



Semiparametric approaches for matched case–control studies with error-in-covariates

Nels G. Johnson¹ · Inyoung Kim²

Received: 7 June 2017 / Accepted: 23 March 2019 / Published online: 3 April 2019

© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The matched case–control study is a popular design in public health, biomedical, and epidemiological research for human, animal, and other subjects for clustered binary outcomes. Often covariates in such studies are measured with error. Not accounting for this error can lead to incorrect inference for all covariates in the model. The methods for assessing and characterizing error-in-covariates in matched case–control studies are quite limited. In this article we propose several approaches for handling error-in-covariates that detect both parametric and nonparametric relationships between the covariates and the binary outcome. We propose a Bayesian approach and two approximate-Bayesian approaches for addressing error-in-covariates that is additive and Gaussian, where the variable measured with error has an unknown, nonlinear relationship with the response. The Bayesian approaches use an approximate latent variable probit model. All methods are developed using the nonparametric method of low-rank thin-plate splines. We assess the performance of each method in terms of mean squared error and mean bias in both simulations and a perturbed example of 1–4 matched case-crossover study.

Keywords Bayesian methods · Latent variable probit · Mixed model · Thin-plate splines

1 Introduction

Case–control studies are retrospective studies where the response variable Y is dichotomous, e.g. presence or absence of some disease or injury. Subjects where $Y = 1$ are called cases and subjects where $Y = 0$ are called controls. Often there are poten-

✉ Inyoung Kim
inyoungk@vt.edu

¹ National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN, USA

² Department of Statistics, Virginia Tech, 410-A Hutcheson Hall, Blacksburg, VA 24061-0439, USA

tial confounding variables that are not of interest. Subjects with similar responses on these variables are considered part of the same stratum S . Matching subjects based on their stratum can reduce the effect of the confounding. A case–control study where 1 case is matched with M controls within the same stratum is called a 1– M matched case–control study (Agresti 2002; Hosmer and Lemeshow 2000). A special case of the matched case–control study is a matched case–crossover study where the stratum is the subject (Woodward 2013). Matched case–control studies are popular in public health, biomedical, and epidemiological applications, e.g., vaccine studies (Whitney et al. 2006), organ transplant studies (Peleg et al. 2007), and studies on traffic safety (Tester et al. 2004).

A semiparametric model for matched case–control studies with covariates measured with error is,

$$P(Y = 1 | \tilde{X}, Z, S) = H^{-1}[m^*(\tilde{X}, Z) + q(S)], \quad (1.1)$$

where $H(\cdot)$ is the link function, $q(S)$ is the effect of stratum S , $m^*(\cdot, \cdot)$ is some function of Z , the covariates measure without error, and \tilde{X} , the covariates measured with error. As matched case–control studies are retrospective studies, H is chosen to be the logit link function since it is the only link function that can be used to recover the prospective model (Scott and Wild 1997). Often the model is analyzed using conditional logistic regression in order to avoid estimating $q(S)$. An alternative approach, and the approach we take in this paper, is to estimate the prospective model directly as a longitudinal binary outcomes and model $q(S)$ as a random effect.

There is some existing work on the error-in-covariates for these models. For matched case–control studies analyzed using conditional logistic regression there is the work of McShane et al. (2001), Guolo and Brazzale (2008) and the related work on partial likelihood models of Huang and Wang (2000, 2001) which also account for functional covariate relationships. There are structural approaches where the unknown true covariate is parametrically modeled (Guolo 2008). These methods require knowledge of the true exposure rate and the measurement error distribution, including parameters. Others use functional approaches where the unobserved true covariate is unknown, but considered to be fixed and consequently no assumption is made regarding the distribution of the unobserved true covariate (Buzas and Stefanski 1996; Stefanski and Carroll 1987). There are texts available for a thorough review of non-Bayesian and Bayesian error-in-covariate methods (Carroll et al. 2006; Gustafson 2003).

There are some related Bayesian methodologies for measurement error in covariates, but none of them handle the clustered binary outcomes of the problem we confront. Berry et al. (2002) used smoothing splines and regression splines in the classical measurement error problem to a linear model set up, but not to the important case of binary data. Carroll et al. (2004) used Bayesian spline-based regression when an instrument is available for all study participants. In addition, both papers assumed that the unknown X is normally distributed. Sinha et al. (2010) proposed a semiparametric Bayesian method for handling measurement error under logistic regression setting. They developed a flexible Bayesian method where not only is the relationship between the disease and exposure variable treated semiparametrically, but also the relationship between the surrogate and the true exposure is modeled semiparametri-

cally. The two nonparametric functions are modeled simultaneously via B-splines. Ryu et al. (2011) also proposed nonparametric regression analysis in a generalized linear model (GLM) framework for data with covariates that are the subject-specific random effects of longitudinal measurements. They proposed to apply Bayesian nonparametric methods including cubic smoothing splines or P-splines for the possible nonlinearity and use an additive model setting. The posterior model space is explored through a Markov chain Monte Carlo (MCMC) sampler. Bartlett and Keogh (2018) first gave an overview of the Bayesian approach to handling covariate measurement error under parametric model setting, and contrast it with regression calibration, arguably the most commonly adopted approach. Bartlett and Keogh (2018) then demonstrated why the Bayesian approach has a number of statistical advantages compared to regression calibration and demonstrate that implementing the Bayesian approach is usually quite feasible for the analyst.

We adapt a fully Bayesian approach for covariate measurement error in semiparametric regression models for normal responses Y (Berry et al. 2002) for use with matched case–control studies, which treats \tilde{X} as a latent variable to be integrated over. This is similar to the work of Sinha et al. (2005) who take a Bayesian approach to error-in-covariates in conditional logistic linear regression models. We also develop two approximate-Bayesian approaches which use a first order Laplace approximation (Tierney and Kadane 1986) to marginalize \tilde{X} out of the likelihood.

For estimating $m^*(\cdot, \cdot)$, we assume $m^*(\tilde{X}, Z) = m(x) + Z\beta_z$, where $m(\cdot)$ is a smooth function that can be approximated by the user's favorite spline method, and where only one variable x is measured with error. Our focus will then be on a semiparametric mixed model approach for estimating $m(x) + Z\beta_z$ that addresses covariate measurement error in x in $1-M$ matched case–control studies. Existing methods for characterizing error-in-covariates in models with clustered binary outcomes cannot estimate nonparametric relationships between the clustered binary outcome and covariates measured with error. Hence, the method we propose are unified approaches in their ability to handle error-in-covariates and detect both parametric and nonparametric relationships between clustered binary outcome and error-in-covariates. The Bayesian approaches are for computational convenience developed using a latent variable probit model (Albert and Chib 1993) with a scaled linear predictor to approximate a logit model (Camilli 1994). All are developed using low-rank thin-plate splines (Ruppert et al. 2003). We show through both simulations and a perturbed example of a $1-4$ matched case–control study that the (fully) Bayesian approach performs similarly to the approximate-Bayesian approaches, except under model misspecification, where it tends to perform better.

This article is organized as follows: In Sect. 2, we describe a semiparametric mixed model with error-in-covariates and estimate it using low-rank thin-plate splines. In Sect. 3, we develop the approximate- and fully Bayesian approaches based on a latent variable probit approximation to a logistic model. In Sect. 4, we conduct a simulation study to compare our methods. In Sect. 5, we apply each approach to a $1-4$ matched case–case control study for juvenile aseptic meningitis. Section 6 contains concluding remarks and possible future work.

2 Semiparametric mixed model with error-in-covariates

We can approximate $m(x)$ using some basis function method such that $m(x) \approx B(x)\beta_B$, where $B(x)$ is a matrix of the basis function and β_B are the basis coefficients. For the purposes of this paper using low-rank thin-plate splines (Ruppert et al. 2003) with order p (chosen to be some natural number) and knots $(\xi_1, \xi_2, \dots, \xi_\kappa)$, $\kappa < N \times (M + 1)$, chosen a priori. This produces the following linear model,

$$\begin{aligned} m(x) &\approx B(x)\beta_B \\ B(x) &= [X^* L_p^*(x)], \\ \beta_B &= [\beta_x^* \beta_L^*]^T \\ X^* &= [1 \ x \ \dots \ x^{p-1}], \\ L_p^*(x) &= [|x - \xi_1|^{2p-1} \ |x - \xi_2|^{2p-1} \ \dots \ |x - \xi_\kappa|^{2p-1}]. \end{aligned}$$

The penalty on β_L^* is treated as the prior in a Bayesian framework. For low-rank thin-plate splines this is $N(0, \sigma_\beta^2 \Omega^{-1})$, where the (r, c) th element of the penalty matrix Ω is $|\xi_r - \xi_c|^{2p-1}$. However, Ω is not positive definite, and thus an invalid covariance matrix. To address this, singular value decomposition is used to find $(\Omega^{-1/2})^T (\Omega^{-1/2}) = \Omega^{-1}$. We scale $\Omega^{1/2} \beta_L^* = \beta_L$ and $L_p^*(x) \Omega^{-1/2} = L_p(x)$ so that $L_p(x)$ is an orthogonal basis with prior distribution (i.e. penalty) $N(0, \sigma_\beta^2 I)$ on β_L (Crainiceanu et al. 2005).

For error-in-covariates in matched case-control studies, we assume that we observe $w_{ijk} = x_{ij} + u_{ijk}$, however x_{ij} is unobserved, the measurement error $u_{ijk} \sim N(0, \sigma_u^2)$, $i = 1, 2, \dots, N$ (the number of strata), $j = 1, 2, \dots, M + 1$ (the number of subjects in each strata), and $k = 1, 2, \dots, K_{ij}$ is the number of replicated measurements for subject j in strata i . In order to properly estimate σ_u^2 , K_{ij} must be greater than or equal to 2 for at least one ij .

To ease computations we approximate the logistic link function by using a latent variable probit model (Albert and Chib 1993) where the linear predictor is scaled by $\sqrt{\pi/8}$ (Camilli 1994). The model is as follows,

$$\begin{aligned} y_{ij} | x_{ij}, Z_{ij}, S_i &\sim \text{Bernoulli}(\pi_{ij}), \\ \pi_{ij} &= \text{logit}^{-1}(\eta_{ij}), \\ &\approx \Phi(l_{ij}), \\ l_{ij} | y_{ij} &\sim \begin{cases} \text{Normal}^+ (\eta_{ij} \sqrt{\pi/8}, 1), & y_{ij} = 1 \\ \text{Normal}^- (\eta_{ij} \sqrt{\pi/8}, 1), & y_{ij} = 0 \end{cases}, \\ \eta_{ij} &= m(x_{ij}) + Z_{ij} \beta_z + q(S_i), \\ &\approx X_{ij}^* \beta_x^* + L_p(x_{ij}) \beta_L + Z_{ij} \beta_z + q(S_i), \\ q(S_i) &\sim \text{Normal}(0, \sigma_q^2), \\ \beta_L &\sim \text{Normal}(0, \sigma_\beta^2), \\ w_{ijk} | x_{ij} &\sim \text{Normal}(x_{ij}, \sigma_u^2), \end{aligned}$$

where l is a latent variable, $\text{Normal}^+(\cdot, 1)$ and $\text{Normal}^-(\cdot, 1)$ are truncated normal distributions, to the left and to the right of zero, respectively.

3 Methods

We develop a fully Bayesian (FB) approach and two approximate-Bayesian (AB) approaches using first order Laplace approximation in Sects. 3.1 and 3.2, respectively.

3.1 Fully Bayesian approach

To improve computations, we let the intercept β_0 be absorbed into $q(S)$. We work with X_{ij} , which is X_{ij}^* without the column of ones, and β_x , which is β_x^* without β_0 . The response Y depends on the regression parameters through l . Thus, a natural parametrization of the likelihood for modeling additive Gaussian measurement error is as follows,

$$L(W, l|Y, x, Z, \beta, \sigma_u^2) \propto \prod_{i=1}^N \prod_{j=1}^{M+1} \left\{ \text{Normal} \left[l_{ij}; \eta_{ij} \sqrt{\pi/8}, 1 \right] \times \left[\delta_{(l_{ij} \geq 0)} \delta_{(y_{ij}=1)} + \delta_{(l_{ij} < 0)} \delta_{(y_{ij}=0)} \right] \times N(W_{ij}; x_{ij}, \sigma_u^2 I_{K_{ij} \times K_{ij}}) \right\}$$

$$\eta_{ij} = Z_{ij} \beta_z + X_{ij} \beta_x + L_p(x_{ij}) \beta_L + q(S_i).$$

As mentioned previously, the prior on β_L should be chosen to be $\pi(\beta_L | \sigma_\beta^2) \sim N(0, \sigma_\beta^2)$, where σ_β^2 is a hyperparameter. In practice, the prior distributions placed on x and on $q(S)$ should be chosen to reflect the data collected. For instance, a flexible approach might model x using a mixture of normals (Carroll et al. 1999). For this article, we choose $\pi(x | \mu_x, \sigma_x^2) \sim N(\mu_x, \sigma_x^2)$ and $\pi[q(S) | \beta_0, \sigma_q^2] \sim N(\beta_0, \sigma_q^2)$, where $\mu_x, \beta_0, \sigma_x^2$, and σ_q^2 are hyperparameters. The prior distributions for the other parameters are: $\pi(\sigma_u^2) \sim IG(\sigma_u^2; A_u, B_u)$, and $\pi[\beta_z, \beta_x] \sim N(\beta_z, \beta_x; g_\beta, t_\beta^2)$. Finally, we take the prior distributions for the hyper parameters as follows: $\pi(\mu_x) \sim N(\mu_x; g_\mu, t_\mu^2)$, $\pi(\sigma_x^2) \sim IG(\sigma_x^2; A_x, B_x)$, $\pi(\sigma_\beta^2) \sim IG(\sigma_\beta^2; A_{\sigma_\beta^2}, B_{\sigma_\beta^2})$, $\pi(\beta_0) \sim N(\beta_0; g_0, t_0^2)$, and $\pi(\sigma_q^2) \sim IG(\sigma_q^2; A_q, B_q)$. Both the likelihood and prior structure are adapted from existing work where Y is continuous (Berry et al. 2002), which defaults to normal priors on mean-like parameters and inverse-gamma priors on variance parameters. In practice, careful choice of an informative prior structure can further improve inference. For example, it may be more appropriate to restrict the support of the prior on σ_u to be less than the value of σ_w since σ_w should be greater than σ_u when errors are additive and Gaussian.

We use Metropolis–Hastings (Metropolis et al. 1953; Hastings 1970) and Gibbs (Geman and Geman 1984) algorithms to sample the joint posterior of these parameters using Markov chain Monte Carlo (MCMC). The joint posterior distribution of x uses a Metropolis–Hastings step, while all other parameters can be sampled using Gibbs steps. The conditional posterior distributions for each parameter along with the proposal distribution for x_{ij} can be found in Appendix A.

3.2 Approximate-Bayesian approaches

Updating each x_{ij} via Metropolis–Hastings can be time consuming computationally, especially for large $N \times M$. We propose two approximate-Bayesian (denoted AB1 and AB2) approaches to reduce computation time, by integrating each x_{ij} out of the conditional likelihood a priori. This integration is intractable due to the spline portion of the linear predictor. We use a first order Laplace approximation (Tierney and Kadane 1986) to solve the integration.

For AB1, we place an improper flat prior on x_{ij} and find:

$$\int L(l_{ij}, W_{ij}|Y_{ij}, x_{ij}, Z_{ij}, S, \beta, \sigma_u^2) dx_{ij} \approx L(l_{ij}|Y_{ij}, x_{ij} = \bar{w}_{ij\cdot}, Z_{ij}, S, \beta),$$

where $\bar{w}_{ij\cdot} = K_{ij}^{-1} \sum_{k=1}^{K_{ij}} w_{ijk}$. See Appendix 2.1 for derivation.

For AB2, we place prior a normal prior on x_{ij} (as in Sect. 3.1) and find:

$$\begin{aligned} &\int L(l_{ij}, W_{ij}|Y_{ij}, x_{ij}, Z_{ij}, S, \beta, \sigma_u^2) N(x_{ij}|\mu_x, \sigma_x^2) dx_{ij} \\ &\approx L(l_{ij}|Y_{ij}, x = \tilde{w}_{ij\cdot}, Z_{ij}, S, \beta), \end{aligned}$$

where $\tilde{w}_{ij\cdot} = \frac{K_{ij} \bar{w}_{ij\cdot} \sigma_x^2 + \mu_x \sigma_u^2}{K_{ij} \sigma_x^2 + \sigma_u^2}$. See Appendix 2.1 for derivation. The parameters μ_x , σ_x^2 , and σ_u^2 need to be estimated a priori. We recommend setting each to the MLE:

$$\begin{aligned} \mu_x &= \bar{w}\dots, \\ \sigma_u^2 &= \sum_{i=1}^N \sum_{j=1}^{M+1} \sum_{k=1}^{K_{ij}} [N(M+1)(K_{ij}-1)]^{-1} (w_{ijk} - \bar{w}_{ij\cdot})^2, \quad \text{and} \\ \sigma_x^2 &= \sum_{i=1}^N \sum_{j=1}^{M+1} [N(M+1)-1]^{-1} (\bar{w}_{ij\cdot} - \mu_x)^2 - \sigma_u^2. \end{aligned}$$

Both approximate-Bayesian methods produce simple ways of handling error-in-covariates, equivalent to using a plug-in estimator w^* and proceeding as if no covariate measurement error. To be clear, we use the following likelihood in the approximate-Bayesian analysis:

$$L(Y|l, w^*, Z, \beta) \propto \prod_{i=1}^N \prod_{j=1}^{M+1} N[l_{ij}; \eta_{w,ij} \sqrt{\pi/8}, 1] \\ \times [\delta_{(l_{ij} \geq 0)} \delta_{(y_{ij}=1)} + \delta_{(l_{ij} < 0)} \delta_{(y_{ij}=0)}], \\ \eta_{w,ij} = Z_{ij} \beta_z + W_{ij}^* \beta_x + L_p(w_{ij}^*) \beta_L + q(S_i),$$

where $w_{ij}^* = \bar{w}_{ij}$. for AB1 and $w_{ij}^* = \frac{m\bar{w}_{ij} \cdot \sigma_x^2 + \mu_x \sigma_u^2}{K_{ij} \sigma_x^2 + \sigma_u^2}$ for AB2, and $W_{ij}^* = (w_{ij,1}^*, w_{ij,2}^*, \dots, w_{ij,p-1}^*)$.

The prior structure we adopt for the rest of this model, i.e. β , $q(S)$, and their hyperparameters, is the same as in Sect. 3.1. We then obtain the same conditional posteriors for them as well.

4 Simulation study

To assess the adequacy of each approach for correcting for covariate measurement error, we conducted a simulation study to address performance in terms of minimizing both the mean squared error (MSE) and the mean bias. We considered the fully Bayesian approach of Sect. 3.1 and the two approximate-Bayesian approach of Sect. 3.2. In Sect. 4.1 we address model performance when the assumptions concerning the covariate measurement error are met. In Sect. 4.2 we address the robustness of each method when there is model misspecification error in the distribution of x and u . In Sect. 4.3 we describe the results.

For all simulations we set $K_{ij} = 2$ for all ij , $M = 4$. We look at only a single covariate z measured without error, with $\beta_z = -0.5$. We simulate $z \sim N(0, 1)$ and $q(S) \sim N(0, 0.1^2)$. We look at two functions $m(x) = x^2/6$ and $m(x) = \sin(\pi x/2)$. The quadratic function was chosen for its simplicity and because quadratic models are usually fit parametrically with a linear term, while the sinusoidal pattern was chosen for its similarity to the relationship found in our juvenile aseptic meningitis data described in Sect. 5. To generate the clustered binary outcomes, sets of $1 + M$ binary outcomes were generated from $P(Y = 1|x, z, S, \beta_z) = \Phi[m(x) + z\beta_z + q(S)]$ until a set is found such that $\sum_{j=1}^{1+M} Y_j = 1$. This was repeated for each of the N strata. And again repeated to produce 100 datasets for each simulation setup.

It should be noted that the choice of probit link function for data generation is arbitrary. We use the logit link function for analysis because the estimated parameters are the same whether the data were collected prospectively or retrospectively, and not because said estimated parameters will be from the “true model,” if such a thing is knowable.

For each Bayesian approach, we use the same prior structure as noted in Sect. 3.1, where $\{A_u, B_u, A_x, B_x, A_{\sigma_\beta^2}, B_{\sigma_\beta^2}, A_q, B_q\} = 0.1$, $\{g_\mu, g_\beta, g_0\} = 0$, and $\{t_\mu^2, t_\beta^2, t_0^2\} = 5^2$. For estimation, we use low-rank thin-plate splines with $\kappa = 10$ knots, chosen at evenly spaced percentiles of \bar{w} , and with order $p = 2$, for all methods. The mean squared error $\sum_{i=1}^N \sum_{j=1}^{M+1} (\hat{\eta}_{ij}^{(c)} - \hat{\eta}_{ij}^{(T)})^2$ and mean bias

$\sum_{i=1}^N \sum_{j=1}^{M+1} (\hat{\eta}_{ij}^{(\cdot)} - \hat{\eta}_{ij}^{(T)})$ is computed for each simulation dataset, where $\hat{\eta}^{(\cdot)}$ is the estimated linear predictor using one of the proposed methods, and $\hat{\eta}^{(T)}$ is the estimated linear predictor for the fully Bayesian approach with perfect measurements for x .

For all simulations, all Bayesian methods were run for 10000 iterations with the first 2000 treated as burn-in. These values were determined graphically from the results of repeated test cases for each simulation combination. Every 10th iteration was kept after burn-in. Acceptance rates for the x_{ij} s averaged around 0.4 across all simulations. Simulations were run using Matlab 2012a (MATLAB 2012) and GNU Octave (Eaton et al. 2008). For code, please contact the authors.

4.1 Correctly specified model

In simulations where the measurement error distribution and distribution of x are correctly specified, we generate each x_{ij} from a standard normal and each u_{ijk} such that $\sigma_u = \{0.1, 0.3, 0.5\}$, corresponding to small, large, and very large amounts of measurement error when $\sigma_x = 1$ (Parker et al. 2010). We also consider small and large sample situations with the number of strata $N = \{25, 100\}$.

4.2 Model mis-specification

We consider three cases of model misspecification, one where only the distribution of x is misspecified, one where only the distribution of u is misspecified, and one where the distribution of both x and u are misspecified:

- $2^{3/2} \times (x + 4) \sim \chi_4^2$ and $u \sim N(0, \sigma_u = 0.5)$
- $x \sim N(0, 1)$ and $u \sim \text{Laplace}[0, \text{scale} = 2^{-3/2}]$
- $2^{3/2} \times (x + 4) \sim \chi_4^2$ and $u \sim \text{Laplace}[0, \text{scale} = 2^{-3/2}]$

The misspecified distributions are chosen such that $\sigma_u/\sigma_x = 0.5$ for all cases. For this set, we consider a moderate sample size with the number of strata $N = 50$.

4.3 Simulation results

Tables 1 and 2 present the results for when the distribution of x and u are correctly specified, for the quadratic $m(x) = x^2/6$ and sinusoidal $m(x) = \sin(\pi x/2)$ cases, respectively. We observe that for the quadratic cases, no method consistently reduces the bias or MSE over the other. However, it should be noted that the fully Bayesian approach is never the worst at reducing MSE. For the sinusoidal cases the fully Bayesian approach is worst at reducing MSE in one case, where $N = 100$ and $\sigma_u = 0.1$, but it is the best choice for reducing MSE and bias for all cases where $\sigma_u = \{0.3, 0.5\}$. These results suggest that the fully Bayesian is at least as good as both approximate-Bayesian approaches, particularly at reducing MSE, and it is not clear which approximate-Bayesian approach is the better than the other.

Tables 3 and 4 present the results for when the distribution of x and u are incorrectly specified, for the quadratic $m(x) = x^2/6$ and sinusoidal $m(x) = \sin(\pi x/2)$ cases, respectively. We observe that for the quadratic cases, the fully Bayesian approach

Table 1 The $MMB \times 10^2$ (mean mean bias) and the $MMSE \times 10^2$ (mean mean squared error) of the linear predictors $\hat{\eta}^{(\cdot)}$ for the case where $m(x) = x^2/6$ for comparing the approximate-Bayesian methods with flat (AB1) and normal (AB2) prior on x and the fully Bayesian method (FB)

σ_u	N		AB1	AB2	FB
0.1	25	MMB	0.0812	-0.0268	-0.3399
		MMSE	0.8200	0.7881	0.7530
	100	MMB	0.0795	0.1776	0.0236
		MMSE	0.3634	0.3899	0.3827
0.3	25	MMB	-0.2167	-0.5135	-1.0846
		MMSE	4.8584	5.0369	4.9435
	100	MMB	0.2223	-0.0003	-0.0768
		MMSE	1.9779	1.9631	1.9111
0.5	25	MMB	-0.2700	0.2242	-2.0456
		MMSE	10.8233	10.3909	10.7041
	100	MMB	0.3681	0.4187	-0.2370
		MMSE	4.5485	4.4818	4.3825

Table 2 The $MMB \times 10^2$ (mean mean bias) and the $MMSE \times 10^2$ (mean mean squared error) of the linear predictors $\hat{\eta}^{(\cdot)}$ for the case where $m(x) = \sin(\pi x/2)$ for comparing the approximate-Bayesian methods with flat (AB1) and normal (AB2) prior on x and the fully Bayesian method (FB)

σ_u	N		AB1	AB2	FB
0.1	25	MMB	0.1143	0.2341	-0.1522
		MMSE	1.6226	1.5989	1.6203
	100	MMB	0.1482	0.2456	0.1580
		MMSE	1.1938	1.1994	1.2200
0.3	25	MMB	2.1800	2.0676	0.3136
		MMSE	10.8182	10.7291	10.4553
	100	MMB	1.6864	1.6714	0.5713
		MMSE	8.2323	8.2445	7.8472
0.5	25	MMB	3.6901	4.1200	0.1591
		MMSE	26.6599	26.3365	24.6970
	100	MMB	3.7132	3.8774	1.2278
		MMSE	18.3059	18.3925	17.4123

Table 3 The $MMB \times 10^2$ (mean mean bias) and the $MMSE \times 10^2$ (mean squared error) of the linear predictors $\hat{\eta}^{(\cdot)}$ for assessing robustness of the approximate-Bayesian methods with flat (AB1) and normal (AB2) prior on x , and the fully Bayesian method (FB) to model misspecification of the distribution of x and u when $m(x) = x^2/6$

Distribution of X	Distribution of U		AB1	AB2	FB
Normal	Laplace	MMB	0.1361	0.2477	-0.9131
		MMSE	6.9562	6.8001	6.7604
χ^2	Normal	MMB	-0.2797	-0.1178	-1.2173
		MMSE	7.4255	7.1806	6.8872
χ^2	Laplace	MMB	-0.1482	-0.1652	-1.3037
		MMSE	6.8544	6.6766	6.3983

Table 4 The $MMB \times 10^2$ (mean mean bias) and the $MMSE \times 10^2$ (mean mean squared error) of the linear predictors $\hat{\eta}^{(c)}$ for assessing robustness of the approximate-Bayesian methods with flat (AB1) and normal (AB2) prior on x , and the fully Bayesian method (FB) to model misspecification of the distribution of x and u when $m(x) = \sin(\pi x/2)$

Distribution of X	Distribution of U		AB1	AB2	FB
Normal	Laplace	MMB	3.7703	3.9309	0.6109
		MMSE	21.5896	21.3823	20.1694
χ^2	Normal	MMB	3.2538	3.4483	− 0.4529
		MMSE	20.7194	20.7156	19.6295
χ^2	Laplace	MMB	3.0307	3.1729	− 0.1932
		MMSE	19.9894	19.9152	18.6234

provides better reduction in MSE than both the approximate-Bayesian approaches for all misspecification types. Similarly, approximate-Bayesian approach AB2 dominates AB1 in terms of MSE. The opposite is observed for bias, where approximate-Bayesian approaches provide better reduction than the fully Bayesian for all misspecification types, though one approximate-Bayesian approach does not dominate the other. However, this is not true for the sinusoidal cases where the fully Bayesian approach reduces both the bias and MSE more than the approximate-Bayesian approaches for all misspecification types. Approximate-Bayesian approach AB1 dominates AB2 in terms of bias, however AB2 dominates AB1 in terms of MSE. These results suggest the approximate-Bayesian approaches might be at least as good at reducing bias, however the fully Bayesian approach is at least as good at reducing MSE, both when the distribution of x or u is misspecified.

5 Application: juvenile aseptic meningitis data

We consider the aseptic meningitis data of ADD CITATION. Aseptic meningitis is a viral infection that causes inflammation of the membrane that covers the brain and spinal chord. It is rarely fatal, but can take about two weeks to recover from fully. The study design was for a 1–4 case-crossover study with 211 subjects (i.e., strata). Case-crossover studies are a special case of matched case–control studies where the stratum is the subject. The turbidity of drinking water (i.e., the amount of suspended matter in the water) is believed to affect the risk of aseptic meningitis. For this study water turbidity was measured by a nephelometer. A nephelometer shoots a beam of light at water then measures the scattered light. It then uses a formula to determine the turbidity, measured in nephelometric turbidity units (NTU). The device is susceptible to miscalibration, and can be thrown off by air bubbles that may make water that does not actually contain any suspended particles appear cloudy. This study design was not setup for multiple measurements of NTU, so for illustrative purposes we treat NTU measurements collected on two separate samples dates as error prone measurements of NTU from the sample data of interest. These measurements are centered and scaled across dates. We use the same model specification as used in our simulations.

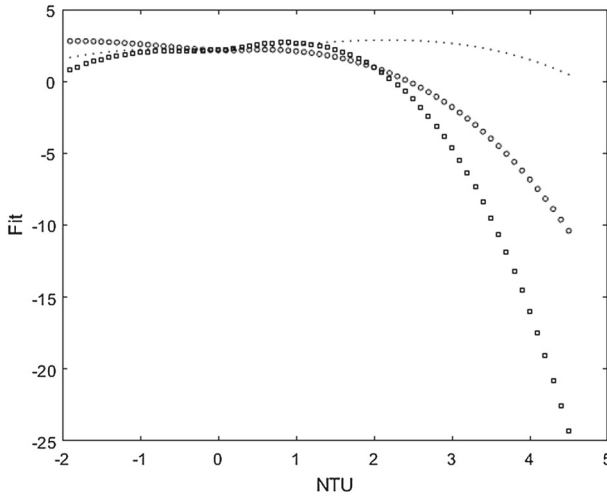


Fig. 1 The posterior-mean fits of $m(x)$ for the aseptic meningitis data, where x is centered and scaled Nephelometric Turbidity Units (NTU). The black dots are the centered (across all methods) fitted values using approximate-Bayesian method 1 (AB1) evaluated over a grid of NTU values. Similarly, the black circles are for approximate-Bayesian method 2 (AB2) and the black squares are for the fully Bayesian method (FB)

For \mathbf{Z} we use the centered and scaled body temperature of the subjects in degrees Celsius.

Figure 1 shows a plot of the posterior mean fitted value (centered across method) for $m(\cdot)$ using the approximate-Bayesian and fully Bayesian methods. The choice of method leads to dramatically different inference concerning the effect of NTU on the probability of acquiring aseptic meningitis, primarily for large measurements of NTU. By using the fully Bayesian method, we do a much better job capturing the decreasing effect of NTU for large values of NTU.

The posterior mean of σ_u from the fully Bayesian model is 0.8854 with a 90% equal tail credible interval of [0.8548, 0.9176]. Given $\sigma_x \approx \sigma_{\bar{w}} = 0.7788$, we are in a large measurement error scenario with $\sigma_u/\sigma_x \approx 1.1369$. From Fig. 2 we can see that the distribution of NTU is not normally distributed. As a result, we made a model misspecification error by placing a normal prior on the distribution of NTU. Given these conditions and our simulations results, we believe the fully Bayesian approach is the best approach for this data.

6 Discussion

We have proposed a fully Bayesian and two approximate-Bayesian approaches for handling a semiparametric mixed model with error-in-covariates for matched case–control studies. These approaches are developed using low-rank thin-plate splines and a latent variable probit model. The strength of these methods is that they can handle both error-in-covariates and explain nonlinear relationships between matched

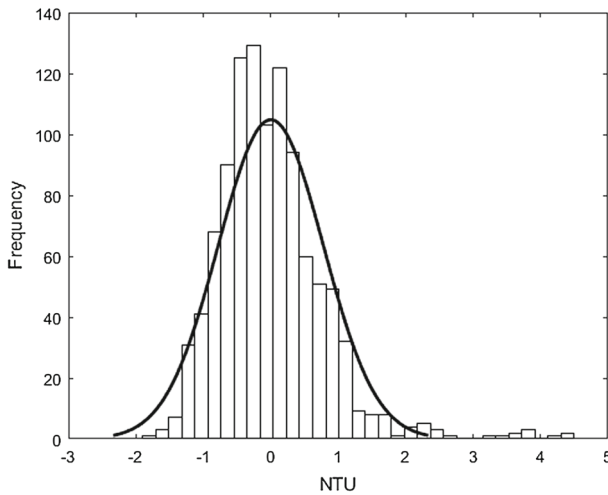


Fig. 2 Histogram of \bar{w} , the subject-specific mean measurement of Nephelometric Turbidity Units (NTU), from the aseptic meningitis data. The black line is a fit from a normal distribution. The measurement for NTU are somewhat symmetric and are unimodal, but normality does not hold

binary outcomes and covariates measured with error. Additionally, we have shown these methods exhibit some robustness to model misspecification of x and u . Based on our knowledge, there is no existing methodology that has been shown to do these in matched case–control studies.

The fully-Bayesian approach treats x as a latent variable and then integrates it out. The approximate-Bayesian approaches use a first order Laplace approximation to the likelihood, marginalizing out x . Deciding which approach to take, AB1, AB2, or FB, can be challenging as it appears somewhat to depend on what the unknown function $m(x)$ is and how well our assumptions are met about the normality of the distribution of x and of u . When all assumptions are met, the fully Bayesian approach tended to perform best more often. However, improvements were often small and not consistent across sample sizes or size of measurement error. As a result, it may not be worth the additional computation for large datasets when there is a reasonable chance it will not actually lead to an improvement. The stronger argument for using the fully Bayesian method is made by its performance under model misspecification, particularly in terms of MSE. If you believe the adage that ‘All models are wrong...,’ then this is the more compelling argument for using the fully Bayesian method, as it outperformed the approximate methods in almost every scenario in terms of mean bias, and every scenario in terms of MSE. Though, again these improvements were often only of modest size. A user may still feel that a faster solution has greater utility than a more accurate one. It is up to the user to consider what is best for their own project.

We note that our approach was developed for the univariate x . Our approach can be generalized for several covariates x measured with error into an additive model,

$$m^*(\tilde{X}, Z) = \sum_{r_1=1}^{R_1} m_{r_1}(x_{r_1}) + \sum_{r_2=1}^{R_2} m_{r_2}(z_{r_2}),$$

where there are R_1 covariates measured with error and R_2 covariates measured without error. Generalization to a nonadditive model will be an interesting and challenging problem because of the unknown interaction structure among unknown covariates. We illustrated our technique using low-rank thin-plate splines, however, it is straight forward to change the spline basis to any other where the smoothness penalty can be thought of as a $N(0, \sigma_\beta^2)$ prior on β_L the spline coefficients (Ruppert et al. 2003).

We assumed that the measurement error u was additive and normally distributed. This assumption creates a computational convenience, as we can choose an inverse-gamma conjugate prior for σ_u^2 so that it can be sampled in a Gibbs step. If we change the distributional assumption on u , this convenience will be lost. More complicated measurement error distributions that may depend on x or Y are worthwhile future research problems. Another choice of computational convenience was to use a rescaled latent variable probit model to approximate the logistic model. This validity of this choice depends the quality of this approximation and the user's personal loss function for tolerating this approximation. However, we gained the ability to use conjugate normal priors on $\beta_x, \beta_z, \beta_L$ and $q(S)$ and sample them using Gibbs steps. Changing the link function would also remove this convenience.

Finally, we assumed the distribution of x was normal. In practice this might not be the case, and was not the case for our data analysis example. We showed the fully and approximate-Bayesian methods were somewhat robust to violations of this assumption, as well as assumptions about the distribution of u . However, flexible methods for properly modeling the distribution of x and u should improve performance of Bayesian error-in-covariates models.

Acknowledgements We would like to thank Pang Du, Leanna House, Scotland Leman, George Terrell, and Matt Williams for their advice and assistance. We would also like to thank Ho Kim for supplying the aseptic meningitis data.

Appendix

A Markov chain Monte Carlo details for implementation

The full posterior conditional distributions are as follows:

- Full conditional for x_{ij} is:

$$[x_{ij}|-] \propto L(l_{ij}, W_{ij}|Y_{ij}, x_{ij}, Z_{ij}, \beta, q(S), \sigma_u^2) \times N(x_{ij}; \mu_x, \sigma_x^2),$$

- Full conditional for σ_u^2 is:

$$[\sigma_u^2|-] \sim IG \left[\sigma_u^2; (1/2) \sum_{i=1}^N \sum_{j=1}^{M+1} K_{ij} + A_u, \right. \\ \left. \sum_{i=1}^N \sum_{j=1}^{M+1} (W_{ij} - x_{ij})^T (W_{ij} - x_{ij})/2 + B_u \right],$$

- Full conditional for μ_x is:

$$[\mu_x|-] \sim N \left\{ \mu_x; \left[t_\mu^2 \sum_{i=1}^N \sum_{j=1}^{M+1} x_{ij} + g_\mu \sigma_x^2 \right] / [N(M+1)t_\mu^2 + \sigma_x^2], \right. \\ \left. t_\mu^2 \sigma_x^2 / [N(M+1)t_\mu^2 + \sigma_x^2] \right\},$$

- Full conditional for σ_x^2 is:

$$[\sigma_x^2|-] \sim IG[\sigma_x^2; N(M+1)/2 + A_x, (x - \mu_x)^T (x - \mu_x)/2 + B_x],$$

- Full conditional for (β_x, β_z) is:

$$[\beta_x, \beta_z|-] \sim MN\{\beta_x, \beta_z; \\ [(X, Z)^T (X, Z) + I/t_\beta^2]^{-1} (X, Z)^T [l - L_p(x)\beta_L - Jq(S)], \\ [(X, Z)^T (X, Z) + I/t_\beta^2]^{-1}\},$$

- Full conditional for β_L is:

$$[\beta_L|-] \sim MN\{\beta_L; \\ [L_p(x)^T L_p(x) + I/\sigma_\beta^2]^{-1} L_p(x)^T [l - X\beta_x - Z\beta_z - Jq(S)], \\ [L_p(x)^T L_p(x) + I/\sigma_\beta^2]^{-1}\},$$

- Full conditional for σ_β^2 is:

$$[\sigma_\beta^2|-] \sim IG(\sigma_\beta^2; \kappa/2 + A_\beta, \beta_L^T \beta_L/2 + B_\beta),$$

- Full conditional for $q(S)$ is:

$$[q(S)|-] \sim MN[q(S); \\ (J^T J + I/\sigma_q^2)^{-1} \{\beta_0/\sigma_q^2 + J^T [l - X\beta_x - Z\beta_z - L_p(x)\beta_L]\}, \\ (J^T J + I/\sigma_q^2)^{-1}],$$

- Full conditional for σ_q^2 is:

$$[\sigma_q^2|-] \sim IG\{\sigma_q^2; N/2 + A_q, [q(S) - \beta_0]^T [q(S) - \beta_0]/2 + B_q\},$$

- Full conditional for β_0 is:

$$[\beta_0|-] \sim N \left\{ \beta_0; \left[t_0^2 \sum_{i=1}^N q(S_i) + g_0 \sigma_q^2 \right] / (N t_0^2 + \sigma_q^2), t_q^2 \sigma_q^2 / (N t_q^2 + \sigma_q^2) \right\},$$

where J is a $N(M+1) \times N$ matrix defined by the Kronecker product $I_{N \times N} \otimes 1_{(M+1) \times 1}$.

When choosing a proposal distribution for x_{ij} , we followed Berry et al. (2002) and used $x_{ij}^{(t)} \sim N(x_{ij}^{(t-1)}, 2^2 \sigma_u^2 / K_{ij})$, where $2\sigma_u / \sqrt{K_{ij}}$ is chosen as the proposal standard deviation because it covers about 95% of the sampling distribution for $\bar{w}_{ij} = K_{ij}^{-1} \sum_{k=1}^{K_{ij}} w_{ijk}$. Alternatively, an automatically tuned proposal distribution (Shaby and Wells 2010) could be used to ensure optimal acceptance rates.

B Derivation of Laplace approximations

2.1 First order Laplace approximation for approximate-Bayesian methods

The goal of this section is to show using first order Laplace approximation that,

$$\int L(l, W|Y, x, Z, S, \beta, \sigma_u^2) dx \approx L(l|Y, x = \bar{w}, Z, S, \beta).$$

Note that:

$$\begin{aligned} L(l_{ij}, W_{ij}|Y_{ij}, x_{ij}, Z_{ij}, S, \beta, \sigma_u^2) &= L(l_{ij}|Y_{ij}, x_{ij}, Z_{ij}, S_i, \beta) \times N(W_{ij}; x_{ij}, \sigma_u^2) \\ &= L(l_{ij}|Y_{ij}, x_{ij}, Z_{ij}, S_i, \beta) (2\pi \sigma_u^2)^{K_{ij}/2} \\ &\quad \times \exp[-(W_{ij} - x_{ij})^T (W_{ij} - x_{ij}) / (2\sigma_u^2)]. \end{aligned}$$

We will write $A(x_{ij}) = L(l_{ij}|Y_{ij}, x_{ij}, Z_{ij}, \beta) (2\pi \sigma_u^2)^{-K_{ij}/2}$ and $h(x_{ij}) = (W_{ij} - x_{ij})^T (W_{ij} - x_{ij}) / (2\sigma_u^2)$. It is easy to show $h(x_{ij})$ has unique maximum \bar{w}_{ij} , since $h(\cdot)$ is a quadratic form, and that the second derivative $h''(x_{ij}) = 1/\sigma_u^2$, both for all ij . Tierney and Kadane (1986) show that we can approximate $\int A(x_{ij}) \exp[-h(x_{ij})] dx_{ij}$ by $A(\tilde{x}) \exp[-h(\tilde{x})] \sqrt{\frac{2\pi}{K_{ij} h''(\tilde{x})}}$, where \tilde{x} is the value that maximizes $h(\cdot)$. We then get:

$$\begin{aligned} \int A(x_{ij}) \exp[-h(x_{ij})] dx_{ij} &\approx L(l_{ij}|Y_{ij}, \bar{w}_{ij}, Z_{ij}, S_i, \beta) \left[K_{ij} (2\pi \sigma_u^2)^{K_{ij}-1} \right]^{-1/2} \\ &\quad \times \exp[-(W - \bar{w})^T (W - \bar{w}) / (2\sigma_u^2)] \\ &\propto L(l_{ij}|Y_{ij}, \bar{w}_{ij}, Z_{ij}, S_i, \beta). \end{aligned}$$

It is then clear that:

$$\begin{aligned}
 & \int L(l, W|Y, x, Z, S, \beta, \sigma_u^2) dx \\
 &= \prod_{i=1}^N \prod_{j=1}^{M+1} \int L(l_{ij}, W_{ij}|Y_{ij}, x_{ij}, Z_{ij}, S_i, \beta, \sigma_u^2) dx_{ij} \\
 &\approx \prod_{i=1}^N \prod_{j=1}^{M+1} \left\{ L(l_{ij}|Y_{ij}, \bar{w}_{ij}, Z_{ij}, S_i, \beta) \left[K_{ij}(2\pi\sigma_u^2)^{K_{ij}-1} \right]^{-1/2} \right. \\
 &\quad \left. \times \exp[-(W_{ij} - \bar{w}_{ij})^T (W_{ij} - \bar{w}_{ij}) / (2\sigma_u^2)] \right\} \\
 &\propto \prod_{i=1}^N \prod_{j=1}^{M+1} L(l_{ij}|Y_{ij}, \bar{w}_{ij}, Z_{ij}, \beta) \\
 &= L(l|Y, \bar{w}, Z, \beta).
 \end{aligned}$$

It follows from a similar argument that,

$$\begin{aligned}
 & \int L(l, W|Y, x, Z, S, \beta, \sigma_u^2) N(x|\mu_x, \sigma_x^2) dx \\
 &\approx L\left(l|Y, x = \frac{K\bar{w}\sigma_x^2 + \mu_x\sigma_u^2}{K\sigma_x^2 + \sigma_u^2}, Z, S, \beta\right).
 \end{aligned}$$

2.2 First order Laplace for E2 approach to E-step

Consider now where the goal is to use first order Laplace approximation to find:

$$\begin{aligned}
 E\{\log[L(Y|x, Z, S, \beta)]\} &= \int \log[L(Y|x, Z, S, \beta)] N(W; x, \sigma_u^2) N(x; \mu_x, \sigma_x^2) dx, \\
 &\approx \log[L(Y|\tilde{x}, Z, S, \beta)], \\
 \tilde{x}_{ij} &= \frac{K_{ij}\bar{w}_{ij}\sigma_x^2 + \mu_x\sigma_u^2}{K_{ij}\sigma_x^2 + \sigma_u^2}.
 \end{aligned}$$

We can rewrite the integration as follows:

$$\begin{aligned}
 & \int \log[L(Y_{ij}|x_{ij}, Z_{ij}, S_i, \beta)] N(W_{ij}; x_{ij}, \sigma_u^2) N(x_{ij}; \mu_x, \sigma_x^2) dx_{ij} \\
 &= \int \log[L(Y_{ij}|x_{ij}, Z_{ij}, S_i, \beta)] (2\pi\sigma_u^2)^{-K_{ij}/2} \\
 &\quad \times \exp\left[-0.5 \sum_{k=1}^{K_{ij}} (w_{ijk} - x_{ij})^2 / \sigma_u^2\right]
 \end{aligned}$$

$$\begin{aligned} & \times (2\pi\sigma_x^2)^{-1/2} \exp\left[-0.5(x_{ij} - \mu_x)^2/\sigma_x^2\right] dx_{ij} \\ & \approx \log[L(Y|\tilde{x}, Z, S, \beta)], \\ & = \int A(x_{ij}) \exp[-h(x_{ij})] dx_{ij}, \end{aligned}$$

where:

$$\begin{aligned} A(x_{ij}) &= \log[L(Y_{ij}|x_{ij}, Z_{ij}, S_i, \beta)](2\pi\sigma_u^2)^{-K_{ij}/2}(2\pi\sigma_x^2)^{-1/2}, \quad \text{and} \\ h(x_{ij}) &= \frac{\sum_{k=1}^{K_{ij}} (w_{ijk} - x_{ij})^2 + (x_{ij} - \mu_x)^2}{2\sigma_x^2\sigma_x^2}. \end{aligned}$$

It should be clear since $h(x_{ij})$ is the sum of two quadratic functions of x_{ij} , that the unique maximum of $h(\cdot)$ is the Bayes estimator $\tilde{x}_{ij} = \frac{K_{ij}\bar{w}_{ij}\sigma_x^2 + \mu_x\sigma_u^2}{K_{ij}\sigma_x^2 + \sigma_u^2}$. Also the second derivative $h''(x_{ij}) = \frac{1}{\sigma_x^2\sigma_x^2}$. It follows then that:

$$\begin{aligned} \int A(x_{ij}) \exp[-h(x_{ij})] dx_{ij} &\approx A(\tilde{x}_{ij}) \exp[-h(\tilde{x}_{ij})] \sqrt{\frac{2\pi\sigma_u^2\sigma_x^2}{K_{ij} + 1}}, \\ &\propto \log[L(Y_{ij}|\tilde{x}_{ij}, Z_{ij}, S_i, \beta)]. \end{aligned}$$

It follows from a similar argument as in Appendix 2.1 that:

$$E\{\log[L(Y|x, Z, S, \beta)]\} \approx \log[L(Y|\tilde{x}, Z, S, \beta)].$$

References

- Agresti A (2002) Categorical data analysis, 2nd edn. Wiley series in probability and statistics. Wiley, Hoboken
- Albert J, Chib S (1993) Bayesian-analysis of binary and polytochotomous response data. *J Am Stat Assoc* 88(422):669–679. <https://doi.org/10.2307/2290350>
- Bartlett J, Keogh R (2018) Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration. *Stat Methods Med Res* 27:1695–1708
- Berry SM, Carroll RJ, Ruppert D (2002) Bayesian smoothing and regression splines for measurement error problems. *J Am Stat Assoc* 97(457):160–169. <https://doi.org/10.1198/016214502753479301>
- Buzas JS, Stefanski LA (1996) A note on corrected-score estimation. *Stat Probab Lett* 28(1):1–8. [https://doi.org/10.1016/0167-7152\(95\)00074-7](https://doi.org/10.1016/0167-7152(95)00074-7)
- Camilli G (1994) Origin of the scaling constant $d = 1.7$, in item response theory. *J Educ Behav Stat* 19(3):293–295
- Carroll R, Roeder K, Wasserman L (1999) Flexible parametric measurement error models. *Biometrics* 55(1):44–54. <https://doi.org/10.1111/j.0006-341X.1999.00044.x>
- Carroll R, Ruppert D, Tosteson T, Crainiceanu C, Karagas M (2004) Nonparametric regression and instrumental variables. *J Am Stat Assoc* 99:736–750
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective, 2nd edn. Monographs on statistics and applied probability. Chapman and Hall/CRC, Boca Raton
- Crainiceanu C, Ruppert D, Wand MP (2005) Bayesian analysis for penalized spline regression using winbugs. *J Stat Softw* 14(1):1–14

- Eaton JW et al (2008) GNU Octave 3.0.5. www.gnu.org/software/octave/
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6(6):721–741
- Guolo A (2008) A flexible approach to measurement error correction in case–control studies. *Biometrics* 64(4):1207–1214. <https://doi.org/10.1111/j.1541-0420.2008.00999.x>
- Guolo A, Brazzale AR (2008) A simulation-based comparison of techniques to correct for measurement error in matched case–control studies. *Stat Med* 27(19):3755–3775. <https://doi.org/10.1002/sim.3282>
- Gustafson P (2003) Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. CRC Press, Boca Raton
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109. <https://doi.org/10.2307/2334940>
- Hosmer DW Jr, Lemeshow S (2000) Applied logistic regression, 2nd edn. Wiley series in probability and statistics. Wiley, Hoboken
- Huang Y, Wang C (2000) Cox regression with accurate covariates unascertainable: a nonparametric-correction approach. *J Am Stat Assoc* 95(452):1209–1219
- Huang Y, Wang C (2001) Consistent functional methods for logistic regression with errors in covariates. *J Am Stat Assoc* 96(456):1469–1482
- MATLAB (2012) Version 7.14.0.739 (R2012a). The MathWorks Inc., Natick
- McShane L, Midthune D, Dorgan J, Freedman L, Carroll R (2001) Covariate measurement error adjustment for matched case–control studies. *Biometrics* 57(1):62–73. <https://doi.org/10.1111/j.0006-341X.2001.00062.x>
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092. <https://doi.org/10.1063/1.1699114>
- Parker PA, Vining GG, Wilson SR, Szarka JL III, Johnson NG (2010) The prediction properties of classical and inverse regression for the simple linear calibration problem. *J Qual Technol* 42(4):332–347
- Peleg AY, Husain S, Qureshi ZA, Silveira FP, Sarumi M, Shutt KA, Kwak EJ, Paterson DL (2007) Risk factors, clinical characteristics, and outcome of nocardia infection in organ transplant recipients: a matched case–control study. *Clin Infect Dis* 44(10):1307–1314. <https://doi.org/10.1086/514340>
- Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric regression. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, New York
- Ryu D, Li E, Mallick B (2011) Bayesian nonparametric regression analysis of data with random effects covariates from longitudinal measurements. *Biometrics* 67:454–466
- Scott AJ, Wild CJ (1997) Fitting regression models to case–control data by maximum likelihood. *Biometrika* 84(1):57–71
- Shaby B, Wells M (2010) Exploring an adaptive metropolis algorithm. Technical report, Department of Statistical Science, Duke University
- Sinha S, Mukherjee B, Ghosh M, Mallick BK, Carroll RJ (2005) Semiparametric Bayesian analysis of matched case–control studies with missing exposure. *J Am Stat Assoc* 100(470):591–601
- Sinha S, Mallick B, Kipnis V, Carroll R (2010) Semiparametric Bayesian analysis of nutritional epidemiology data in the presence of measurement error. *Biometrics* 66:444–454
- Stefanski LA, Carroll RJ (1987) Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* 74(4):703–716. <https://doi.org/10.1093/biomet/74.4.703>
- Tester J, Rutherford G, Wald Z, Rutherford M (2004) A matched case–control study evaluating the effectiveness of speed humps in reducing child pedestrian injuries. *Am J Public Health* 94(4):646–650. <https://doi.org/10.2105/AJPH.94.4.646>
- Tierney L, Kadane J (1986) Accurate approximations for posterior moments and marginal densities. *J Am Stat Assoc* 81(393):82–86. <https://doi.org/10.2307/2287970>
- Whitney CG, Pilishvili T, Farley MM, Schaffner W, Craig AS, Lynfield R, Nyquist A-C, Gershman KA, Vazquez M, Bennett NM, Reingold A, Thomas A, Glode MP, Zell ER, Jorgensen JH, Beall B, Schuchat A (2006) Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: a matched case–control study. *Lancet* 368(9546):1495–1502. [https://doi.org/10.1016/S0140-6736\(06\)69637-2](https://doi.org/10.1016/S0140-6736(06)69637-2)
- Woodward M (2013) *Epidemiology: study design and data analysis*, 3rd edn. Chapman & Hall, Boca Raton