

SOFTWARE

Open Access



unmconf : an R package for Bayesian regression with unmeasured confounders

Ryan Hebdon^{1*}, James Stamey¹, David Kahle¹ and Xiang Zhang²

Abstract

The inability to correctly account for unmeasured confounding can lead to bias in parameter estimates, invalid uncertainty assessments, and erroneous conclusions. Sensitivity analysis is an approach to investigate the impact of unmeasured confounding in observational studies. However, the adoption of this approach has been slow given the lack of accessible software. An extensive review of available R packages to account for unmeasured confounding list deterministic sensitivity analysis methods, but no R packages were listed for probabilistic sensitivity analysis. The R package `unmconf` implements the first available package for probabilistic sensitivity analysis through a Bayesian unmeasured confounding model. The package allows for normal, binary, Poisson, or gamma responses, accounting for one or two unmeasured confounders from the normal or binomial distribution. The goal of `unmconf` is to implement a user friendly package that performs Bayesian modeling in the presence of unmeasured confounders, with simple commands on the front end while performing more intensive computation on the back end. We investigate the applicability of this package through novel simulation studies. The results indicate that credible intervals will have near nominal coverage probability and smaller bias when modeling the unmeasured confounder(s) for varying levels of internal/external validation data across various combinations of response-unmeasured confounder distributional families.

Keywords Bayesian methods, R package, Unmeasured confounding, Sensitivity analysis, Epidemiology

Introduction

Estimating the causal relationship between an exposure/treatment variable and a desired outcome is often of general interest in observational studies. While randomized clinical trials are recognized as the gold standard in investigating this relationship, observational studies have long been important in healthcare research. Since the subjects are not randomized to treatments, addressing problems due to selection bias and unmeasured confounding are vital for making appropriate inferences [1, 2]. For the issue of unmeasured confounding, researchers

in non-randomized studies often claim the ignorability assumption (i.e., causal framework can only address bias due to measured/observed confounders), where it is assumed any bias from unmeasured confounding is negligible. This dismissal can lead to bias in parameter estimates and erroneous conclusions [3–6].

Sensitivity analysis, or quantitative bias analysis (QBA), provide tools to account for the potential existence of unmeasured confounders in observational research [7–9]. Under QBA, the investigator may choose to use a deterministic or probabilistic approach to quantify the potential direction and impact of the bias. Deterministic QBA specifies a range of values for bias parameters, ϕ , and then calculates the bias-adjusted estimate of the exposure effect assuming ϕ for all combinations of the specified values of ϕ . A bias parameter is defined here to be any parameter that explains the association between the unmeasured confounder and another variable.

*Correspondence:

Ryan Hebdon

Ryan_Hebdon@baylor.edu

¹ Department of Statistical Science, Baylor University, Waco, TX, USA

² CSL Behring, CSL Limited, King of Prussia, PA, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Probabilistic QBA, on the other hand, involves the analyst's explicit modeling of probable occurrences of various combinations of ϕ . The bias parameters are assigned a joint probability distribution (a joint prior distribution in Bayesian terminology) to depict the analyst's uncertainty in the true value of the bias parameters. Monte Carlo sensitivity analysis (MCSA), considered a partial Bayesian analysis and a simple approach to probabilistic QBA, iteratively samples the bias parameters from the joint distribution and fits those sampled values into the same sensitivity analysis formulas used in a fixed-value analysis [10]. Bayesian sensitivity analyses have also been proposed, where the bias parameters are assumed to be unknown and are assigned mildly informative priors in order to avoid non-identifiable models [11, 12]. These methods all rely on additional sources of information on the relationships between the unmeasured confounder and both the response and exposure variables. This information can be from internal validation data, external validation data, or elicited from subject matter experts.

Adoption of sensitivity analysis methods has been slow due to the lack of easily accessible software as well as a focus of methods for binary outcomes and exposures. To aid in the awareness of software developed, a literature search of available software for sensitivity analysis on unmeasured confounders was performed on articles published since 2010 [13]. Twelve packages were implemented in R. These packages, such as `treatSens` [14, 15], `causalsens` [16], `sensemkr` [17], `EValue` [18], and `konfound` [19], implement deterministic QBA. The R package, `episensr`, appears to be the only known package that performs probabilistic QBA via MCSA to account for unmeasured confounders [20]. However, only summary-level data can be supplied to this package. With record-level analysis missing from the package, Fox et al. [21] later adopt MCSA for record-level data and provide modifiable R scripts in a supplementary appendix to be tailored to an analyst's data set. They end their discussion with an encouragement to the user to explore formal Bayesian approaches to overcome some obstacles in their MCSA code.

To address these limitations, we have developed an R package called `unmconf` that uses a Bayesian regression framework to assess the impact of unmeasured confounding in observational studies. To our knowledge, `unmconf` implements the first available package for a fully Bayesian sensitivity analysis and expands beyond the events of binary outcomes and exposures. `unmconf` is available through CRAN at <https://cran.r-project.org/web/packages/unmconf/index.html>. Bayesian sensitivity analysis is often viewed as difficult to implement due to requiring special Markov Chain Monte Carlo (MCMC) software packages and checking convergence of the fitted

model. `unmconf` overcomes these common challenges through a handful of user friendly functions that resemble the `glm()` framework on the front end. The package requires that the user has Just Another Gibbs Sampler (JAGS) installed on their computer, but a user does not need to be proficient in this software.

The introduced package can facilitate sensitivity analyses by leveraging informative priors to explore the influence of unknown unmeasured confounders. Should validation data be accessible, the package enables users to adjust inferences for either one or two unmeasured confounders. In cases where more than two unmeasured confounders are present, the package generates editable JAGS code, offering scalability to accommodate additional confounders. In "Methods" section, we briefly review the statistical model behind `unmconf`. "Example" section compares our model's results to the MCSA found in Fox et al. [21]. With this new software, we provide simulation studies to elaborate on the applicability of the package in "Simulation" section. Lastly, in "Discussion" section, we discuss the strengths and limitations of the underlying model in `unmconf`.

Methods

Motivating example

Consider a study where interest is in the relationship between body mass index (BMI) and hypertension among adults, with age, gender, and cholesterol level as other covariates recorded in the study. For the case where $n = 1000$, BMI, hypertension, age, and gender are fully observed for all subjects but cholesterol is only tracked for 10% of observations. If one wished to perform causal inference using `glm()` in R, the model would estimate the regression coefficients with a message in the summary saying, "900 observations deleted due to missingness". A summary will output the results, but the conclusions drawn from the subset of 100 observations may or may not accurately capture the true, latent relationship between hypertension and BMI. Additional information through the other observations are discarded due to the missing data in cholesterol. Rather than discarding the data, `unmconf` estimates cholesterol as part of the overall Bayesian unmeasured confounding model. Of course, in some cases the information on cholesterol may come from other sources, either data or expert opinion, and the standard `glm()` function cannot address the unmeasured confounder at all.

The choice of methodologies for addressing unmeasured confounders relies on the availability of information regarding the unmeasured confounders and the objectives of assessing such confounding. The level of information can vary, ranging from expert opinion to either internal or external data sources. Internal validation is

available when the unmeasured confounder is ascertained for a, typically, small subset of the main study data. In certain cases, only internal validation data may be accessible and information on the unmeasured confounder is only known for a subset of the patients. Provided that the internal validation data is a well representative sample, a small subsample of the unmeasured confounder variable can effectively calibrate estimation in Bayesian analysis [22, 23]. External validation data is data from previous studies where the unmeasured confounder is fully observed. External data methods make the cautious assumption of transportability across study populations in comparing them to the main study. The external data often captures the relationship between the unmeasured confounder and the response yet lacks information on the exposure. A combination of external data and mildly informative priors can calibrate the estimation through the transportability assumption. `unmconf` handles both of these events for up to two unmeasured confounders. The Bayesian approach we apply here is related to missing data methods where the unmeasured confounder is treated as missing data and is imputed similar to missing at random (MAR) models.

Bayesian unmeasured confounding model

For the statistical model, we denote the continuous or discrete response Y , the binary main exposure variable X , the vector of p other perfectly observed covariates \mathbf{C} , and the unmeasured confounder(s) relating to both Y and X U . In the event of more unmeasured covariates, we denote them U_1, U_2 , and so forth; these unmeasured confounders can be either binary or normally distributed.

In the scenario where there is a single unmeasured confounder, Lin et al. [24] suggest the factorization $f(y, u|x, c) = f(y|x, u, c)f(u|x, c)$, which yields

$$\begin{aligned} Y|x, c, u &\sim D_Y(g_Y^{-1}(\mathbf{v}'\boldsymbol{\beta} + \lambda u), \xi_y) \\ U|x, c &\sim D_U(g_U^{-1}(\mathbf{v}'\boldsymbol{\gamma}), \xi_u), \end{aligned} \tag{1}$$

where $\mathbf{v}' = [x \ \mathbf{c}']'$ denotes the vector of the main exposure variable and all of the perfectly observed covariates. (1) is defined as a Bayesian unmeasured confounding model with one unmeasured confounder, where the distribution for Y pertains to the response model and the distribution for U pertains to the unmeasured confounder model. This model is completed by the specification of a link function g_* and some family of distributions D_* . Additional parameters for certain distributions—if any—are denoted ξ_y and ξ_u . Examples of these would be σ^2 for the variance of a normal distribution or α for the shape parameter in the gamma distribution. For the cases of binomial and Poisson distributions, these parameters are absent. `unmconf` allows the user to work with a

response from the normal, Poisson, gamma, or binomial distribution and unmeasured confounder(s) from the normal or binomial distribution. The package supports the identity (normal), log (Poisson or gamma), and logit (Bernoulli) link functions. Here we build a conditional distribution model on Y given treatment, perfectly observed confounders, and unmeasured confounders in addition to a marginal distribution model on U given treatment and measured confounders. The joint modeling is able to provide an adjusted estimate of the treatment-outcome effect [12, 22, 23, 25]. This goes beyond previous works only allowing for either a binary or continuous response [13].

`unmconf` also extends beyond previous software packages by allowing for a second unmeasured confounder. For the second unmeasured confounder, the joint distribution can be factorized as $f(y, u_1, u_2|x, c) = f(y|x, c, u_1, u_2)f(u_1|x, c, u_2)f(u_2|x, c)$, giving the Bayesian unmeasured confounding model:

$$\begin{aligned} Y|x, c, u_1, u_2 &\sim D_Y(g_Y^{-1}(\mathbf{v}'\boldsymbol{\beta} + \lambda'\mathbf{u}), \xi_y) \\ U_1|x, c, u_2 &\sim D_{U_1}(g_{U_1}^{-1}(\mathbf{v}'\boldsymbol{\gamma} + \zeta u_2), \xi_{u_1}) \\ U_2|x, c &\sim D_{U_2}(g_{U_2}^{-1}(\mathbf{v}'\boldsymbol{\delta}), \xi_{u_2}), \end{aligned} \tag{2}$$

where again $\mathbf{v}' = [x \ \mathbf{c}']'$ so that the coefficients of all perfectly observed covariates in the response model are $\boldsymbol{\beta}$'s, and λ denotes the coefficients on the unmeasured confounders in the response model aggregated together as $\mathbf{u} = [u_1 \ u_2]'$. For the first unmeasured confounder model in (2), the coefficients on all perfectly observed covariates are $\boldsymbol{\gamma}$'s, and ζ denotes the coefficient on the second unmeasured confounder in this model. The remaining parameters in $\boldsymbol{\delta}$ denote the relationship between U_2 and perfectly known covariates. Any parameter that models an association with U_1 or U_2 is defined as a bias parameter and requires either validation data or an informative prior to be estimated. The bottom two equations in (2) do not require conditional dependence between the two unmeasured confounders but rather grants the user flexibility to apply knowledge of one unmeasured confounder to estimate the other. Likewise, a multivariate distribution for $U_1, U_2|x, c$ would align with this framework. If a multivariate distribution is desired on U_1, U_2 , the user would need to edit the JAGS code provided by `unmconf`. This same construction is generalizable to an arbitrary number of unmeasured confounders, the notation simply becomes a bit more cumbersome. We introduce the concept here, however, because the scenario of two unmeasured confounders is not uncommon, shows the general character of the construction, and illustrates how the implementation provided by `unmconf` works. The model in (2) is referenced throughout this paper. Given that we are interested in the exposure effect on Y

Table 1 Default prior structure for regression coefficients and nuisance parameters in `unmconf`

Parameter	Default prior
$\beta_*, \lambda_*, \gamma_*, \zeta_*, \delta_*$	<code>dnorm(0, 0.1)</code> when model is discrete; <code>dnorm(0, 0.001)</code> when model is continuous
τ_*	<code>dgamma(0.001, 0.001)</code> when * is normal
α_y	<code>dgamma(0.1, 0.1)</code> when response is gamma

conditional on X, C, U , β_x is the primary parameter of interest for this study.

$$f(\mathbf{u}_{1\text{mis}}, \mathbf{u}_{2\text{mis}}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{x}, \mathbf{C}, \mathbf{u}_{1\text{obs}}, \mathbf{u}_{2\text{obs}}) \propto \pi(\boldsymbol{\theta}) \prod_{i=1}^n f(y_i | x_i, \mathbf{c}_i, \mathbf{u}_{1i}, \mathbf{u}_{2i}) f(u_{1i} | x_i, \mathbf{c}_i, \mathbf{u}_{2i}) f(u_{2i} | x_i, \mathbf{c}_i).$$

Prior distributions

Prior distributions for the model parameters will be jointly defined as $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \zeta, \boldsymbol{\delta})$ with bias parameters $\boldsymbol{\phi} = (\boldsymbol{\lambda}, \boldsymbol{\gamma}, \zeta, \boldsymbol{\delta}, \xi_{u_1}, \xi_{u_2})$. The bias parameters in an unmeasured confounder model can only be inferred through either validation data or informative priors. The default prior structure in `unmconf` is in Table 1. The regression coefficients have a relatively non-informative prior with a mean of 0 and precision (inverse of the variance) of 0.1 when the response is discrete. When the response is continuous, the regression coefficients have a relatively non-informative prior with a mean of 0 and precision of 0.001. To further customize the analysis, users can specify custom priors using the `priors` argument within the modeling function, `unm_glm()`. The format for specifying custom priors is `c("{parameter} [{covariate}]" = "{distribution}")`. Example code eliciting informative priors is provided in “[Example](#)” section.

Families specified for the response and unmeasured confounder(s) may present nuisance parameters, necessitating the inclusion of their prior distributions as well. The precision parameter, τ_* , on a normal response or normal unmeasured confounder will have a `Gamma(0.001, 0.001)` as the default prior. Priors can also be elicited in terms of σ_* or σ_*^2 through `priors`. The nuisance parameter, α_y , for a gamma response has a gamma distribution as the prior with both scale and rate set to 0.1. The aforementioned nuisance parameters are tracked and posterior summaries are provided as a default setting in the package, but this can be modified.

Posterior inference

When data is observed, scalar-valued objects in (2) become vectors and vectors become matrices, one for each response indexed $i = 1, \dots, n$. Thus, data is constituted of the observed response values \mathbf{y} , exposure values \mathbf{x} , covariates \mathbf{C} . The exposure values and covariates can be combined columnwise to form the matrix \mathbf{V} whose rows correspond to the same observation. Similarly, the unmeasured or partially unmeasured confounders would be denoted \mathbf{u}_1 and \mathbf{u}_2 , perhaps aggregated together as the columns of a matrix \mathbf{U} . Combining the Bayesian unmeasured confounding model with two unmeasured confounders in (2) with $\pi(\boldsymbol{\theta})$, we get the joint posterior distribution,

We sample from this posterior distribution via Gibbs sampling to obtain approximate marginal posterior distributions for the parameters of interest. For modeling purposes, unmeasured confounders can be viewed as missing variables and thus are treated as parameters in the Bayesian paradigm. Bayesian computation in a missing data problem is based on the joint posterior distribution of the parameters and missing data conditioned on the modeling assumptions and observed data [26]. Using (2), we compute the joint posterior of $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\gamma}, \zeta, \boldsymbol{\delta}, \mathbf{u}_{1\text{mis}}, \mathbf{u}_{2\text{mis}})$ given the observed $(\mathbf{y}, \mathbf{x}, \mathbf{c}, \mathbf{u}_{1\text{obs}}, \mathbf{u}_{2\text{obs}})$, where the subscripts “mis” and “obs” indicate the values that were missing and observed, respectively. The posterior simulation then uses two or three Gibbs sampling steps, depending on the number of unmeasured confounders.

Example

We illustrate the applicability of our package by first considering a simple example found in Fox et al. [21]. We consider the data provided in their paper, modeling a binary unmeasured confounder, binary response, and binary exposure. Summary-level data from the paper is displayed in Table 2. No additional measured

Table 2 Observed record-level data

	Exposed	Unexposed
Disease	40	20
No Disease	60	80
Total	100	100
Risk	0.4	0.2
Risk Ratio	2.0	

confounders are accounted for. The relationship between the exposure and response, accounting for the presence of an unmeasured confounder, is of interest.

Fox et al. [21] define $P(C + |E+)$ as the probability of the unmeasured confounder among those with the exposure, $P(C + |E-)$ as the probability of the unmeasured confounder among those without the exposure, and RR_{CE} as the relative risk. They set $P(C + |E+) \sim \text{Beta}(10, 20)$, $P(C + |E-) \sim \text{Beta}(5, 20)$, and $RR_{CE} \sim \text{trapezoidal}(\text{min} = 1.5, \text{mod1} = 1.7, \text{mod2} = 2.3, \text{max} = 2.5)$ to be the distributions on the bias parameters. The algorithm for the probabilistic QBA by Fox et al. [21] is detailed in their paper. They base their Monte Carlo sensitivity analysis $m = 100,000$ times on record-level data and results are summarized by the median, 2.5th percentile, and 97.5th percentile of the distribution in terms of a relative ratio. The run time was 11 minutes. We note that the summary-level code by Fox et al. [21] runs much faster, but we use their record-level code for comparison with `unmconf` given that model is built on a regression framework.

We apply the Bayesian unmeasured confounding model in (1) using `unmconf`. The distributions on Y and U are Bernoulli with no C covariates. This simplifies to

$$Y|x, u \sim \text{Bernoulli}(\text{logit}^{-1}(\beta_1 + \beta_x x + \lambda u))$$

$$U|x \sim \text{Bernoulli}(\text{logit}^{-1}(\gamma_1 + \gamma_x x)).$$

Fox et al. [21] generated the unmeasured confounder through sampling from the bias parameters' distributions. With U completely missing in our model, informative priors are needed on the bias parameters to converge to meaningful results. We use the information from Fox et al. [21] and apply the conditional means prior approach of [27, 28] to determine priors for the bias parameters. Note that the package uses the parameterization of JAGS for the normal, that is, mean and precision, where precision is the reciprocal of the variance. Using the $P(C + |E-) \sim \text{Beta}(5, 20)$ and $P(C + |E+) \sim \text{Beta}(10, 20)$, we induce $\gamma_1 \sim N(-1.5, \tau_{\gamma_1} = 3.7)$ and $\gamma_x \sim N(.747, \tau_{\gamma_x} = 2.31)$ priors for the unmeasured confounder/exposure model. We also require a prior for λ . Using $RR_{CE} \sim \text{trapezoidal}(\text{min} = 1.5, \text{mod1} = 1.7, \text{mod2} = 2.3, \text{max} = 2.5)$, we induce $\lambda \sim N(0.99, \tau_\lambda = 26.70)$.

We fit the model using our package across 4 chains, each with 25,000 iterations of which 4,000 were burn-in for a total of 100,000 posterior draws. We note that the MCMC algorithm does not require 25,000 iterations for the sake of convergence. When comparing time, we set the iterations to be the same for the posterior sampling as Fox et al. [21] did for MCSA. This took about 50 seconds compared to the 11 minutes from their code. The

Table 3 Results in terms of odds ratio for each sensitivity analysis

Model	Data type	2.5th	Median	97.5th	Run time
Fox et al. [21]	Original data	1.075	2.258	5.714	11 min.
<code>unmconf</code>	Original data	1.168	2.350	4.837	50 sec.
<code>unmconf</code>	Original data w/ more diffuse priors	1.157	2.353	4.808	56 sec.
Fox et al. [21]	10x original data	1.559	2.246	3.288	40 min.
<code>unmconf</code>	10x original data	1.623	2.390	3.267	20.5 min.

2.5th percentile, median, and 97.5th percentile from their simulation along with the 95% credible interval and posterior median from `unmconf` are displayed in Table 3 in terms of the odds ratio. Despite using a trapezoidal for the risk ratio for Fox et al. [21] versus the normal distribution for the logistic regression parameter in our model, inferences are very similar. We add the code to convey the user friendliness of this package.

```
df <-
  tibble(y = c(rep(1, 40), rep(0, 60), rep(1, 20), rep(0, 80)),
         x = c(rep(c(1, 0), each = 100))) |>
  mutate(u = as.numeric(NA))

mod <- unm_glm(
  form1 = y ~ x + u,
  form2 = u ~ x,
  family1 = binomial(),
  family2 = binomial(),
  data = df,
  n.iter = 25000, n.adapt = 4000, thin = 1, n.chains = 4,
  priors = c("gamma[1]" = "dnorm(-1.5, 3.7)",
            "gamma[x]" = "dnorm(.747, 2.31)",
            "lambda[u]" = "dnorm(0.9872587, 26.69809)")

unm_summary(mod) |>
  mutate(across(mean: `97.5%`, ~ exp(.x)))
```

To investigate whether priors on the non-bias parameters impacted the width of the intervals, we performed the analysis again with a precision of 0.01 on the regression parameters' priors instead of the default precision of 0.1 for the Bayesian model. Similarly, fitting the model via MCMC across 4 chains with 25,000 iterations of which 4,000 were warm-up, the model again took about 56 seconds. The 2.5th percentile, median, and 97.5th percentile from the Fox et al. [21] simulation are displayed in Table 3, in terms of the odds ratio, along with the 95% credible interval and posterior median from `unmconf`. As expected, the 95% credible interval is slightly wider with the more diffuse priors.

To see if the two methods have similar results for larger samples, we multiplied the counts in Table 2 by 10. Using this larger record-level data, we again used the Fox et al. [21] code with 100,000 iterations. The analysis took approximately 40 minutes to run. The analysis using

`unmconf`, with 25,000 iterations and a burn in of 4,000 across 4 chains took just under 20.5 minutes to run. The results from both packages are displayed in Table 3. Not surprisingly, for this larger data set, the 95% intervals are more similar than in the small sample case.

The code supplied by Fox et al. [21] is currently structured to handle the case of modeling with no measured covariates. If a researcher’s data matches this code structure, then there is no concern from the user to conduct a sensitivity analysis. If other covariates are in a data set, then the authors assume that the researchers interested in conducting a simulation are proficient enough with R to modify the scripts in order to match their data. This may not always be the case. Our framework for modeling in `unmconf` allows for user ease in adjusting the model as needed and adding measured covariates. The modeling function, `unm_glm()`, carries a structure much like `glm()` in R, and the researchers can simply add measured covariates to the right hand side of the equation if desired.

Simulation

A useful aspect of the `unmconf` package is that it provides a convenient framework to perform simulation studies to determine how large validation sample sizes need to be. In practice, these sizes might be fixed, but researchers could see in advance how precise inference will be and if supplementing with other external information might be needed. To assess the performance of the proposed Bayesian unmeasured confounding model with different levels of information on unmeasured confounder(s), coverage probabilities, average bias, and average lengths of 95% quantile-based credible intervals were compared against a naive model for simulated data sets. The naive model here ignores all unmeasured confounders and is of the form

$$Y|x, c \sim D_Y(g_Y^{-1}(v'\beta), \xi_y) = D_Y(g_Y^{-1}(\beta_0 + \beta_x x + \beta_{c_1} c_1 + \dots + \beta_{c_p} c_p), \xi_y).$$

In logistic regression, we note that the difference between the parameter estimate from the naive model and the parameter estimate from the confounder-adjusted model can come from a combination of confounding bias and the noncollapsibility effect [29]. To avoid confusion of what is meant by bias we consider the difference between the conditional odds ratio of the estimated model (either naive or corrected) and the “true” conditional odds ratio for the full model. For further understanding of noncollapsibility and quantifying the bias that derives from confounding bias in logistic regression models, refer to Schuster et al. [30],

Pang et al. [31], and Janes et al. [32]. We compare the Bayesian unmeasured confounding model and the naive model under the presence of internal and external validation data.

Sensitivity analysis – internal validation data

Performance metrics of coverage, bias, and length were assessed for combinations of $n = 500, 1000, 1500, 2000$, internal validation data of $p = 15\%, 10\%, 5\%, 2\%$, $response = norm, bin, gam, pois$, $u_1 = norm, bin$, and $u_2 = norm, bin$ across $m = 1000$ data sets, where p denotes the fraction of the main study used for validation. The data is generated as follows. First, a single, perfectly measured covariate, z , was modeled as a standard normal for the desired sample size, n . The unmeasured confounders were then generated independently as either a normal distribution with a mean and variance of 1 or a Bernoulli distribution with probability of success being 0.7 depending on the family requested in the simulation. The binary exposure variable is generated conditioned on the unmeasured confounders and perfectly measured covariate with the parameters $\theta_E = (\eta_1 = -1, \eta_z = .4, \eta_{u_1} = .75, \eta_{u_2} = .75)$ in the inverse logit link of the exposure model. Lastly, the response model is either generated normally with variance of 1, Bernoulli, Poisson, or Gamma with shape parameter $\alpha_y = 2$, given the requested family, with parameters $\theta_R = (\beta_1 = -1, \beta_x = .75, \beta_z = .75, \lambda_{u_1} = .75, \lambda_{u_2} = .75)$ in the inverse link function. Thus, we have the design points $\theta = (\theta_R, \theta_E)$ for our data generation. Once all variables are modeled for our data sets, we denote a proportion, $1 - p$, of the unmeasured confounders observations as missing. Our JAGS model results are based on 40000 posterior samples across 4 chains with a burn in of 4000. The algorithm for this simulation study is as follows:

1. Select parameter values for θ .
2. Generate m data sets on combinations of n, p, u_1, u_2 using `runm()` for selected parameter values.
3. Build the model and call JAGS using `unm_glm()` with `n.iter = 10000`, `n.adapt = 4000`, `n.thin = 1`, `n.chains = 4`.
4. For each data set, monitor all parameters and whether the “true” parameters are contained in the 95% credible intervals.
5. Compute the average posterior mean, average posterior standard error, coverage, average bias, and average credible interval length across replications.

6. Assess convergence of the Bayesian model.

We note that, in step 1 above, using either validation data or informative priors to set initial values for the bias parameters on the “correct side of 0” leads to better results in terms of convergence. The simulation fits models with default priors, thus no informative priors. Output is only displayed for a binary response and two binary unmeasured confounders. We leave the Bayesian sensitivity analysis for all other combinations of responses (normal, binary, Poisson, gamma) and unmeasured confounders (normal, binary) as supplemental material.

For all levels of internal validation investigated, we obtain near nominal coverage. The coverage for the naive model is generally well below nominal, as noted in Table 4. For small sample sizes and small internal validation proportions, the average bias is similar to the naive model. Rarely a study is performed with only 2% internal validation for a sample size of $n = 500$ being low, but we choose to incorporate the results to display that the model may not be robust under extreme circumstances. The credible interval length largely increases with the decrease in proportion of internal validation

data acquired, which demonstrates why smaller validation samples have higher coverage in our study. For internal validation data at 2%, the 95% credible intervals may not be of practical significance due to the range of values covered when the true value of β_x is 0.75. For instance, the average length of the 95% credible intervals when the response is binary, the two unmeasured confounders are binary, and the sample size is 500 was 3.601. As the sample size increases and internal validation data increases for any combination of response and unmeasured confounders, the bias shows a trend of decreasing towards zero. For these parameters values of θ , the naive model overestimates the truth to a larger magnitude relative to any instance where the unmeasured confounders are accounted for in the model. This may not be the same direction of bias for other values of θ . The median credible interval length and median bias are also displayed in Table 4.

Sensitivity analysis – external validation data

Performance metrics of coverage, bias, and length were assessed for combinations of $n = 500, 1000, 1500, 2000$, external validation data with $p = 50\%, 100\%$ of the

Table 4 Breakdown of 95% Credible Intervals when $\beta_x = 0.75$ for the case of a binary response and two binary unmeasured confounders

n	Validation	Coverage	Avg. CI length	Med. CI length	Avg. bias	Med. bias
500	Naive	0.846	0.781	0.780	0.198	0.202
500	2%	0.994	3.601	3.453	0.189	0.192
500	5%	0.978	2.223	2.093	0.161	0.169
500	10%	0.951	1.451	1.408	0.080	0.077
500	15%	0.968	1.169	1.150	0.061	0.063
500	Full	0.954	0.815	0.814	0.020	0.022
1000	Naive	0.728	0.550	0.549	0.189	0.187
1000	2%	0.978	2.390	2.266	0.168	0.172
1000	5%	0.959	1.291	1.242	0.081	0.093
1000	10%	0.956	0.877	0.862	0.036	0.041
1000	15%	0.951	0.749	0.740	0.025	0.023
1000	Full	0.942	0.572	0.572	0.006	0.008
1500	Naive	0.629	0.449	0.448	0.187	0.188
1500	2%	0.961	1.757	1.663	0.104	0.122
1500	5%	0.954	0.948	0.926	0.042	0.057
1500	10%	0.942	0.677	0.669	0.016	0.023
1500	15%	0.939	0.590	0.586	0.015	0.017
1500	Full	0.938	0.466	0.466	0.004	0.004
2000	Naive	0.523	0.388	0.388	0.188	0.187
2000	2%	0.958	1.430	1.374	0.096	0.113
2000	5%	0.955	0.772	0.753	0.038	0.050
2000	10%	0.946	0.566	0.560	0.019	0.022
2000	15%	0.953	0.502	0.500	0.011	0.012
2000	Full	0.948	0.403	0.404	0.002	0.001

original sample size, response = norm, bin, gam, pois, $u_1 = \text{norm, bin}$, and $u_2 = \text{norm, bin}$ across $m = 1000$ data sets. For the simulation with external validation data, the main study data has no information on the unmeasured confounders. Typically, the external data has information on the unmeasured confounder-response relationship but not the exposure. Thus, informative priors on the bias parameters for this relationship are needed to achieve convergence. That is, using (2), $\gamma_1, \gamma_x, \delta_1$, and δ_x . For the sensitivity analysis, we again model z from a standard normal for the desired sample size, n . The unmeasured confounders were then generated independently as either a standard normal or a Bernoulli distribution with probability of success being 0.5 depending on the family requested in the simulation. Using the notation for the design points, θ , in “Sensitivity analysis – internal validation data” section, $\theta_E = (\eta_1 = -1, \eta_z = .4, \eta_{u_1} = .75, \eta_{u_2} = .75)$ and $\theta_R = (\beta_1 = -1, \beta_x = .75, \beta_z = .75, \lambda_{u_1} = .75, \lambda_{u_2} = .75)$. Once all variables are modeled for our data sets, we designate a proportion, $1 - p$, of the unmeasured confounders observations as missing. We additionally generate external validation data of size np where there is no treatment effect in θ_E (i.e., $\beta_x = 0$). Then, we combine the main study data and external validation data into one large data set. Our JAGS model results are based on 40000 posterior samples across 4 chains, with a burn in of 4000. The algorithm for this simulation study is as follows:

1. Select parameter values for θ .
2. Generate m data sets on combinations of n, p, u_1, u_2 using `runm()` for selected parameter values.
3. Elicit informative priors on γ_X and δ_X .
4. Build the model and call JAGS using `unm_glm()` with `n.iter = 10000`, `n.adapt = 4000`, `n.thin = 1`, `n.chains = 4`. Specify the informative priors through the argument, `priors`.
5. For each data set, monitor all parameters and whether the “true” parameters are contained in the 95% credible intervals.
6. Compute the average posterior mean, average posterior standard error, coverage, average bias, and average credible interval length across replications.
7. Assess convergence of the Bayesian model.

In the algorithm above, the unmeasured confounders were either $u_* \sim N(0, 1)$ or $u_* \sim \text{Bernoulli}(0.5)$ depending on the family requested in the simulation. The simulation fits models with default priors and no informative priors beyond those for γ_X and δ_X . Here, $\gamma_X \sim N(.65, \sigma^2 = 1/.3^2)$ and $\delta_X \sim N(.65, \sigma^2 = 1/.3^2)$. We again leave the the Bayesian sensitivity analysis for all other combinations of responses and unmeasured confounders as supplemental material.

For all levels of external validation investigated, we obtain near nominal coverage. The slight over-coverage likely derives from the increase in uncertainty when the model accounts for missing data through the unmeasured confounders. The naive model under performs in terms of coverage, as noted in Table 5. The credible interval lengths appear to be relatively similar for the naive

Table 5 Breakdown of 95% Credible Intervals when $\beta_x = 0.75$ for the case of a binary response and two binary unmeasured confounders

n	Validation	Coverage	Avg. CI length	Med. CI length	Avg. bias	Med. bias
500	Naive	0.846	0.781	0.780	0.198	0.202
500	250	0.966	0.771	0.766	0.009	0.005
500	500	0.960	0.717	0.715	-0.007	-0.002
500	Full	0.954	0.815	0.814	0.020	0.022
1000	Naive	0.728	0.550	0.549	0.189	0.187
1000	500	0.975	0.559	0.557	0.005	-0.001
1000	1000	0.963	0.524	0.523	-0.022	-0.028
1000	Full	0.942	0.572	0.572	0.006	0.008
1500	Naive	0.629	0.449	0.448	0.187	0.188
1500	750	0.974	0.470	0.467	-0.005	-0.007
1500	1500	0.962	0.443	0.442	-0.020	-0.020
1500	Full	0.938	0.466	0.466	0.004	0.004
2000	Naive	0.523	0.388	0.388	0.188	0.187
2000	1000	0.972	0.419	0.417	-0.007	-0.006
2000	2000	0.971	0.396	0.395	-0.025	-0.028
2000	Full	0.948	0.403	0.404	0.002	0.001

model as with external validation data. The naive model performs much worse in terms of bias, often overestimating the true effect. Yet, the external validation data supplemented with informative priors tends to remove the bias for even half of the number of original observations for all sample sizes tested. Average and median credible interval length as well as average and median bias are also displayed in Table 5.

Discussion

When performing causal inference in observational studies, accounting for the presence of unmeasured confounders has often been overlooked by researchers due to the lack of easily accessible software. Quantitative bias analysis, both deterministic and probabilistic, can quantify the magnitude and direction of the bias on the exposure effect when ignoring the presence of an unmeasured confounder by considering possible scenarios for the unmeasured confounder. Previous work focuses on deterministic QBA, likely due to its simplicity, with no R packages published for probabilistic QBA. Fox et al. [21] contribute R scripts in the supplemental material of their work to provide the first known openly accessible probabilistic QBA with unmeasured confounders. `unmconf` provides a package on CRAN to help resolve the disconnect between methodology for Bayesian sensitivity analysis with unmeasured confounders and its implementation through easily accessible software. A more thorough introduction to the package is provided in the package vignette, accessible via the command `vignette("unmconf", package = "unmconf")`. The package is also documented via R's standard documentation system and provides several examples therein.

We note the limitations of the package through its inability to model more than two unmeasured confounders. However, we combat that limitation by enabling the user to extract the JAGS model from the modeling function in order to adjust as needed. With additional (important) unmeasured confounders, a causal inference assessment may not be feasible. This package is not meant to compare to other Bayesian regression modeling packages such as `brms`. Rather, it merely overcomes the obstacle of modeling unmeasured confounders, which `brms` does not. For future work, we aim to create a similar function structure using the programming language Stan. Stan implements Hamiltonian Monte Carlo and the No-U-Turn Sampler (NUTS), which tends to converge more quickly for high-dimensional models. We hope to have provided a detailed process that can be utilized in epidemiological research to address unmeasured confounders through the discussed Bayesian unmeasured confounding model.

Acknowledgements

Not applicable.

Authors' contributions

R.H. wrote the main manuscript text. All authors reviewed and contributed to the manuscript. J.S. and X.Z. conceived the idea of this project. D.K. and R.H. wrote the foundation of the software package. All authors contributed to the package development. All authors have approved the submitted version of the manuscript.

Funding

Stamey, Kahle, and Hebdon funded by research contract from CSL Behring.

Availability of data and materials

This is an R package on CRAN. The link to the package is here: <https://cran.r-project.org/web/packages/unmconf/index.html>. The package comes with a vignette with working examples.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 19 February 2024 Accepted: 27 August 2024

Published online: 07 September 2024

References

1. Cochran WG. Controlling bias in observational studies: a review. Cambridge University Press; 2006. pp. 30–58. <https://doi.org/10.1017/cbo9780511810725.005>.
2. Rosenbaum PR, Rubin DB. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *J Am Stat Assoc*. 1984;79(387):516–24. <https://doi.org/10.1080/01621459.1984.10478078>.
3. Steenland K. Monte Carlo Sensitivity Analysis and Bayesian Analysis of Smoking as an Unmeasured Confounder in a Study of Silica and Lung Cancer. *Am J Epidemiol*. 2004;160(4):384–92. <https://doi.org/10.1093/aje/kwh211>.
4. Arah OA. Bias Analysis for Uncontrolled Confounding in the Health Sciences. *Annu Rev Public Health*. 2017;38:23–38. <https://doi.org/10.1146/annurev-publhealth-032315-021644>.
5. Fewell Z, Davey Smith G, Sterne JAC. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2007;166(6):646–55. <https://doi.org/10.1093/aje/kwm165>.
6. Groenwold RHH, Sterne JAC, Lawlor DA, Moons KGM, Hoes AW, Tilling K. Sensitivity analysis for the effects of multiple unmeasured confounders. *Ann Epidemiol*. 2016;26(9):605–11. <https://doi.org/10.1016/j.annepidem.2016.07.009>.
7. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. 2006;15(5):291–303. <https://doi.org/10.1002/pds.1200>.
8. Uddin MJ, Groenwold RHH, Ali MS, de Boer A, Roes KCB, Chowdhury MAB, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *Int J Clin Pharm*. 2016. <https://doi.org/10.1007/s11096-016-0299-0>.
9. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *Int J Epidemiol*. 2009;38(6):1662–1673. <https://doi.org/10.1093/ije/dyp278>.
10. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969–85. <https://doi.org/10.1093/ije/dyu149>.

11. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med*. 2007;26(11):2331–47. <https://doi.org/10.1002/sim.2711>.
12. Gustafson P, McCandless LC, Levy AR, Richardson S. Simplified Bayesian Sensitivity Analysis for Mismeasured and Unobserved Confounders. *Biometrics*. 2010;66(4):1129–37. <https://doi.org/10.1111/j.1541-0420.2009.01377.x>.
13. Kawabata E, Tilling K, Groenwold R, Hughes R. Quantitative bias analysis in practice: Review of software for regression with unmeasured confounding. 2022. <https://doi.org/10.1101/2022.02.15.22270975>.
14. Carnegie NB, Harada M, Hill JL. Assessing Sensitivity to Unmeasured Confounding Using a Simulated Potential Confounder. *J Res Educ Eff*. 2016;9(3):395–420. <https://doi.org/10.1080/19345747.2015.1078862>.
15. Dorie V, Harada M, Carnegie NB, Hill J. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Stat Med*. 2016;35(20):3453–70. <https://doi.org/10.1002/sim.6973>.
16. Blackwell, M. A Selection Bias Approach to Sensitivity Analysis for Causal Effects. *Political Analysis*. Cambridge University Press; 2014;22(2):169–82. <https://doi.org/10.1093/pan/mpt006>. Accessed 17 Jan 2024.
17. Cinelli C, Ferwerda J, Hazlett C. Sensemakr: Sensitivity Analysis Tools for OLS in R and Stata. *SSRN Electron J*. 2020. <https://doi.org/10.2139/ssrn.3588978>.
18. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*. 2017;167(4):268. <https://doi.org/10.7326/m16-2607>.
19. Xu R, Frank KA, Maroulis SJ, Rosenberg JM. konfound: Command to quantify robustness of causal inferences. *Stata J Promot Commun Stat Stata*. 2019;19(3):523–50. <https://doi.org/10.1177/1536867x19874223>.
20. Fox MP, MacLehose RF, Lash TL. Best Practices for Quantitative Bias Analysis. In: *Applying Quantitative Bias Analysis to Epidemiologic Data*. Cham: Springer International Publishing; 2021. pp. 441–452. Series Title: *Statistics for Biology and Health*. https://doi.org/10.1007/978-3-030-82673-4_13.
21. Fox MP, MacLehose RF, Lash TL. SAS and R code for probabilistic quantitative bias analysis for misclassified binary variables and binary unmeasured confounders. *Int J Epidemiol*. 2023. <https://doi.org/10.1093/ije/dyad053>.
22. Faries D, Peng X, Pawaskar M, Price K, Stamey JD, Seaman JW. Evaluating the Impact of Unmeasured Confounding with Internal Validation Data: An Example Cost Evaluation in Type 2 Diabetes. *Value Health*. 2013;16(2):259–66. <https://doi.org/10.1016/j.jval.2012.10.012>.
23. Stamey JD, Beavers DP, Faries D, Price KL, Seaman JW. Bayesian modeling of cost-effectiveness studies with unmeasured confounding: a simulation study. *Pharm Stat*. 2013;13(1):94–100. <https://doi.org/10.1002/pst.1604>.
24. Lin DY, Psaty BM, Kronmal RA. Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies. *Biometrics*. 1998;54(3):948. <https://doi.org/10.2307/2533848>.
25. Zhang X, Faries DE, Boytsov N, Stamey JD, Seaman JW. A Bayesian sensitivity analysis to evaluate the impact of unmeasured confounding with external data: a real world comparative effectiveness study in osteoporosis. *Pharmacoepidemiol Drug Saf*. 2016;25(9):982–92. <https://doi.org/10.1002/pds.4053>.
26. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall/CRC; 2013. <https://doi.org/10.1201/b16018>.
27. Bedrick EJ, Christensen R, Johnson W. A New Perspective on Priors for Generalized Linear Models. *J Am Stat Assoc*. 1996;91(436):1450–60. <https://doi.org/10.1080/01621459.1996.10476713>.
28. Christensen R, Johnson W, Branscum A, Hanson TE. *Bayesian Ideas and Data Analysis*. Boca Raton: CRC Press; 2010. <https://doi.org/10.1201/9781439894798>.
29. Mood C. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *Eur Sociol Rev*. 2010;26(1):67–82. <https://doi.org/10.1093/esr/jcp006>.
30. Schuster NA, Twisk JWR, Ter Riet G, Heymans MW, Rijnhart JJM. Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC Med Res Methodol*. 2021;21(1):136. <https://doi.org/10.1186/s12874-021-01316-8>.
31. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res*. 2016;25(5):1925–37. <https://doi.org/10.1177/0962280213505804>.
32. Janes H, Dominici F, Zeger S. On quantifying the magnitude of confounding. *Biostatistics*. 2010;11(3):572–82. <https://doi.org/10.1093/biostatistics/kxq007>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.