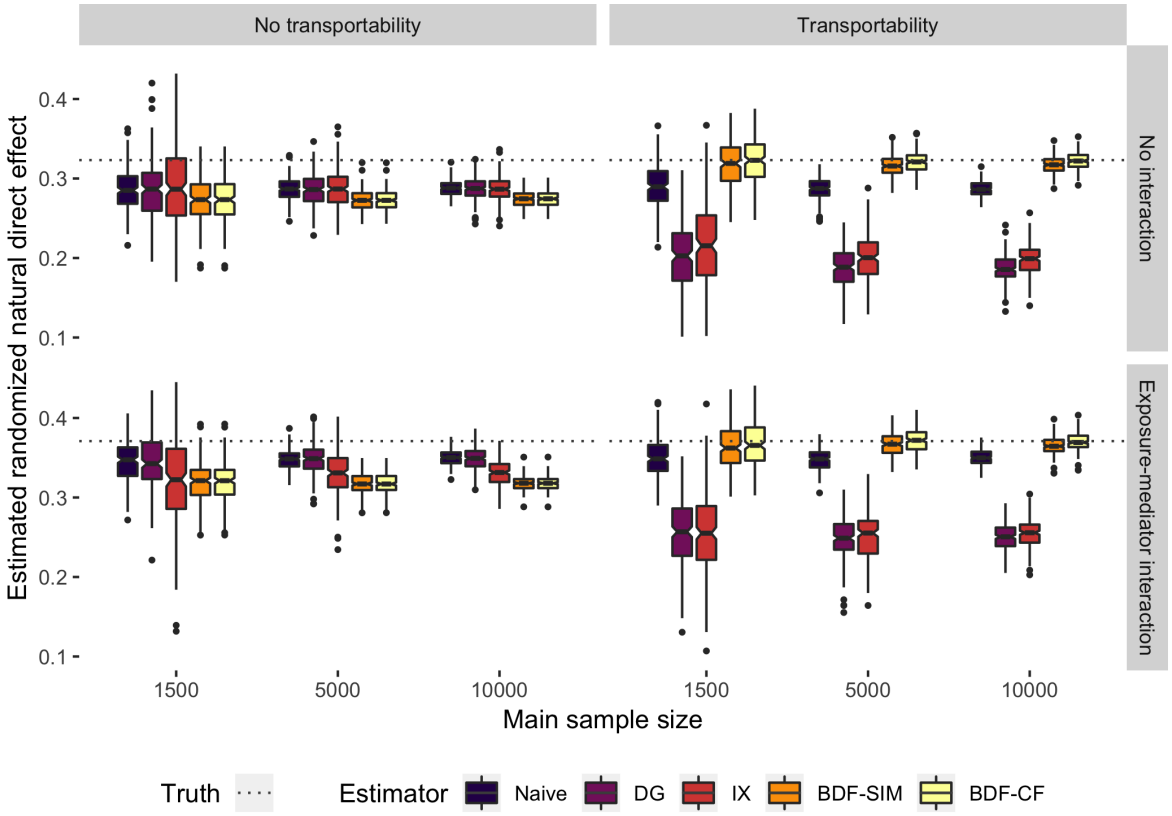
When U is not exposure-induced, $\Delta_{U,A} = 0$.

For violations of transportability, $\beta_U = \alpha_U = 0$ was used for generation of the small data set. Otherwise, $\beta_U = \alpha_U = 1.5$ in order to induce strong mediator-outcome confounding by U.

$$\begin{aligned} Z_1 &\sim \text{Bernoulli}(0.5) \\ Z_2|Z_1 &\sim \text{Bernoulli}(0.5) \\ A|Z_1,Z_2 &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left(-0.2+0.5Z_1+0.7Z_2\right)\right) \\ U|A,Z_1,Z_2 &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left(-0.4+\Delta_{U,A}1.5A\right)\right) \\ M|U,A,Z_1,Z_2 &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left(-1.5+0.3Z_1+0.2Z_2+0.7A+\beta_U U\right)\right) \\ Y|M,U,A,Z_1,Z_2 &\sim \text{Bernoulli}\left(\text{logit}^{-1}\left(-2+0.3Z_1+0.2Z_2+A+0.8M+\Delta_{Y,AM}AM+\alpha_U U\right)\right) \end{aligned}$$

H Additional simulation results

H.1 Boxplots and coverage table for exposure-induced U, including the closed-form estimator variant BDF-CF



Supplemental Figure 3: Randomized natural direct effects estimated with naive, delta-gamma (DG) correction, interaction (IX) correction, simulation-based Bayesian data fusion (BDF-SIM), and closed-form Bayesian data fusion (BDF-CF) estimators, with and without exposure-mediator interaction and causal transportability between main and external data sets. The unmeasured confounder U is exposure-induced.

Coverage probabilities for the closed-form variants in the case of exposure-induced mediator-outcome confounding are included below in Supplemental Table 1.

Supplemental Table 1: Coverage percentages for 95% confidence and credible intervals for naive, delta-gamma (DG) and interaction (IX) frequentist corrections, simulation-based (BDF-SIM) and closed-form (BDF-CF) Bayesian data fusion estimators, calculated in 200 replicates with exposure-induced mediator-outcome confounding

Transportability	Interaction	Sample sizes	Naive	DG	IX	BDF-SIM	BDF-CF
Yes	No	(150, 1500)	73.5	5.0	15.5	93.5	92.5
Yes	No	(500, 5000)	25.0	0.0	0.0	94.0	95.0
Yes	No	(1000, 10000)	2.5	0.0	0.0	91.5	94.5
Yes	Yes	(150, 1500)	69.0	7.0	12.5	92.5	91.5
Yes	Yes	(500, 5000)	19.5	0.0	0.0	95.5	96.0
Yes	Yes	(1000, 10000)	2.5	0.0	0.0	93.5	94.5
No	No	(150, 1500)	66.5	64.0	54.0	50.0	50.0
No	No	(500, 5000)	26.0	31.5	34.0	3.0	2.5
No	No	(1000, 10000)	2.5	10.5	13.0	0.0	0.0
No	Yes	(150, 1500)	59.5	72.5	45.0	53.0	49.5
No	Yes	(500, 5000)	18.0	55.0	33.0	3.5	3.5
No	Yes	(1000, 10000)	3.5	44.0	13.0	0.0	0.0

H.2 Credible interval widths

Mean confidence and credible interval widths for the simulated scenarios are given in Supplemental Table 2. BDF-SIM is the simulation-based estimator described in the main text, whereas BDF-CF uses the closed-form g-formula expressions alluded to in Web Appendix D instead of the posterior prediction of potential outcomes used in BDF-SIM. The Bayesian estimators tend to have widths comparable to frequentist analogs when there is no transportability, but tend to have wider intervals when there is substantial bias in the external data. BDF-CF may perform slightly better than BDF-SIM (i.e., have narrower intervals) in small samples due to smaller Monte Carlo error.

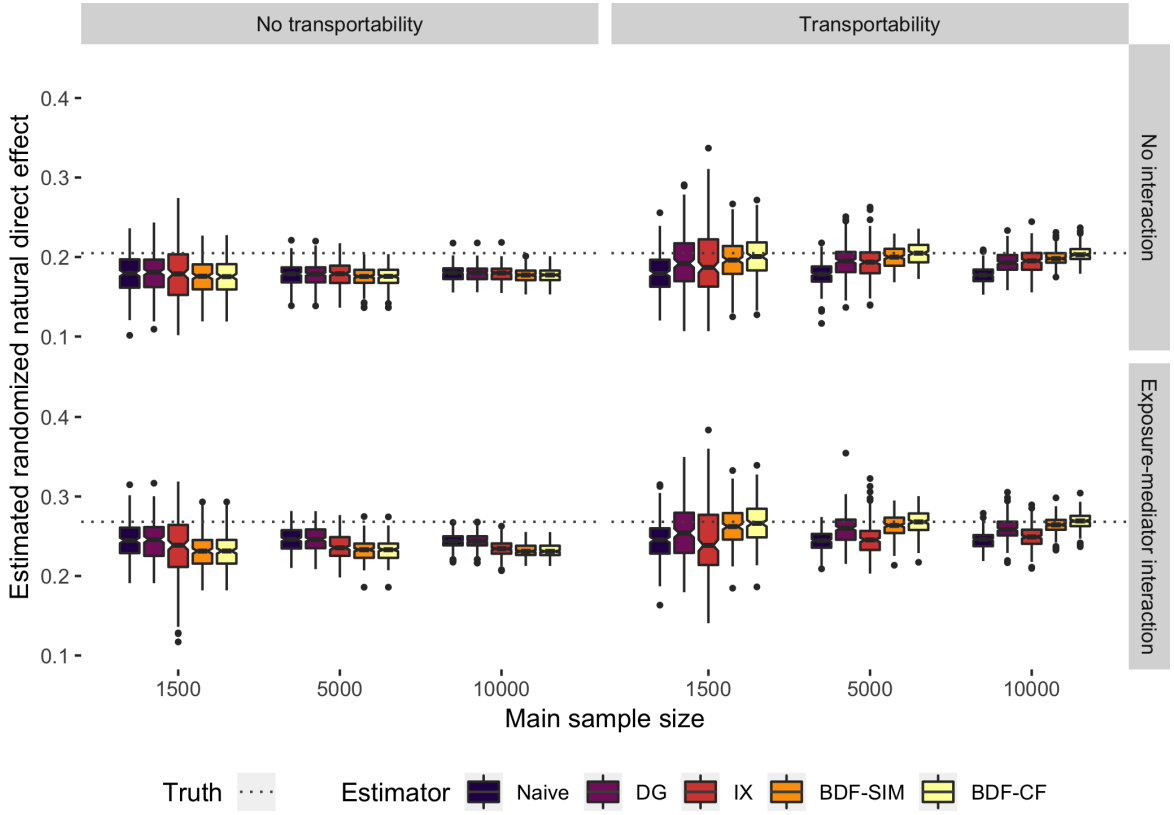
Supplemental Table 2: Widths of 95% confidence and credible intervals for naive, delta-gamma (DG) and interaction (IX) frequentist corrections, simulation-based (BDF-SIM) and closed-form (BDF-CF) Bayesian data fusion estimators, calculated in 200 replicates with exposure-induced mediator-outcome confounding

Transportability	Interaction	Sample sizes	Naive	DG	IX	BDF-SIM	BDF-CF
Yes	No	(150, 1500)	0.101	0.100	0.100	0.108	0.107
Yes	No	(500, 5000)	0.055	0.055	0.055	0.059	0.058
Yes	No	(1000, 10000)	0.039	0.039	0.039	0.042	0.042
Yes	Yes	(150, 1500)	0.101	0.098	0.101	0.112	0.110
Yes	Yes	(500, 5000)	0.056	0.054	0.056	0.061	0.061
Yes	Yes	(1000, 10000)	0.039	0.038	0.039	0.043	0.043
No	No	(150, 1500)	0.100	0.101	0.101	0.100	0.099
No	No	(500, 5000)	0.055	0.055	0.055	0.054	0.054
No	No	(1000, 10000)	0.039	0.039	0.039	0.038	0.038
No	Yes	(150, 1500)	0.101	0.099	0.102	0.103	0.101
No	Yes	(500, 5000)	0.056	0.054	0.055	0.056	0.054
No	Yes	(1000, 10000)	0.039	0.038	0.039	0.039	0.038

H.3 Performance when unmeasured confounder is not exposure-induced

In the absence of transportability, all of the bias corrections considered perform poorly. However, the BDF-SIM and BDF-CF estimates are more confident about their (biased) estimates. In the top right panel of Supplemental Figure 4, one can see that the BDF approaches reduce bias better than the frequentist

corrections in smaller sample sizes, but that the difference in performance is virtually eliminated with a main data set size of $n = 10,000$. Interestingly, the IX correction does *not* exhibit good performance when there is a strong exposure-mediator interaction, particularly in small samples. This is likely due to external data being too small to provide sufficiently precise estimates of the the terms in the bias correction formula.



Supplemental Figure 4: Randomized natural direct effects estimated with naive, delta-gamma (DG) correction, interaction (IX) correction, simulation-based Bayesian data fusion (BDF-SIM), and closed-form Bayesian data fusion (BDF-CF) estimators, with and without exposure-mediator interaction and causal transportability between main and external data sets. The unmeasured confounder U is *not* exposure-induced.

H.4 Transportability violations and the role of the inflation factor

The central assumption of parametric causal transportability (Assumption **??**) can be violated in multiple ways, and the complete transportability across data sets is only guaranteed by design in rare circumstances, such as random subsampling for a validation substudy. The inflation factor proposed in Section **??** is intended to make inference more robust to discrepancies in the data generation processes underlying the main and external data sources. To demonstrate the impact of the inflation factor, we performed additional simulations for the case where external information and internal information are similarly informative, i.e., $n_1 = n_2$.

We considered three data generation scenarios, all with $n_1 = n_2 = 5000$, strong exposure-induced confounding by U, and a large A-M interaction in the Y outcome model. With *full transportability*, the data generation processes are identical. For the *potentially recoverable transportability violation*, the true data generation processes were selectively altered such that the external data had $(\beta_A, \alpha_A, \alpha_{AM}) = (0, 0, 0)$ rather than the $(0.7, 1.0, 1.0)$ in the main data. The primary data set contains some information on all of the discrepant parameters, as A and M are fully observed, but the transportability violation induces a large change in the estimand of interest, the $rNDE$, from 0.371 in the main data to 0.180 in the external data. The

Supplemental Table 3: Coverage percentages for 95% credible intervals for simulation-based (BDF-SIM) Bayesian data fusion estimators, calculated in 200 replicates with exposure-induced mediator-outcome confounding and $n_1 = n_2 = 5000$. Full transportability has identical data generating processes in the external and main data sets, potentially recoverable violations have discrepancies only in variance-inflated parameters, and non-recoverable violations have discrepancies in coefficients for the unmeasured confounder U.

	Inflation factor σ			
Transportability scenario	1	10	50	100
Full transportability	94.0	97.5	92.0	94.0
Potentially recoverable transportability violation	0.0	49.0	90.5	92.0
Non-recoverable transportability violation	0.0	4.5	20.0	23.5

Supplemental Table 4: Widths of 95% credible intervals for simulation-based (BDF-SIM) Bayesian data fusion estimators, calculated in 200 replicates with exposure-induced mediator-outcome confounding and $n_1 = n_2 = 5000$. Full transportability has identical data generating processes in the external and main data sets, potentially recoverable violations have discrepancies only in variance-inflated parameters, and non-recoverable violations have discrepancies in coefficients for the unmeasured confounder U.

	Inflation factor σ			
Transportability scenario	1	10	50	100
Full transportability	0.046	0.058	0.061	0.061
Potentially recoverable transportability violation	0.046	0.058	0.060	0.061
Non-recoverable transportability violation	0.044	0.058	0.059	0.059

non-recoverable transportability violation matches the situation in the main text, where parameters that are completely unobservable in the main data, (β_U, α_U), are $(0, 0)$ in the external data and $(1.5, 1.5)$ in the main data; the true underlying $rNDE$ is 0.052 in the external data and 0.371 in the main data. Within each data generation scenario, we ran 200 simulation replicates and calculated the $rNDE$ using BDF-SIM and prior inflation factors of $\sigma \in \{1, 10, 50, 100\}$.

Supplemental Table 3 shows that both types of transportability violations we considered are sufficiently severe to cause poor interval coverage in the absence of prior variance inflation. Specifically, the 95% credible intervals from the no-inflation scenario (i.e., $\sigma = 1$) have zero coverage. In the potentially recoverable scenario, larger inflation factor values increase coverage. However, even with $\sigma = 100$, coverage is slightly less than nominal at 92%. Intervals get slightly wider (Supplemental Table 4) with larger σ, but are not so wide as to be completely noninformative, with a mean width of ≈ 0.06 on the risk difference scale. In the non-recoverable scenario, coverage remains below 25% for all values of σ we considered. These findings suggest the use of inflation factors to make the analysis more robust to certain violations of transportability. However, as demonstrated in the main text, non-transportability with respect to U-related parameters *cannot* be overcome with prior variance inflation.

References

Chen, M.-H. and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis* **1,** 551–574.

Ibrahim, J. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statistical Science* **15,** 46–60.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Keil, A., Daza, E., Engel, S., Buckley, J., and Edwards, J. (2015). A Bayesian approach to the g-formula. *Statistical Methods in Medical Research* page 0962280217694665.

VanderWeele, T. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology* **21,** 540.

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction.* Oxford University Press.