Bartlett, JW; Keogh, RH (2016) Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. Statistical methods in medical research. ISSN 0962-2802 DOI: https://doi.org/10.1177/0962280216667764

Downloaded from: http://researchonline.lshtm.ac.uk/2901283/

DOI: 10.1177/0962280216667764

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Article

Statistical Methods in Medical Research
0(0) 1–14

DOI: 10.1177/0962280216667764
smm.sagepub.com
SAGE

Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration

Jonathan W Bartlett[1] and Ruth H Keogh[2]

Abstract

Bayesian approaches for handling covariate measurement error are well established and yet arguably are still relatively little used by researchers. For some this is likely due to unfamiliarity or disagreement with the Bayesian inferential paradigm. For others a contributory factor is the inability of standard statistical packages to perform such Bayesian analyses. In this paper, we first give an overview of the Bayesian approach to handling covariate measurement error, and contrast it with regression calibration, arguably the most commonly adopted approach. We then argue 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. Next, we describe the closely related maximum likelihood and multiple imputation approaches and explain why we believe the Bayesian approach to generally be preferable. We then empirically compare the frequentist properties of regression calibration and the Bayesian approach through simulation studies. The flexibility of the Bayesian approach to handle both measurement error and missing data is then illustrated through an analysis of data from the Third National Health and Nutrition Examination Survey.

Keywords

Measurement error, Bayesian inference, regression calibration, multiple imputation

1 Introduction

Many epidemiological studies are affected by measurement error in one or more of the covariates of interest. It is well known that error in covariates results in biased estimates of true covariate(s)-outcome associations and in a loss of power to detect such associations.[1] In this paper, we focus on correcting for the effects of measurement error in continuous covariates in three models which are commonly used in epidemiological analysis; linear regression models for continuous outcomes, logistic regression models for binary outcomes, and Cox proportional hazards models for survival or time to event outcomes.

Throughout most of the paper, we will focus on the situation in which there is one main exposure of interest, which is subject to measurement error, and one or more other covariates to be adjusted for, which are assumed to be measured without error. The variable which is measured with error could equally be one of the confounders, and indeed the approaches we describe also extend to the more general case of multiple covariates measured with error. While exposure measurement error is commonly prioritized, measurement error in confounders is also a serious and highly prevalent issue, and causes estimates of exposure effects to only be partially adjusted for the poorly measured confounder. Error in continuous variables can take a number of forms. The most simple and most commonly assumed form is the classical measurement error model, under which the measured exposure is equal to the true exposure plus an independent random error term. Under this model, the measured exposure is an

[1]Statistical Innovation Group, AstraZeneca, Cambridge, UK
[2]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

Corresponding author:
Jonathan W Bartlett, Statistical Innovation Group, AstraZeneca, Riverside 2, Granta Park, Cambridge CB21 6GP, UK.
Email: jwb133@googlemail.com

unbiased measure of the true exposure. The error terms are assumed to have zero mean and, typically, constant variance.

To make corrections for the effects of covariate measurement error in regression models requires some information about the relationship between the true exposure and the measured exposure, i.e. regarding the parameters of a measurement error model. One way of gaining information about the error model is to use a validation study within the main study sample, in which the true exposure is observed alongside the measured exposure. It is often not feasible or even possible, however, to obtain a validation sample and a more common alternative is to obtain one or more replicate observations of the exposure for a subset of individuals within the main study sample. We refer to this as a replication study. In this paper, we focus on replication studies.

Many methods have been described for correcting for the effects of measurement error in regression models.[1] The most widely used correction method is regression calibration (RC), which is popular due to its simplicity and applicability in different types of regression models. In RC, the true exposure, which is unobserved in the main study sample, is replaced when fitting the outcome regression model by the expected value of the true exposure, conditional on the measured exposure and the other error-free covariates for each individual. RC gives consistent estimates of the true associations between the explanatory variables and the outcome in a linear regression model, and approximately consistent estimates in non-linear models, including logistic regression models[2,3] and Cox proportional hazards models.[4]

RC has some drawbacks, however. First, for non-linear models estimates can have moderately large biases even when the sample size is large, particularly if the effect size (odds ratio or hazard ratio) is large.[5] Second, RC does not automatically accommodate uncertainty in the parameters indexing the measurement process. Measures of uncertainty require use of approximate methods (the ‘delta method’ approach), bootstrapping methods, which are computationally intensive, or estimating equation methods, whose validity relies on asymptotic conditions and are complex to implement in practice. Third, extending the basic RC approach to more complex situations, such as when the outcome model is assumed to depend on non-linear functions of the true covariate,[6] or when the measurement error model is more complex (e.g. heteroscedastic error[7]), is not trivial.

The Bayesian approach has been often advocated as a natural route to accommodating sources of uncertainty, including measurement error, misclassification, and missing data. Early papers include those by Richardson and Gilks,[8,9] who described a Bayesian approach to handling measurement error. By taking a Bayesian approach to handle covariate measurement error, uncertainty in the parameters indexing the measurement process is automatically accommodated. Like the method of maximum likelihood (ML), the posterior distributions involved typically involve intractable likelihoods, but this difficulty is obviated by Markov Chain Monte Carlo (MCMC) methods, which are now implemented in a number of standard and Bayesian-specific packages. A further strength is that these software packages allow one to define and fit quite complex user defined Bayesian models, meaning that there is great flexibility in adapting the modelling assumptions to the situation at hand. Lastly, and in contrast to methods such as ML or multiple imputation (MI), whose inferences typically rely on various large sample assumptions (e.g. to handle nuisance parameters or in deriving simple imputation combination rules), Bayesian methods do not. In the setting of covariate measurement error, estimators which allow for the error typically have skewed sampling distributions, and this is automatically accommodated in a Bayesian approach, since the entire posterior distribution is simulated.

Despite excellent book length treatments of covariate measurement error methods,[1] including one specifically focusing on Bayesian methods,[10] in our view it nevertheless continues to be underused by the epidemiological and clinical research communities. This may be for a number of reasons, but principal among them may be the apparent need to move from a frequentist to a Bayesian inferential approach, and the fact that standard statistical packages have (with exceptions) not enabled such Bayesian models to be fitted. To the first of these reasons, as has been noted by others (e.g. Little[11]), Bayes procedures often have good frequentist properties, and indeed in small samples can have better frequentist properties than ML methods. As such, one may be able to use a Bayesian method without necessarily adopting the Bayesian inferential paradigm. To this end, we present simulation results to examine the frequentist properties of the Bayesian approach to covariate measurement error, using certain default priors. To the second reason, major steps forward have been made over the last 25 years in terms of accessible MCMC software, such that software and computational power are usually not a hinderance to using a Bayesian approach. Moreover, we make all of our code available online to facilitate increased use of the Bayesian approach.

In Section 2, we begin by describing the assumed setup and notation for the covariate measurement error problem. Next, in Section 3, we review the RC approach. In Section 4, we describe the Bayesian approach, both in terms of modelling choices and statistical properties, and its practical implementation. We contrast the Bayesian approach with ML and MI in Section 5. In Section 6, we evaluate the frequentist properties of RC and Bayesian analysis in a series of simulation studies of the most common outcome model types. In Section 7, we present results of illustrative analyses using data from the Third National Health and Nutrition Examination Survey (NHANES III). We conclude in Section 8 with a discussion.

2 Setup and notation

In this section, we describe the general setup used for the remainder of the paper.

2.1 Outcome model

We assume data are available for an i.i.d. sample of n individuals. For individual i, we let Y_i, X_i, and $\mathbf{Z}_i$, respectively, denote the outcome, true covariate which is subject to measurement error and error-free covariates. We consider three types of outcome models for Y_i (i) linear, (ii) logistic, and (iii) Cox proportional hazards regression. We assume that the outcome model includes only main effects of X_i and $\mathbf{Z}_i$. For a linear regression outcome model, we thus assume that

$$Y_i = \beta_0 + \beta_X X_i + \beta_Z^T \mathbf{Z}_i + \epsilon_i \tag{1}$$

where $\epsilon_i \sim N(0, \sigma^2)$ is an independent normally distributed residual error. For a logistic regression outcome model, we assume that

$$\text{logit}\{P(Y_i = 1 | X_i, \mathbf{Z}_i)\} = \beta_0 + \beta_X X_i + \beta_Z^T \mathbf{Z}_i \tag{2}$$

Lastly, in the case of a censored time to event outcome, the outcome $Y_i = (T_i, D_i)$ where T_i denotes the observed event or censoring time and D_i denotes the event indicator. We then assume a Cox proportional hazards outcome model, such that the hazard given X_i and $\mathbf{Z}_i$ is given by

$$h(t | X_i, \mathbf{Z}_i) = h_0(t) \exp(\beta_X X_i + \beta_Z^T \mathbf{Z}_i)$$

where $h_0(t)$ denotes the baseline hazard function. In the standard frequentist analysis based on the Cox proportional hazards model, the baseline hazard is left unspecified, and inferences about the hazard ratio parameters (β_X, β_Z) are made via a partial likelihood.[12]

2.2 Measurement error model

We assume that for each study individual, an error-prone measurement W_{i1} is available, rather than the covariate of interest X_i. We assume a classical error model

$$W_{i1} = X_i + U_{i1}$$

where $E(U_{i1} | X_i) = 0$. We also assume that the errors U_{i1} are independent of all other random variables. This implies that the error is non-differential with respect to the outcome Y_i.

To permit estimation of the measurement error variance, we assume the existence of an internal replication sub-study. This means that for a randomly selected group of individuals a second error-prone measurement $W_{i2} = X_i + U_{i2}$ is obtained, where the error U_{i2} is assumed independent of U_{i1}. We let $\mathbf{W}_i$ denote the vector of error-prone measurements on individual i, and let N_i denote its length, which is two for those individuals in the replication sub-study and one for those not. In the following, we specify further assumptions as required by RC and Bayesian methods.

3 Regression calibration

In the simplest version of RC, the outcome model is fitted as usual, with the unobserved X_i replaced by an estimate of $E(X_i | W_{i1}, \mathbf{Z}_i)$.[13] Typically, the latter conditional expectation is assumed to be linear in W_{i1} and $\mathbf{Z}_i$, and it can be estimated by linearly regressing W_{i2} on W_{i1} and $\mathbf{Z}_i$ in the individuals from the internal replication sub-study.

We note that this version of RC does *not* rely on an assumption that the two errors U_{i1} and U_{i2} have the same variance.

If one is willing to make additional assumptions, a somewhat more efficient version of RC can be used, in which X_i is replaced by an estimate of $E(X_i|\mathbf{W}_i, \mathbf{Z}_i)$, and the parameters involved in the latter are estimated using all study individuals. A common assumption is to assume that $X_i|\mathbf{Z}_i \sim N(\gamma_0 + \gamma_Z^T\mathbf{Z}_i, \sigma^2_{X|Z})$ and that the measurement errors U_{i1} and U_{i2} are normally distributed with mean zero and common variance σ^2_U. The parameters can be estimated by ML as a random-intercepts mixed model for the $\mathbf{W}_i$ conditional on $\mathbf{Z}_i$. It then follows from standard properties of the multivariate normal distribution that $X_i|\mathbf{W}_i, \mathbf{Z}_i$ is normally distributed, with

$$\begin{aligned} E(X_i|\mathbf{W}_i, \mathbf{Z}_i) &= \gamma_0 + \gamma_Z^T\mathbf{Z}_i + \frac{\sigma^2_{X|Z}}{\sigma^2_{X|Z} + \sigma^2_U/N_i}\left(\overline{W}_i - (\gamma_0 + \gamma_Z^T\mathbf{Z}_i)\right) \\ \mathrm{Var}(X_i|\mathbf{W}_i, \mathbf{Z}_i) &= \sigma^2_{X|Z}\left(1 - \frac{\sigma^2_{X|Z}}{\sigma^2_{X|Z} + \sigma^2_U/N_i}\right) \end{aligned} \tag{3}$$

where $\overline{W}_i$ denotes the mean of individual *i*'s N_i error-prone measurements.

As noted in the introduction, RC gives consistent parameter estimates in the case of a linear outcome model. For logistic and Cox outcome models, RC is approximately consistent. Armstrong first gave justification for RC in generalized linear models under the assumption that $\mathrm{Var}(X_i|\mathbf{W}_i)$ is 'small', using the delta method.[2] This condition will be satisfied when the measurement error variance is 'small'. Rosner et al.[3] later justified its use in logistic regression under the assumptions that the outcome is rare and that $X_i|\mathbf{W}_i$ is normal. Subsequently, Kuha[14] showed that RC could be justified as an approximate method for logistic regression provided that $\beta_X^2\mathrm{Var}(X_i|\mathbf{W}_i)$ is small, without the rare outcome assumption. This condition holds when β_X is small or the measurement error variance is small. For a Cox proportional hazards outcome model, RC can be justified when the event rate is low or the measurement error variance is small.[4,15]

For valid inferences, the estimation of the parameters involved in $E(X_i|\mathbf{W}_i, \mathbf{Z}_i)$ should be allowed for. One approach is to use bootstrapping. Alternatively, it is possible to construct sandwich variance estimators by stacking the estimating equations used in the two stages.[1] One drawback with this approach is that the resulting Wald type symmetric confidence intervals do not reflect the asymmetric sampling distribution of the RC estimator, which may lead to confidence interval coverage which deviates from the nominal level.

4 Bayesian approach

In this section, we describe the key elements of a Bayesian analysis of the covariate measurement error problem.

4.1 Model specification

First, we specify a joint parametric model for $(Y_i, X_i, W_{i1}, W_{i2}|\mathbf{Z}_i)$. We condition on the fully observed $\mathbf{Z}_i$, thereby avoiding the need to model its distribution. Assuming that the measurement error is non-differential with respect to both Y_i and $\mathbf{Z}_i$, this joint model can be decomposed as

$$f(Y_i|X_i, \mathbf{Z}_i, \beta, \eta)f(\mathbf{W}_i|X_i, \sigma^2_U)f(X_i|\mathbf{Z}_i, \gamma) \tag{4}$$

The first component is the outcome model, which contains regression parameters of primary interest $\beta = (\beta_0, \beta_X, \beta_Z)$ in the case of linear or logistic regression and $\beta = (\beta_X, \beta_Z)$ in the case of Cox regression, and possibly additional parameters η (e.g. a residual variance in the case of a linear regression outcome model). The second component is the measurement model, and as described previously the simplest assumption is that the error-prone measurements W_{ij} follow a classical error model, with independent normally distributed errors $U_{ij} \sim N(0, \sigma^2_U)$. The final component specifies a model for the unobserved covariate X_i, conditional on $\mathbf{Z}_i$, with a default choice being a normal linear regression model. We return later to questions of robustness and to model extensions to relax such distributional assumptions. In the case of a Cox proportional hazards outcome model, the outcome Y_i has two components (T_i, D_i), and the additional parameters, η, denote the baseline hazard function $H_0(t)$.

4.2 Prior specification

In the Bayesian approach, we must specify priors for the model parameters. The first 'Bayesian' analyses made use of flat or constant priors, based on the notion that these represent a priori ignorance regarding the value(s) of the model parameter(s).[16] The key issue with such priors is that while a flat prior expresses ignorance on one scale, a transformation of the parameter implies a non-flat prior on the transformed parameter. The latter half of the 20th century witnessed the growth of the subjective Bayesian approach, in which the analyst carefully choose the priors to represent their beliefs about the model parameters in advance of seeing the data. Arguably the majority of Bayesian analyses which are now performed by researchers make use of so called non-informative or reference priors.[17] Such priors do not (and cannot) represent total ignorance about the model parameters, but can be viewed as default priors that one might use when subjective prior information is either not available, or one does not want to use such information in the analysis. The intention of such priors is usually that they have minimal impact on inferences.

For the joint model in equation (4), prior independence is typically assumed for the parameters in the three sub-models. For the outcome model regression coefficients β and the coefficients in γ, a common default prior is a very diffuse normal prior centred at zero. For the variance parameters, the conjugate inverse Gamma distribution has traditionally been advocated. In the context of adjustment for covariate measurement error, Gustafson has proposed using a $Ga(0.5, 0.5)$ prior for the precision (reciprocal of variance) parameters.[10] This prior equates to the likelihood that would be obtained from one observation, with the best guess for the precision of one.

Bayesian analysis of the Cox model requires specification of a prior for the baseline cumulative hazard process $H_0(t)$ in addition to priors for the regression coefficients β and the other sub-model parameters. The prior distribution for the baseline cumulative hazard process $H_0(t)$ is assumed to be independent of the other priors, including that for β. Here, we use a Gamma process prior for $H_0(t)$ as described by Kalbfleisch[18] and Sinha et al.,[19] denoted $H_0(t) \sim \mathcal{GP}(cH_0^*, c)$, where $H_0^*(t)$ is a prior guess at the mean and c is a parameter which represents the confidence in that guess, with small values of c corresponding to a diffuse prior. We let $t_{(1)} < t_{(2)} < \cdots < t_{(n^*)}$ denote the ordered observed event times. Under the assumption that the hazard is degenerate at 0 except at the observed event times T_i where $D_i = 1$ it follows from the Gamma process prior for $H_0(t)$ that the increments in the cumulative baseline hazard from time $t_{(j)}$ to time $t_{(j+1)}$ $(j = 1, \ldots, n^* - 1)$ have independent Gamma distributions; $dH_0(t_{(j)}) \sim \text{Gamma}(c(H^*(t_{(j+1)}) - H^*(t_{(j)}), c)$. In the later application of this approach, we use $H^*(t_{(j+1)}) - H^*(t_{(j)} = r(t_{(j+1)} - t_{(j)})$ where r is a guess at the event rate per unit time. It can be shown[18,19] that under the Gamma process prior for the cumulative hazard the likelihood for $(\beta, H_0(t), c)$ tends to the partial likelihood in the limit as c tends to 0, and that it tends to the full likelihood with $H_0(t) = H^*$ as c tends to infinity.

4.3 Posterior inference and simulation

Given specification of the model and priors, Bayesian inference is then based on the posterior distributions of the model parameters. For the purposes of point estimation, the posterior mean is commonly used. To form a 95% credible interval for a particular parameter, we take the 2.5% and 97.5% centiles of the posterior distribution. An advantage of this in the present context of adjustment for covariate measurement error is that asymmetry in the posterior distribution, which typically occurs when adjusting for covariate measurement error, is automatically accounted for in credible intervals.

Except for very specific choices of model and prior, in general the posteriors are not available analytically. Instead, we can utilize MCMC methods to simulate draws from the posteriors distributions (see e.g. part III of Gelman et al.[20]). The most common approach is the method of Gibbs sampling, in which taking each parameter in turn, a new value is drawn from its full conditional distribution given all other quantities. Often these conditional distributions do not belong to standard parametric families, necessitating the use of more sophisticated sampling techniques (see Robert[17] and Gelman et al.[20] for further details). However, these are implemented in the software packages we describe in the following, such that the analyst need not generally concern themselves with the details.

4.4 Frequentist properties

Under certain regularity conditions, as the sample size tends to infinity, the choice of prior has no impact on the posterior distribution, since the latter is then dominated by the likelihood function. Consequently, Bayes estimators and uncertainty intervals enjoy the same large sample properties as maximum likelihood methods: the Bayes posterior mean estimator is consistent, asymptotically normal, and efficient.[20]. In reality of course, all samples are finite, and the choice of prior can sometimes have a material affect on inferences.

Importantly, however, in small samples or sparse data situations, Bayesian methods can have better frequentist properties than ML procedures, particularly if sensible priors are adopted.[21]

4.5 Software

The explosion of Bayesian data analyses being performed over the last few decades is largely thanks to both the MCMC methods developed and their implementation in accessible software. Chief among these is the WinBUGS software package, developed in the 1990s.[22] It allows the user to define, using a simple language syntax or graphical interface, the model and priors. MCMC methods are then automatically chosen by the package, depending on the specified model and priors. One can then run the MCMC sampler, and after a sufficient number of burn-in iterations, draws can be saved as draws from the respective posterior distributions. More recently, new packages have been developed, with developments in various directions. These include the OpenBUGS project (www.openbugs.net) and Stan (mc-stan.org). In the simulations described later, we make use of the Just Another Gibbs Sampler (JAGS) program,[23] whose model language is very similar to the BUGS language used by WinBUGS, and which can be easily called from R.

5 Maximum likelihood and multiple imputation

5.1 Maximum likelihood

ML estimation and inference is based on the likelihood function, but unlike the Bayesian approach, does not involve specification of prior distributions for parameters. ML methods enjoy many favourable frequentist properties – assuming correct model specification the ML estimator is consistent, asymptotically normal, and efficient. In the specific context of adjustment for covariate measurement error, a drawback of ML is that the likelihood function typically involves intractable integrals,[1] such that numerical methods such as numerical integration are required in order to obtain estimates. The same obstacle is overcome in the Bayesian approach through the use of MCMC methods. A further drawback of ML in the present context is that in small samples inference based on symmetric Wald based confidence intervals may perform poorly due to the lack of regularity of the likelihood function. Software to fit user defined models which allow for covariate measurement error is also somewhat limited. Lastly, the absence of prior distributions prevents the incorporation of external information which may sometimes be available regarding the measurement process.

5.2 Multiple imputation

MI has become an extremely popular approach for handling missing data and has also been advocated as an approach for handling measurement error, in which X_i is multiply imputed.[24–27] There is a very close connection between MI and 'direct' Bayesian inference. In its originally devised form, MI is based on repeatedly drawing imputations of missing values from their posterior distribution based on a Bayesian imputation model. The analysis model of interest is then fitted, typically using ML, to each of the complete datasets. The resulting estimates and standard errors are then pooled using rules developed by Rubin.[28] MI can most directly be viewed as an approximation to a full Bayesian analysis,[29] although its frequentist properties can of course also be evaluated.[30] From the Bayesian perspective, application of MI and Rubin's rules can be viewed as a particular route to performing a Bayesian analysis, in which one effectively assumes that the posterior distributions for the parameters are normally distributed.

As described by Carpenter and Kenward[29] in the context of missing data, there are a number of settings where use of MI may be preferable to a direct Bayesian analysis. However, in the context of covariate measurement error we argue that the advantages of a direct Bayesian approach far outweigh the disadvantages, relative to MI. First, when only replicate error-prone measurements are available, standard software for performing MI cannot be applied, since X_i is missing for all individuals. Second, standard parametric imputation models which might be used to impute X_i in general may not be compatible with the assumed outcome model.[31] This will in particular occur when the outcome model is itself non-linear, or the imprecisely measured covariate is assumed to have a non-linear effect on the outcome. Third, when allowance is made for covariate measurement error, as noted earlier, the posterior distributions for the outcome model parameters are typically skewed in small to moderate sample sizes, such that symmetric credible/confidence intervals constructed using Rubin's rules may perform poorly, either from a subjective Bayesian perspective or in a frequentist evaluation. Lastly, we note that software for performing Bayesian inference will typically also permit saving of the imputed values of X_i as a by-product, such that if the analyst really wants imputed datasets they can still be obtained.

6 Simulations

In this section, we present simulation results for the cases of a linear, logistic, and Cox proportional hazards outcome model, comparing the popular RC approach with the Bayesian approach. We adopt the standard frequentist type simulation setup, in which datasets are repeatedly generated using fixed population parameter values.

6.1 Linear regression

We first present simulation results for a simple linear regression outcome model. Datasets of size $n = 1000$ were simulated, with covariates X_i and Z_i drawn from a bivariate normal distribution with means 0, variances 1, and covariance 0.25. Continuous outcomes Y_i were generated from the linear regression model given in equation (1), with $\beta_0 = 0$ and $\beta_X = \beta_Z = 1$. The normally distributed residual variance σ^2 was chosen in order to given $R^2 = 0.1, 0.5, 0.9$. Each individual had an error-prone measurement $W_{i1} = X_i + U_{i1}$, with $U_{i1} \sim N(0, \sigma_U^2)$. A random subset of 10% of individuals had a second error-prone measurement, with the same error variance σ_U^2. This variance was chosen such that the unconditional reliability $\rho = \sigma_X^2/(\sigma_X^2 + \sigma_U^2)$ took values $0.5, 0.7, 0.9$, corresponding to low, moderate, and high reliability.

We first estimated the outcome model parameters using RC, by fitting a random-intercepts model for the error-prone measurements, with Z_i entering as a fixed effect, as described in Section 3. Next, we fitted a Bayesian model, calling JAGS from R using the rjags package. We adopted non-informative priors for all model parameters, following those proposed by Gustafson.[10] Specifically, we assumed independent normal priors for $\beta_0, \beta_X, \beta_Z, \gamma_0, \gamma_Z$, with mean 0 and variance 10,000, and inverse gamma $IG(0.5, 0.5)$ priors for each of the variance parameters. As discussed by Gustafson, the latter prior can be thought of as being equivalent to a best guess for the variance of one, coming from a single observation. We ran five parallel chains with 1000 burn-in iterations and 5000 main iterations. If the Rubin–Gelman convergence statistic Rhat was greater than 1.05 for any of $\beta_0, \beta_X, \beta_Z$, we extended the chains until this was met.

Table 1 shows the simulation results, with 1000 simulations per scenario. For Bayes, convergence to stationarity (as defined previously) was achieved in all simulations, although additional iterations were sometimes required to achieve this, particularly when reliability was 0.5. RC had slight upward bias for β_X when the reliability of the error-prone measurements was 0.5, and was unbiased for the higher reliability values. The Bayes mean estimator was upwardly biased for reliability equal to 0.5 and $R^2 = 0.1$ and $R^2 = 0.5$. Inspection of the estimates showed that the sampling distribution of the Bayes mean estimator had greater skew than the RC estimator, with the larger estimated values inducing the upward bias. For reliability of 0.5 and 0.7, and $R^2 = 0.9$, the Bayes estimators had lower empirical standard deviation (SD) than the RC estimator and consequently had lower mean squared error in these scenarios. For the other scenarios, the RC and Bayes estimators performed similarly. Lastly, the 95% Bayesian credible intervals had frequentist coverage close to 95%.

Table 1. Linear regression results.

Reliability	R^2	RC mean (SD)	Bayes mean (SD)	Bayes CI
0.5	0.1	1.03 (0.28)	1.17 (0.39)	0.94
0.5	0.5	1.03 (0.19)	1.15 (0.27)	0.92
0.5	0.9	1.03 (0.17)	0.98 (0.07)	0.98
0.7	0.1	1.01 (0.20)	1.05 (0.22)	0.95
0.7	0.5	1.01 (0.09)	1.04 (0.10)	0.94
0.7	0.9	1.00 (0.07)	1.01 (0.05)	0.97
0.9	0.1	1.00 (0.16)	1.02 (0.16)	0.95
0.9	0.5	1.00 (0.06)	1.01 (0.06)	0.94
0.9	0.9	1.00 (0.03)	1.01 (0.03)	0.94

1000 simulations per scenario. Empirical means and SDs for estimates of β_X, and coverage of 95% Bayesian credible intervals. RC: regression calibration; SD: standard deviation.

We emphasize that the performance of the Bayesian estimator depends on the choice of priors. In particular, the use of more informative priors for the measurement error variance and the variance for $X|Z$ could be used to reduce the bias and variability of the Bayesian estimators. To investigate this, we performed a further set of simulations, performing the Bayesian analysis with three different sets of priors. The first (Bayes 1) used the same priors as described previously. The second (Bayes 2) used the same priors except that the prior for $\sigma^2_{X|Z}$ was replaced by a $Beta(3, 1)$ prior for the conditional (on Z) reliability $\rho_{|Z} = \sigma^2_{X|Z}/(\sigma^2_{X|Z} + \sigma^2_U)$. This prior corresponds to assuming that the (conditional) reliability is unlikely to be small (≈ 0.015 chance of being less than 0.25), and it can be expected to stabilize estimates of this parameter (and therefore also the outcome model parameter estimates). The third (Bayes 3) instead used a $\text{Beta}(10, 10(1 - \rho^*_{|Z})/\rho^*_{|Z})$ prior for $\rho_{|Z}$, where $\rho^*_{|Z}$ denotes the true value of the parameter for each scenario. The mean of this prior is $\rho^*_{|Z}$, and so this corresponds to using a prior that is centred at the correct value of this parameter. In order to allow the different priors to have a larger influence, the sample size was reduced to 250 subjects, and only 25 subjects had a second error-prone replicate measurement W_{i2}.

Table 2 shows the results. When the (unconditional) reliability was 0.5, RC showed some upward bias. This is due to occasionally the estimated conditional reliability being very small, leading to a large estimate of β_X. Bayes 1 performed worse than RC with this reliability, except for $R^2 = 0.9$, where Bayes 1 had little bias and was much less variable. This can be attributed to the fact that the Bayes approach essentially imputes the unobserved X using all other variables, and when R^2 is large, the outcome Y enables X to be imputed with little variability. Bayes 2 had similar bias to RC when the reliability was 0.5 and R^2 was 0.1 or 0.5, but had much lower variability. As we would expect Bayes 3 had even lower bias than Bayes 1 and Bayes 2 (as well as RC), and was less variable. Results for reliability equal to 0.7 were qualitatively similar, again demonstrating that the use of stronger priors reduced bias and variability. For reliability of 0.9, the three Bayes estimators performed very similarly, but had slight upward bias (in contrast to RC, which had little bias).

6.2 Logistic regression

Next, we performed simulations with a logistic regression outcome model. The covariates X_i and Z_i, and error-prone measurements were generated as described previously. The binary outcome Y_i was then generated according to a logistic regression model (2). The intercept β_0 was chosen so that $P(Y_i = 1) = 0.2$ approximately. We performed simulations with log odds ratios $\beta_X = 0.1, 0.5, 2$, representing small, moderate, and large effects of X_i. As before, we set $\beta_Z = \beta_X$.

RC was implemented as described previously. For the Bayesian approach, we again used independent normal priors for each of β_0, β_X, and β_Z. For β_0 we used, as before, the non-informative prior $\beta_0 \sim N(0, 10000)$. For β_X and β_Z, we adopted the $N(0, 1.38)$ prior suggested by Hamra et al.[32] As described by Hamra et al., the use of such a mildly informative prior can help in terms of stabilizing estimates. This prior corresponds to assuming, a priori, that we are 95% sure that the odds ratios $\exp(\beta_X)$ and $\exp(\beta_Z)$ lie between 0.1 and 10, an assumption that is

Table 2. Linear regression results with small sample size.

Reliability	R^2	RC mean (SD)	Bayes 1 mean (SD)	Bayes 1 CI	Bayes 2 mean (SD)	Bayes 2 CI	Bayes 3 mean (SD)	Bayes 3 CI
0.5	0.1	1.20 (1.33)	2.06 (1.58)	0.93	1.23 (0.73)	0.97	1.13 (0.51)	0.97
0.5	0.5	1.20 (1.36)	1.62 (0.51)	0.90	1.21 (0.38)	0.96	1.15 (0.24)	0.97
0.5	0.9	1.21 (1.40)	0.95 (0.10)	0.98	0.91 (0.10)	0.96	0.96 (0.08)	1.00
0.7	0.1	1.03 (0.41)	1.30 (0.63)	0.96	1.14 (0.46)	0.97	1.08 (0.41)	0.97
0.7	0.5	1.02 (0.20)	1.27 (0.35)	0.92	1.15 (0.26)	0.95	1.08 (0.18)	0.96
0.7	0.9	1.02 (0.17)	1.00 (0.08)	0.98	0.99 (0.08)	0.98	1.00 (0.07)	0.99
0.9	0.1	1.01 (0.33)	1.08 (0.36)	0.96	1.08 (0.35)	0.96	1.05 (0.34)	0.96
0.9	0.5	1.00 (0.12)	1.07 (0.13)	0.94	1.07 (0.13)	0.94	1.05 (0.12)	0.95
0.9	0.9	1.00 (0.05)	1.04 (0.05)	0.93	1.04 (0.05)	0.92	1.03 (0.05)	0.94

1000 simulations per scenario. Empirical means and SDs for estimates of β_X, and coverage of 95% Bayesian credible intervals. RC: regression calibration; SD: standard deviation.

Bayes 1: $IG(0.5,0.5)$ priors for σ^2_U and $\sigma^2_{X|Z}$.

Bayes 2: $IG(0.5,0.5)$ prior for σ^2_U and $Beta(3, 1)$ prior for $\rho_{|Z}$.

Bayes 2 $IG(0.5,0.5)$ prior for σ^2_U and $Beta(10,10(1 - \rho^*_{|Z})/\rho^*_{|Z})$ prior for $\rho_{|Z}$.

Table 3. Logistic regression outcome model simulation results, with 1000 simulations per scenario.

Reliability	β_X	RC mean (SD)	Bayes mean (SD)	Bayes CI
0.5	0.1	0.10 (0.12)	0.12 (0.14)	0.94
0.5	0.5	0.51 (0.15)	0.58 (0.19)	0.94
0.5	2	1.64 (0.31)	1.94 (0.32)	0.97
0.7	0.1	0.10 (0.10)	0.11 (0.10)	0.95
0.7	0.5	0.50 (0.11)	0.52 (0.12)	0.94
0.7	2	1.74 (0.20)	2.01 (0.27)	0.97
0.9	0.1	0.10 (0.09)	0.10 (0.09)	0.94
0.9	0.5	0.50 (0.10)	0.51 (0.10)	0.95
0.9	2	1.91 (0.17)	2.01 (0.20)	0.96

Monte-Carlo means and SDs for estimates of β_X from regression calibration (RC) and Bayes, and empirical coverage of 95% Bayesian credible intervals. SD: standard deviation.

arguably generally reasonable in most epidemiology studies (provided the predictor has been suitably standardized). For the γ parameters and the variance parameters, we assumed the same priors as in the case of the first set of linear regression simulations.

Table 3 shows the results of the simulations. For $\beta_X = 0.1$, RC and Bayes performed very similarly, both being essentially unbiased and having similar empirical SD. For $\beta_X = 0.5$ and reliability of 0.5, while RC is unbiased, Bayes showed some upward bias and was more variable than RC. For reliability ratios of 0.7 and 0.9, the performance was similar for both. For $\beta_X = 2$ and reliability of 0.5, RC showed downward bias, consistent with the known properties of RC for logistic regression, in that bias is larger for large covariate effects. In contrast, Bayes showed only a slight downward bias. The bias of RC reduced, again as expected, as the reliability was increased to 0.7 and then 0.9, although some downward bias remained even for the latter case. In contrast, Bayes estimates were essentially unbiased. Lastly, the Bayesian 95% credible intervals had approximately 95% coverage across all scenarios.

6.3 Cox proportional hazards regression

Lastly, we performed simulations for time-to-event data based on a Cox proportional hazard model. The covariates X_i and Z_i were generated as described for linear regression. Event times T_i were generated according to the Weibull hazard model $h(t|X_i, Z_i) = \kappa\lambda t^{\kappa-1}e^{\beta_X X_i+\beta_Z Z_i}$. We used $\kappa = 2$, and λ was chosen so that approximately 10% of individuals had an event time before the end of follow-up which was fixed at time 10 (e.g. 10 years); the remaining individuals were censored at time 10 ($D_i = 0$). We performed simulations with log hazard ratios $\beta_X = 0.1, 0.5, 2$ and as before we set $\beta_Z = \beta_X$.

RC was performed as described previously. For the Bayesian approach, we used independent normal priors for β_X and β_Z, and we chose the $N(0, 1.38)$ prior as was used in the logistic regression simulations, corresponding here to an assumption that we are 95% sure that the hazard ratios $\exp(\beta_X)$ and $\exp(\beta_Z)$ lie between 0.1 and 10. For the γ parameters and the variance parameters, we assumed the same priors as in the case of linear and logistic regression. As outlined in Section 4.2, we assumed a process prior for the baseline cumulative hazard which implies Gamma priors for the increments in the hazards; $dH_0(t_{(j)}) \sim \text{Gamma}(c(H^*(t_{(j+1)}) - H^*(t_{(j)})), c)$. We used $c = 0.001$, representing low confidence in the prior mean of the Gamma Process, $H_0^*(t)$. We used $H^*(t_{(j+1)}) - H^*(t_{(j)}) = r(t_{(j+1)} - t_{(j)})$ with $r = 0.01$, since the data were simulated so that 10% of individuals have the event during 10 time units of follow-up. The analysis requires the data to be specified using a counting process format, and this is illustrated in example code given online. Due to the higher computational burden of fitting the Cox model, only three parallel chains were used, and 100 (rather than 1000) simulations were performed for each scenario.

Table 4 shows the simulation results. For $\beta_X = 0.1$ and $\beta_X = 0.5$, RC and Bayes performed very similarly across all three reliability values. For $\beta_X = 2$ and reliability of 0.5, RC showed bias toward the null, in line with previous simulation evidence.[33] This bias was reduced as the reliability increased, although there was still downward bias with reliability of 0.9. In contrast, the Bayes estimator was much less biased. The Bayesian credible intervals had approximately 95% coverage across all nine scenarios. To explore the impact of an increased proportion of individuals failing, an additional simulation study was conducted in which 50% of individuals were observed to

Table 4. Cox regression outcome model simulation results, with 100 simulations per scenario.

Reliability	β_X	RC mean (SD)	Bayes mean (SD)	Bayes CI
0.5	0.1	0.10 (0.09)	0.10 (0.09)	0.98
0.5	0.5	0.49 (0.11)	0.48 (0.11)	0.94
0.5	2	1.49 (0.15)	1.92 (0.20)	0.92
0.7	0.1	0.11 (0.09)	0.11 (0.09)	0.98
0.7	0.5	0.49 (0.11)	0.48 (0.11)	0.93
0.7	2	1.67 (0.16)	1.98 (0.18)	0.97
0.9	0.1	0.11 (0.10)	0.11 (0.10)	0.96
0.9	0.5	0.51 (0.10)	0.50 (0.10)	0.96
0.9	2	1.84 (0.15)	1.96 (0.15)	0.95

Monte-Carlo means and SDs for estimates of β_X from regression calibration (RC) and Bayes, and empirical coverage of 95% Bayesian credible intervals. SD: standard deviation.

fail (and with a smaller sample size of $n = 100$, each of whom had two error-prone replicates). The results (not shown) were qualitatively very similar to those in Table 4.

7 Illustrative example

To illustrate the potential flexibility and advantages of the Bayesian approach, we consider data from the NHANES III. NHANES III was a survey conducted in the US between 1988 and 1994 in 33,994 individuals aged two months and older. We consider a model relating known risk factors for cardiovascular disease (CVD) measured at the original survey to subsequent hazard of CVD. Mortality status at the end of 2011 is available through linkage to the US National Death Index, with cause of death classified using the ICD-10 system.

Here, we consider illustrative analyses of the data available on those aged 60 years and above at the time of the original survey. Fitting Cox models to large datasets is very slow using JAGS, particularly for large datasets. We therefore considered inference for a Weibull regression model for hazard for death due to CVD, with age, sex, smoking status, diabetes status, and systolic blood pressure (SBP) at the time of the survey as covariates

$$h(t) = rt^{r-1}\exp(\beta_0 + \beta_1 \text{sbp}_i + \beta_2 \text{sex}_i + \beta_3 \text{age}_i + \beta_4 \text{smoker}_i + \beta_5 \text{diabetes}_i)$$

where r is a shape parameter, $\beta_1, \ldots, \beta_5$ are log hazard ratios, and t is time since the original survey for each individual. We take the first SBP measurement taken at the original survey, sbp_{i1}, to be an error-prone measurement of each individual's underlying SBP, subject to classical error. An approximate 5% subset of individuals was (non-randomly) selected to participate in a second examination, during which SBP was again measured. This second exam took place on average 17.5 days after the first exam. We assume this second measurement of SBP, sbp_{i2}, to be an independent error-prone measurement of each individual's underlying SBP.

After deleting seven individuals who were missing diabetes status, data were available on 6519 individuals. Of these, by the end of 2011 1469 (22.5%) had died due to CVD, 3641 had died from other causes, and 1409 were still alive. In the Weibull model for hazard for CVD, we treat deaths from causes other than CVD as censorings; 5033 (77.2%) had an SBP measurement available from the first examination. An SBP measurement at the second examination was available in 401 (6.2%) of individuals. Unfortunately, smoking status was only recorded in 3433 (52.3%) of individuals. The analysis thus required handling of both the measurement error in the SBP measurements and the substantial missingness in the smoking and SBP variables.

7.1 Naive analyses

We first fitted the Weibull regression using sbp_{i1} (ignoring measurement error) to the 2667 complete cases, whose estimates are shown in the first column of Table 5. Strong evidence was found for independent associations between each of the covariates and hazard of death due to CVD, with associations in the expected directions. To check that a Weibull assumption was appropriate, we additionally fitted a Cox proportional hazards model with the same covariates. The estimates of the log hazard ratios between the two models were very similar, suggesting a Weibull assumption is reasonable here. Secondly, we performed a global test of the proportional hazards assumption using the Schoenfeld residuals following fitting the Cox model, which gave $p = 0.08$, indicating

Table 5. Log hazard ratios estimates and 95% confidence/credible intervals for the NHANES III data.

Covariate	Naive CCA	Naive Bayes	RC CCA	Bayes adj. CCA	Bayes adj. full
SBP (per 20 mmHg)	0.085 (0.014, 0.157)	0.086 (0.015, 0.160)	0.115 (0.014, 0.221)	0.114 (0.017, 0.211)	0.122 (0.059, 0.186)
Male	0.49 (0.30, 0.67)	0.49 (0.32, 0.67)	0.49 (0.32, 0.68)	0.49 (0.31, 0.69)	0.46 (0.36, 0.57)
Age (per 10 years)	0.88 (0.77, 0.99)	0.87 (0.76, 0.99)	0.87 (0.76, 0.99)	0.87 (0.75, 0.98)	1.01 (0.94, 1.09)
Smoker	0.26 (0.07, 0.46)	0.25 (0.06, 0.45)	0.26 (0.07, 0.45)	0.26 (0.06, 0.46)	0.24 (0.07, 0.41)
Diabetes	0.50 (0.29, 0.72)	0.50 (0.28, 0.72)	0.50 (0.28, 0.72)	0.50 (0.27, 0.71)	0.68 (0.55, 0.81)

CCA: complete case analysis performed using 2667 individuals, full analysis performed using 6519 individuals; RC: regression calibration; naïve: ignoring measurement error; adj.: adjusting for measurement error in SBP; SBP: systolic blood pressure.

no evidence to reject an assumption of proportional hazards. Through fitting a logistic regression model for the missingness indicator of the smoking variable, we found evidence that smoking was more likely to be missing for females, older individuals, diabetics, and those individuals with longer follow-up times. The latter finding suggests that the complete case analysis (CCA) may be biased.

Next, we fitted the same Weibull model to the complete cases, again ignoring measurement error, using the Bayesian approach. We assumed an exponential prior for the shape parameter r with parameter 0.001. Rather than placing a prior on the log hazard ratios, we placed independent $N(0, 10^6)$ priors on $-\beta_k/r$, since this leads to improved MCMC mixing.[34] Five independent chains were used, with 5000 burn-in iterations and 5000 main sample iterations. The estimates are 95% credible intervals are shown in Table 5. In line with theory, due to the large sample size, the Bayesian estimates and intervals were almost identical to those from ML.

7.2 Regression calibration

We then applied RC to the complete cases. To do this, we fitted a linear mixed model to the available SBP measurements, with a random effect for individual and fixed effects of sex, age, smoking, and diabetes. We also included a fixed effect to allow for a systematic shift in mean between the first and second exams. The resulting predicted true SBP values at exam one were then used as a covariate to fit the Weibull regression model. We used 2000 non-parametric bootstrap samples to obtain percentile 95% confidence intervals for the estimates, in order to take into account the two stage estimation process. Based on the mixed model fit to the complete cases, the estimated reliability (conditional on the error-free covariates) was 0.75. Adjusting for measurement error using RC led to the estimated log hazard ratio for SBP increasing, from 0.085 to 0.115, as expected by approximately 4/3 (1 divided by the reliability 0.75). Estimates for the other covariates did not materially change.

In order to apply RC to the full dataset, we contemplated use of its use in combination with MI. This is problematic, however. First, one could use MI to impute the missing smoking, sbp_{i1} and sbp_{i2} values. For example, one could apply the full conditional specification MI approach, imputing the smoking variable using logistic regression and the SBP variables using linear regression models. In these models, one must include the error-free covariates, plus the outcome. In the case of a time to event outcome modelled using a proportional hazards model, an approximately compatible imputation model for covariates includes the event indicator and an estimate of the cumulative hazard function.[35] Having generated the imputed datasets, RC could then be applied to each imputed datasets. However, in order to apply Rubin's rules, one requires valid within imputation estimates. As described in Section 3, for RC these can only be obtained by bootstrapping or by programming large sample theory estimating equation variance estimators. While both could in principle be programmed, they are not entirely straightforward to implement, and so we do not pursue MI in combination with RC.

7.3 Bayesian analyses adjusting for covariate measurement error

Next, we modified the Bayesian CCA to accommodate measurement error. We assumed that each individual's true underlying SBP around the time at which the first measurement was obtained, sbp_i, was normally distributed conditional on smoking, sex, age, and diabetes, with $N(0, 10^4)$ priors on the regression coefficients and a $Ga(0.5, 0.5)$ prior on the precision parameter. For the first SBP measurement, sbp_{i1}, we assumed

$$\text{sbp}_{i1} = \text{sbp}_i + U_{i1}$$

with $U_{i1} \sim N(0, \sigma_U^2)$. For the second SBP measurement, sbp_{i2}, we assumed

$$\text{sbp}_{i2} = \nu + \text{sbp}_i + U_{i2}$$

where ν is a parameter to allow for a systematic shift in mean between the two exams, and $U_{i2} \sim N(0, \sigma_U^2)$. The errors U_{i1} and U_{i2} are assumed to be independent. For σ_U^{-2} a $Ga(0.5, 0.5)$ prior was assumed, and for ν a $N(0, 10^4)$ prior was assumed. The posterior mean and credible intervals are shown in Table 5, under 'Bayes adj. CCA'. The results were very similar to those based on RC CCA, except that the credible interval for SBP was slightly narrower.

A strength of the Bayesian approach is its flexibility to simultaneously handle missing data and measurement error. To accommodate missingness in the smoking and SBP variables under a missing at random assumption, we assumed a model for the distribution of smoking, conditional on the fully observed error-free covariates sex, age, and diabetes. In our analysis, we assumed a logistic model for this conditional distribution

$$\text{logit}\{P(\text{smoker}_i = 1)\} = \alpha_0 + \alpha_1 \text{sex}_i + \alpha_2 \text{age}_i + \alpha_3 \text{diabetes}_i$$

with independent mean zero normal priors for the regression coefficients, each with variance 10,000. A major advantage of the Bayesian approach here is that the missing smoking and underlying SBP values are imputed by the Gibbs sampler, using the conditional distributions implied by a single well-specified joint model for the data. The posterior means were somewhat different to the RC and Bayes complete case analyses, and as expected the credible intervals were narrower, due to the inclusion of observed data from 3852 individuals. The changes in coefficient estimates may be indicative of bias in the complete case analyses.

8 Discussion

In this paper, we have empirically compared the frequentist properties of RC and Bayesian approaches to handling covariate measurement error. Our simulations demonstrate that for what might be considered a fairly typical epidemiological study setup, the methods often perform very similarly. As such, we believe that the Bayesian approach for measurement error adjustment may be as useful for the frequentist as for the Bayesian statistician. When the reliability of error-prone measurements was low, the Bayes estimator performed somewhat worse than RC. However, for larger effect sizes, RC was biased for logistic and Cox regression, while the Bayes estimator showed much less bias. A critical point to bear in mind is that there are infinitely many Bayes estimators, corresponding to the different choices of prior distributions – use of different priors could lead to, depending on the true data generating mechanism, better or worse performance. While some analysts dislike the Bayesian approach because of the requirement to specify priors, they give the analyst the opportunity to exploit external information about model parameters, potentially leading to more precise estimates.

We have highlighted the fact that Bayesian estimators enjoy the same large sample frequentist properties as the method of ML, and also described the relationships between these approaches and the popular MI approach. Software for MI cannot be directly applied to handle covariate measurement error when replication data are available. Moreover, even when validation data are available, the covariate imputation models included in MI implementations may not be compatible with the analyst's outcome model.[31] A further strength of the Bayesian approach is that uncertainty intervals automatically allow for the skewness typically found in covariate measurement error adjusted estimators.

As has been noted by many authors before, a key strength of the Bayesian approach is its flexibility to handle more complicated models and data structures. As we have demonstrated in Section 7, the Bayesian approach can readily accommodate both covariate measurement error and missing data. Moreover, more complex measurement error models can in principle be used, for example, to allow for heteroscedastic error, systematically biased measurements, or more flexible modelling of the true covariate's distribution.[1] The flexibility of the Bayesian approach also lends itself to the problem of adjusting for covariate measurement error when the true covariate is assumed to have a complex non-linear association with the outcome.[36] In this paper, we have focused on the setting whereby internal replication data are available; the Bayesian approach readily handles the situation where validation data are instead available.

Nonetheless, the Bayesian approach has a number of drawbacks. As a fully parametric approach, a natural concern is sensitivity of inferences to distributional assumptions, particularly those about the unobserved true covariate and measurement errors. In this regard, the Bayesian approach can utilize more flexible model specifications, for example, by modelling the unobserved true covariate using a normal mixture model.[37] An important practical issue is that although the software available for fitting complex analyst defined models using the Bayesian approach has seen dramatic developments over the last 25 years,[22] fitting certain models (e.g. Cox proportional hazards models) can still take tremendously long. Although this concern will be progressively mitigated by increasing computational power, it is arguably still a material drawback. Further research and effort are therefore warranted to develop software implementations of the Bayesian approach which mitigate this.

Funding

This paper was written while JWB was a member of the Department of Medical Statistics, London School of Hygiene and Tropical Medicine and was supported by a Medical Research Council Fellowship [MR/K02180X/1]. RHK was supported by a Medical Research Council Fellowship [MR/M014827/1].

Supplementary material

R and JAGS code demonstrating each of the simulation setups, and code and data for the illustrative analysis, are provided at the GitHub repository: https://github.com/jwb133/bayesMeasurementError

References

1. Carroll RJ, Ruppert D, Stefanski LA, et al. *Measurement error in nonlinear models*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall/CRC, 2006.
2. Armstrong B. Measurement error in the generalised linear model. *Comm Stat Simulat Comput* 1985; **14**: 529–544.
3. Rosner B, Spiegelman D and Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990; **132**: 734–743.
4. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982; **69**: 331–342.
5. Wang CY, Hsu L, Feng D, et al. Regression calibration in failure time regression. *Biometrics* 1997; **53**: 131–145.
6. Strawbridge AD. *Modelling non-linear exposure-disease relationships in a large individual participant meta-analysis allowing for the effects of exposure measurement error*. PhD thesis, University of Cambridge, UK, 2011.
7. Guo Y and Little RJ. Regression analysis with covariates that have heteroscedastic measurement error. *Stat Med* 2011; **30**: 2278–2294.
8. Richardson S and Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *Am J Epidemiol* 1993; **138**: 430–442.
9. Richardson S and Gilks WR. Conditional independence models for epidemiological studies with covariate measurement error. *Stat Med* 1993; **12**: 1703–1722.
10. Gustafson P. *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. Boca Raton, FL, USA: Chapman & Hall/CRC, 2003.
11. Little R. Calibrated Bayes, for statistics in general, and missing data in particular. *Stat Sci* 2011; **26**: 162–174.
12. Cox DR. Regression models and life-tables. *J Roy Stat Soc B* 1972; **34**: 187–220.
13. Keogh RH and White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Stat Med* 2014; **33**: 2137–2155.
14. Kuha J. Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Stat Med* 1994; **13**: 1135–1148.
15. Hughes MD. Regression dilution in the proportional hazards model. *Biometrics* 1993; **49**: 1056–1066.
16. Berger J. The case for objective Bayesian analysis. *Bayesian Anal* 2006; **1**: 385–402.
17. Robert CP. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. New York: Springer, 2007.
18. Kalbfleisch JD. Non-parametric Bayesian analysis of survival time data. *J Roy Stat Soc B* 1978; **40**: 214–221.
19. Sinha D, Ibrahim JG and Chen M. A Bayesian justification of Cox's partial likelihood. *Biometrika* 2003; **90**: 629–641.
20. Gelman A, Carlin JB, Stern HS, et al. *Bayesian data analysis*, 2nd ed. Boca Raton: Chapman & Hall/CRC, 2004.
21. Greenland S and Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med* 2015; **34**: 3133–3143.
22. Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: evolution, critique and future directions. *Stat Med* 2009; **28**: 3049–3067.
23. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing* 20 March 2003, Vol. 124, p. 125.

24. Cole SR, Chu H and Greenland S. Multiple-imputation for measurement-error correction. *Int J Epidemiol* 2006; **35**: 1074–1081.
25. Freedman LS, Midthune D, Carroll RJ, et al. A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. *Stat Med* 2008; **27**: 5195–5216.
26. Messer K and Natarajan L. Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. *Stat Med* 2008; **27**: 6332–6350.
27. Guo Y and Little RJ. Bayesian multiple imputation for assay data subject to measurement error. *J Stat Theor Pract* 2013; **7**: 219–232.
28. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley, 1987.
29. Carpenter JR and Kenward MG. *Multiple imputation and its application*. Chichester: John Wiley & Sons Ltd., 2013.
30. Wang N and Robins JM. Large-sample theory for parametric multiple imputation procedures. *Biometrika* 1998; **85**: 935–948.
31. Bartlett JW, Seaman SR, White IR, et al. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Meth Med Res* 2015; **24**: 462–487.
32. Hamra GB, MacLehose RF and Cole SR. Sensitivity analyses for sparse-data problems using weakly informative Bayesian priors. *Epidemiology* 2013; **24**: 233–239.
33. Xie SX, Wang CY and Prentice RL. A risk set calibration method for failure time regression by using a covariate reliability sample. *J Roy Stat Soc B* 2001; **63**: 855–870.
34. Post by Martyn Plummer, JAGS discussion forum. https://sourceforge.net/p/mcmc-jags/discussion/610036/thread/d5249e71/#8c3b (accessed 21 December 2015).
35. White IR and Royston P. Imputing missing covariate values for the Cox model. *Stat Med* 2009; **28**: 1982–1998.
36. Berry SM, Carroll RJ and Ruppert D. Bayesian smoothing and regression splines for measurement error problems. *J Am Stat Assoc* 2002; **97**: 160–169.
37. Richardson S, Leblond L, Jaussent I, et al. Mixture models in measurement error problems, with reference to epidemiological studies. *J Roy Stat Soc A* 2002; **165**: 549–566.