Biostatistics (2016), **17**, 2, *pp.* 277–290
doi:10.1093/biostatistics/kxv043
Advance Access publication on November 3, 2015

Measurement error models with interactions

DOUGLAS MIDTHUNE*

Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, 9609 Medical Center Drive, Room 5E122, Bethesda, MD 20892, USA

midthund@mail.nih.gov

RAYMOND J. CARROLL

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA and School of Mathematical Sciences, University of Technology, Sydney, Broadway, NSW 2007, Australia

LAURENCE S. FREEDMAN

Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer 52161, Israel

VICTOR KIPNIS

Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, 9609 Medical Center Drive, Room 5E118, Bethesda, MD 20892, USA

Summary

An important use of measurement error models is to correct regression models for bias due to covariate measurement error. Most measurement error models assume that the observed error-prone covariate (W) is a linear function of the unobserved true covariate (X) plus other covariates (Z) in the regression model. In this paper, we consider models for W that include interactions between X and Z. We derive the conditional distribution of X given W and Z and use it to extend the method of regression calibration to this class of measurement error models. We apply the model to dietary data and test whether self-reported dietary intake includes an interaction between true intake and body mass index. We also perform simulations to compare the model to simpler approximate calibration models.

Keywords: Interactions; Measurement error; Mixed models; Nonlinear mixed models; Nutritional epidemiology.

1. Introduction

One of the important uses of measurement error models is to correct estimated regression parameters for bias due to covariate measurement error. In this setting, we have a response variable Y, covariates X and Z, and a surrogate W, which is a measurement of X that includes error. We have a "risk" model that specifies the conditional distribution of Y given (X, Z) and a measurement error model that specifies the conditional

*To whom correspondence should be addressed.

distribution of W given (X, Z). The problem is to estimate the parameters in the risk model when Y, W, and Z (but not X) are observed.

There is a large body of literature on methods to address this problem for linear (Fuller, 1987) and nonlinear (Carroll *and others*, 2006; Buonaccorsi, 2010) risk models, including methods based on maximum likelihood, regression calibration (Prentice, 1982; Carroll and Stefanski, 1990), conditional scores (Stenfanski and Carroll, 1987), moment reconstruction (Freedman *and others*, 2004), and multiple imputation (Cole *and others*, 2006). In addition, many authors have considered the problem of correcting for measurement error when the risk model includes interaction terms. Fuller (1987) gives an example of a linear risk model that includes an interaction between X and Z, while Carroll *and others* (2006) show how to use conditional scores estimation for linear and logistic regression models with interactions. Huang *and others* (2005) consider interactions in the special case that Z is categorical, while Murad and Freedman (2007) consider the case when the risk model includes an interaction between two continuous covariates that are both measured with error.

In contrast, there has been relatively little attention paid to the case when the measurement error model (rather than risk model) includes interaction terms. Prentice *and others* (2002) proposed a model for W that includes interactions between scalar X and a vector of covariates Z,

$$W = \beta_0 + \beta_1 X + \boldsymbol{\beta}_2^{\mathrm{T}} Z + \boldsymbol{\beta}_3^{\mathrm{T}} ZX + e, \tag{1.1}$$

where e is random error with $E(e \mid X, Z) = 0$. They proposed the model for self-reported dietary intake data, noting there was evidence that measurement error in self-reported intake may depend on personal characteristics such as body mass, age, and social desirability factors. Equation (1.1) models the mean of W as a linear regression on X, Z, and ZX; in theory, more complex relationships could be posited.

Sugar *and others* (2007) developed methods for correcting parameter estimates in logistic regression under measurement error model (1.1), but restricted their attention to the case when Z is a vector of categorical variables. This allowed them to partition the data into subsets in which Z is constant, so that within the kth subset the measurement error model simplifies to $W = \beta_{k0} + \beta_{k1} X + e$. In their paper, they extended the methods of regression calibration and conditional scores estimation to this class of measurement error models.

Neuhouser *and others* (2002) also considered model (1.1), this time allowing Z to be continuous. They claimed that under model (1.1), and under normality assumptions for (X, e) given Z, the conditional expectation of X given W and Z is given by

$$E(X \mid W, Z) = \lambda_0 + \lambda_1 W + \lambda_2^{\mathrm{T}} Z + \lambda_3^{\mathrm{T}} ZW. \tag{1.2}$$

They used (1.2) to develop calibration equations for total energy intake, protein intake and percent energy from protein that included a potential interaction between self-reported intake from a food frequency questionnaire (FFQ) (W) and body mass index (BMI) (Z).

In this paper, we derive the conditional distribution of X given W and continuous Z under model (1.1), and show that $E(X \mid W, Z)$ is in general different from, and more complex than, (1.2). We also extend regression calibration to this class of models. In Section 2, we consider the case when X and W are scalars, while in Section 3, we extend the model to multivariate X and W. In Section 4, we investigate how interactions in (1.1) affect estimation of risk parameters in linear risk models. In Section 5, we fit the model to dietary intake data in the Observing Protein and Energy Nutrition (OPEN) study (Subar *and others*, 2003) and look for evidence of an interaction between true intake and BMI. In Section 6, we perform simulations to compare the performance of regression calibration under model (1.1) to simpler approximate calibration models such as (1.2). We conclude with a short discussion that includes consideration of some alternative approaches.

2. Measurement error model with scalar X and W

2.1 *Model and main results*

For the ith subject in a study, let Y_i be a response variable, X_i be the exposure of interest, and Z_i be a $q \times 1$ vector of covariates. We want to estimate the parameters in a generalized linear model relating Y_i to X_i and Z_i, which we call the risk model,

$$E(Y_i \mid X_i, Z_i) = h(\delta_0 + \delta_1 X_i + \delta_2^{\mathrm{T}} Z_i), \tag{2.1}$$

where $h(x)$ is the inverse link function, δ_0 and δ_1 are scalars, and δ_2 is a $q \times 1$ vector of regression coefficients. For example, if Y_i is binary, $h(x)$ could be the logistic distribution function. We do not observe X_i but instead observe W_i, which is a measure of X_i that includes error. We assume the following measurement error model for W_i:

$$W_i = \beta_0 + \beta_1 X_i + \beta_2^{\mathrm{T}} Z_i + \beta_3^{\mathrm{T}} Z_i X_i + e_i, \tag{2.2}$$

where random error e_i is normally distributed with mean zero and variance σ_e^2, and e_i is independent of X_i and Z_i. Model (2.2) is a special case of (1.1) in which e_i has a normal distribution. We need to specify the distribution of e_i to be able to define the conditional distribution of W_i and X_i given Z_i.

Our goal is to use regression calibration to correct estimated regression parameters in the risk model for bias due to measurement error in W_i. In regression calibration, one substitutes the predicted covariate $E(X_i \mid W_i, Z_i)$ for unknown X_i in risk model (2.1) and then fits the resulting risk model. Under the assumption that W_i has nondifferential error (i.e. that W_i and Y_i are conditionally independent given X_i and Z_i), regression calibration provides consistent risk estimates for linear risk models and nearly consistent estimates for many generalized linear risk models (Carroll *and others*, 2006). Because of its simplicity and wide applicability, regression calibration is one of the most widely used measurement error correction methods.

To estimate $E(X_i \mid W_i, Z_i)$, we will need, in addition to model (2.2), a model for the conditional distribution of X_i given Z_i. We will assume that

$$X_i = \gamma_0 + \gamma_1^{\mathrm{T}} Z_i + u_i, \tag{2.3}$$

where u_i is normally distributed with mean zero and variance σ_u^2 and u_i is independent of Z_i and e_i.

The following result is proved in Appendix A.1 in supplementary material available at *Biostatistics* online.

Proposition 1 Under models (2.2) and (2.3), with e_i and u_i independent and normally distributed with zero means and variances σ_e^2 and σ_u^2, respectively, the conditional distribution of X_i given W_i and Z_i is normal with mean

$$E(X_i \mid W_i, Z_i) = \gamma_0 + \gamma_1^{\mathrm{T}} Z_i + \gamma_2(Z_i)\{W_i - E(W_i \mid Z_i)\}, \tag{2.4}$$

and variance

$$\mathrm{var}(X_i \mid W_i, Z_i) = \sigma_u^2\{1 - \rho^2(Z_i)\},$$

where

$$E(W_i \mid Z_i) = \beta_0 + \beta_1\gamma_0 + (\beta_2 + \gamma_1\beta_1 + \beta_3\gamma_0)^{\mathrm{T}} Z_i + Z_i^{\mathrm{T}}\beta_3\gamma_1^{\mathrm{T}} Z_i,$$

$$\gamma_2(Z_i) = \frac{\mathrm{cov}(X_i, W_i \mid Z_i)}{\mathrm{var}(W_i \mid Z_i)} = \frac{(\beta_1 + \beta_3^{\mathrm{T}} Z_i)\sigma_u^2}{(\beta_1 + \beta_3^{\mathrm{T}} Z_i)^2\sigma_u^2 + \sigma_e^2},$$

$$\rho^2(Z_i) = \mathrm{corr}^2(X_i, W_i \mid Z_i) = \frac{(\beta_1 + \beta_3^{\mathrm{T}} Z_i)^2\sigma_u^2}{(\beta_1 + \beta_3^{\mathrm{T}} Z_i)^2\sigma_u^2 + \sigma_e^2}.$$

Observe that the conditional variance of X_i is a function of Z_i.

When used in regression calibration, (2.4) is sometimes called a calibration model or calibration equation. The conditional distribution of X_i given W_i and Z_i can also be used in other correction methods such as maximum likelihood and conditional scores. Equations (1.2) and (2.4) are not equivalent unless $\beta_3 = 0$, so that a measurement error model with an interaction in X_i and Z_i leads to a calibration model with functions of W_i and Z_i that are more complex than a simple interaction model.

If X_i is observed on a subset of the subjects, then models (2.2) and (2.3) can be fitted and used to calculate the predicted values in (2.4). Otherwise, one needs to observe repeated measures of an unbiased reference measure $\tilde{R}_i = (R_{i1}, \ldots, R_{iJ_i})^{\mathrm{T}}$, where J_i is the number of repeated measurements for the ith subject, and $J_i > 1$ for at least a subset of the subjects. We assume that

$$R_{ij} = X_i + \xi_{ij}, \tag{2.5}$$

where within-person errors ξ_{ij} are independent of each other and of (X_i, e_i, u_i), and are normally distributed with zero mean and variance σ_ξ^2. Typically, references R_{ij} are more expensive to measure than W_i, so that W, Z, and Y are measured in the main study, while W, Z, and $\tilde{R}_i$ are measured in a smaller calibration study.

In Appendix A.2 in supplementary material available at *Biostatistics* online, we show that the parameters in (2.2), (2.3), and (2.5) are identifiable. In Appendix A.3, we show how to use a nonlinear mixed effects modeling program to estimate the parameters in (2.2), (2.3), and (2.5) when W_i, Z_i, and $\tilde{R}_i$ are observed in a calibration substudy.

In practice, the interaction term in model (2.2) can lead to multicollinearity and large standard errors (s.e.) for the estimated regression coefficients. To avoid this, some authors suggest centering the covariates by replacing X_i and Z_i with $X_i - \mu_X$ and $Z_i - \mu_Z$ in (2.2) (Afshartous and Preston, 2011). This reparameterization changes the interpretation of the regression coefficients in (2.2) but does not affect the parameter estimates for risk model (2.1), although one needs to keep in mind that (2.4) is now the conditional expectation of $X_i - \mu_X$.

2.2 *An alternative model*

It is worth considering at this point the kind of measurement error model that would lead to calibration model (1.2). Let

$$X_i = \lambda_0 + \lambda_1 W_i + \lambda_2^{\mathrm{T}} Z_i + \lambda_3^{\mathrm{T}} Z_i W_i + u_i^*, \tag{2.6}$$

$$W_i = \gamma_{0*} + \gamma_{1*}^{\mathrm{T}} Z_i + e_i^*, \tag{2.7}$$

where u_i^* and e_i^* are normal with mean zero and are independent of each other and of W_i and Z_i. Using similar reasoning as that described in Section 2.1, one can show that the conditional distribution of W_i

given X_i and Z_i is normal with mean

$$E(W_i \mid X_i, Z_i) = \gamma_{0*} + \gamma_{1*}^{\mathrm{T}} Z_i + \gamma_{2*}(Z_i)\{X_i - E(X_i|Z_i)\},$$

and variance

$$\mathrm{var}(W_i \mid X_i, Z_i) = \sigma_{e^*}^2\{1 - \rho_*^2(Z_i)\},$$

where

$$\gamma_{2*}(Z_i) = \frac{(\lambda_1 + \lambda_3^{\mathrm{T}} Z_i)\sigma_{e^*}^2}{(\lambda_1 + \lambda_3^{\mathrm{T}} Z_i)^2\sigma_{e^*}^2 + \sigma_{u^*}^2}, \quad \rho_*^2(Z_i) = \frac{(\lambda_1 + \lambda_3^{\mathrm{T}} Z_i)^2\sigma_{e^*}^2}{(\lambda_1 + \lambda_3^{\mathrm{T}} Z_i)^2\sigma_{e^*}^2 + \sigma_{u^*}^2}.$$

In general, therefore, the measurement error model and regression calibration model cannot simultaneously be linear regressions with simple interaction terms.

3. Multivariate measurement error model

In this section, we extend the measurement error model introduced in Section 2 to the case when X and W are vectors. Let X_i be a $p \times 1$ vector of unobserved covariates, W_i be the corresponding vector of observed covariates that are measured with error, and Z_i be a $q \times 1$ vector of covariates that are measured without error. We assume a measurement error model that allows interactions between X_i and Z_i,

$$W_i = B_0 + B_1 X_i + B_2 Z_i + B_3(Z_i \otimes X_i) + e_i, \tag{3.1}$$

where B_0 is a $p \times 1$ vector of intercepts, B_1 a $p \times p$ matrix of regression coefficients, B_2 a $p \times q$ matrix of coefficients, B_3 a $p \times pq$ matrix of interaction terms, and $\otimes$ is the Kronecker product. To include only a subset of the possible interactions, one can set the other components of B_3 equal to zero. Within-person error e_i is a multivariate normal random vector with zero mean and covariance matrix Σ_e, and e_i is independent of X_i and Z_i.

As in the scalar case in Section 2, we also need to assume a model for the conditional distribution of X_i given Z_i. We will assume that

$$X_i = \Gamma_0 + \Gamma_1 Z_i + u_i, \tag{3.2}$$

where Γ_0 is a $p \times 1$ vector of intercepts, Γ_1 a $p \times q$ matrix of regression coefficients, u_i a multivariate normal random vector with zero mean and covariance matrix Σ_u, and u_i is independent of Z_i and e_i. As in Section 2, in order to fit model (3.1) and (3.2) one would need to observe X_i or repeat observations of an unbiased reference measure on a subset of the subjects.

The following result is proved in Appendix A.4 in supplementary material available at *Biostatistics* online.

Proposition 2 Under models (3.1) and (3.2), with e_i and u_i independent and normally distributed with zero means and covariance matrices Σ_e and Σ_u, respectively, the conditional distribution of X_i given W_i and Z_i is multivariate normal with mean

$$E(X_i \mid W_i, Z_i) = \Gamma_0 + \Gamma_1 Z_i + \Gamma_2(Z_i)\{W_i - E(W_i \mid Z_i)\},$$

and covariance matrix

$$\mathrm{cov}(X_i \mid W_i, Z_i) = \Sigma_u - \Sigma_u K^{\mathrm{T}}(Z_i)\{K(Z_i)\Sigma_u K^{\mathrm{T}}(Z_i) + \Sigma_e\}^{-1} K(Z_i)\Sigma_u,$$

where $E(W_i \mid Z_i) = B_0 + B_1\ \Gamma_0 + (B_2 + B_1\ \Gamma_1)\ Z_i + B_3\ \{Z_i \otimes (\Gamma_0 + \Gamma_1\ Z_i)\}$, $K(Z_i) = B_1 + B_3\ (Z_i \otimes I_p)$, and $\Gamma_2(Z_i) = \mathrm{cov}(X_i, W_i \mid Z_i)\mathrm{cov}(W_i \mid Z_i)^{-1} = \Sigma_u K^{\mathrm{T}}(Z_i)\{K(Z_i)\Sigma_u K^{\mathrm{T}}(Z_i) + \Sigma_e\}^{-1}$.

As in the scalar case, the calibration model is more complex than a linear regression with a simple interaction term.

4. LINEAR RISK MODELS

Regression calibration is known to produce consistent estimates when the risk model is linear regression and the calibration model is correctly specified. In Sections 2 and 3, we showed that measurement error models with interaction terms lead to complex calibration models. In this section, we investigate whether simpler approximate calibration models can produce consistent estimates in linear risk models when the true measurement error model includes interactions. For simplicity, we consider the case when X_i and Z_i are both scalar. The risk model is given by the linear regression of Y_i on V_i,

$$Y_i = \delta_0 + \delta^{\mathrm{T}} V_i + \eta_i, \tag{4.1}$$

where V_i is a vector of covariates, $\delta = \mathrm{cov}^{-1}(V_i)\mathrm{cov}(V_i, Y_i)$ is the vector of regression coefficients, and η_i is random error that is uncorrelated with V_i and has mean zero and constant variance. We are interested in two cases: $V_i = (X_i, Z_i)^{\mathrm{T}}$, a risk model without an interaction term; and $V_i = (X_i, Z_i, Z_i X_i)^{\mathrm{T}}$, a risk model that includes an interaction.

Let Q_i be a vector of observed covariates. The best linear approximation (in the mean square sense) of the true regression of X_i on Q_i is

$$X_i = \lambda_0 + \lambda^{\mathrm{T}} Q_i + u_i^*, \tag{4.2}$$

where $\lambda = \mathrm{cov}^{-1}(Q_i)\mathrm{cov}(Q_i, X_i)$. Again, we are interested in two cases: $Q_i = (W_i, Z_i)^{\mathrm{T}}$ and $Q_i = (W_i, Z_i, Z_i W_i)^{\mathrm{T}}$. Let $X_{ci} = \lambda_0 + \lambda^{\mathrm{T}} Q_i$, and let $V_{ci} = (X_{ci}, Z_i)^{\mathrm{T}}$ if $V_i = (X_i, Z_i)^{\mathrm{T}}$ or $V_{ci} = (X_{ci}, Z_i, Z_i X_{ci})^{\mathrm{T}}$ if $V_i = (X_i, Z_i, Z_i X_i)^{\mathrm{T}}$. The approximate risk model based on calibration model (4.2) is

$$Y_i = \delta_{c0} + \delta_c^{\mathrm{T}} V_{ci} + \eta_{ci}. \tag{4.3}$$

If W has nondifferential measurement error, then

$$\delta_c = \mathrm{cov}^{-1}(V_{ci})\mathrm{cov}(V_{ci}, Y_i) = \mathrm{cov}^{-1}(V_{ci})\mathrm{cov}(V_{ci}, \delta^{\mathrm{T}} V_i) = \mathrm{cov}^{-1}(V_{ci})\mathrm{cov}(V_{ci}, V_i)\delta.$$

This implies that regression calibration based on approximate model (4.2) leads to consistent estimation of (nonzero) δ if and only if $\mathrm{cov}(V_{ci}, V_i) = \mathrm{cov}(V_{ci})$.

The following result is proved in Appendix A.5 in supplementary material available at *Biostatistics* online.

PROPOSITION 3 Under models (2.2) and (2.3), with (X_i, Z_i) bivariate normally distributed, and e_i normally distributed with mean zero and independent of (X_i, Z_i), the following are true:

(i) If $V_i = (X_i, Z_i)$, then $\mathrm{cov}(V_{ci}, V_i) = \mathrm{cov}(V_{ci})$.
(ii) If $V_i = (X_i, Z_i, Z_i X_i)$, then $\mathrm{cov}(V_{ci}, V_i) = \mathrm{cov}(V_{ci})$ if and only if $\beta_3 = 0$.

Proposition 3 implies that the estimated regression parameters in a linear risk model based on approximate calibration model (4.2) will be consistent if the risk model does not include an interaction with unobserved covariate X_i, but will be inconsistent if the risk model includes such an interaction unless the regression coefficient for the interaction term in measurement error model (2.2) equals zero.

Table 1. *Estimated regression coefficients for measurement error models with and without interaction between true dietary intake and BMI*; *OPEN study*

			Covariates in measurement error model				
Nutrient	Gender	Model	X (s.e.)	Z (s.e.)	XZ (s.e.)	AIC	p-value
Energy	Male	1	0.66 (0.19)	−0.46 (0.19)	–	73.1	
		2	0.65 (0.19)	−0.52 (0.19)	1.23 (0.88)	73.2	0.17
	Female	1	0.14 (0.23)	0.19 (0.16)	–	70.9	
		2	0.15 (0.23)	0.20 (0.17)	−0.13 (0.80)	72.9	0.86
Protein	Male	1	0.82 (0.18)	−0.45 (0.19)	–	209.6	
		2	0.81 (0.18)	−0.51 (0.20)	1.87 (0.80)	206.0	0.02
	Female	1	0.85 (0.28)	−0.19 (0.19)	–	285.7	
		2	0.84 (0.28)	−0.22 (0.19)	0.61 (0.74)	287.0	0.40

Model 1 is without interaction, and Model 2 is with interaction. AIC = −2(log-likelihood − number of parameters) (smaller is better). The p-value is for the likelihood-ratio test comparing models 1 and 2.

For linear risk models that include interactions, we refer to $\mathrm{cov}^{-1}(V_{ci})\mathrm{cov}(V_{ci}, V_i)$ as the "bias matrix" for the approximate model.

5. The OPEN study

In this section, we evaluate measurement error in self-reported dietary intake in the OPEN study and look for evidence of an interaction between true intake and BMI. The design of the OPEN study is described in Subar *and others* (2003). Briefly, 484 subjects (261 men, 223 women) were recruited into the study and asked to complete two self-report dietary instruments: an FFQ and a 24-h dietary recall. Two biomarker measures of dietary intake were also collected: 24-h urinary nitrogen for protein intake and doubly labeled water for total energy intake. These biomarkers have been shown in feeding studies to provide approximately unbiased measures of true intake (Bingham and Cummings, 1985; Schoeller, 1988). The urinary nitrogen biomarker was measured twice for each individual, about 10 days apart. The doubly labeled water biomarker was measured once for each individual, and was measured a second time two weeks later in a small subset of 25 individuals.

Kipnis *and others* (2003) evaluated the measurement error structure of FFQ-reported intakes of energy and protein in OPEN, using the biomarkers as reference measures and a measurement error model that did not include interactions. In the present analysis, we allow for an interaction between true intake and BMI. Let W be log-transformed FFQ-reported intake of energy or protein, R be the corresponding log-transformed biomarker measurements, and Z be the logarithm of BMI. As an initial step, we center W, R, and Z by subtracting their means; this is done to avoid multicollinearity in models with interaction terms, as discussed in Section 2. We then calculate maximum likelihood estimates of the parameters in model (2.2) using the SAS NLMIXED procedure (see Appendix A.3 in supplementary material available at *Biostatistics* online for details). The model can also be fitted using the nlme package in R. We fit two versions of the model; model 1 assumes no interaction ($\beta_3 = 0$), while model 2 allows for the interaction.

Table 1 presents the results of the analysis, including likelihood-ratio tests of model 1 vs. model 2. Men and women were analyzed separately. For energy intake, there is no evidence of an interaction between X and Z, while for protein intake, there is evidence of an interaction in males ($p = 0.02$), but not in females ($p = 0.4$). The standard errors (s.e.) for the interaction terms in Table 1 are rather large, indicating only a limited power to detect interactions in studies of this size.

Table 2. *Estimated regression coefficients for calibration models with and without interaction between reported dietary intake and BMI; OPEN study*

			Covariates in calibration model				
Nutrient	Gender	Model	W (s.e.)	Z (s.e.)	WZ (s.e.)	AIC	*p*-value
Energy	Male	1	0.08 (0.02)	0.56 (0.06)	–	73.1	
		2	0.08 (0.02)	0.56 (0.06)	−0.03 (0.15)	75.0	0.75
	Female	1	0.02 (0.03)	0.43 (0.05)	–	70.9	
		2	0.02 (0.03)	0.44 (0.05)	−0.17 (0.14)	71.4	0.22
Protein	Male	1	0.16 (0.03)	0.56 (0.09)	–	209.6	
		2	0.16 (0.03)	0.53 (0.09)	0.22 (0.23)	210.7	0.34
	Female	1	0.14 (0.04)	0.41 (0.08)	–	285.7	
		2	0.14 (0.04)	0.41 (0.08)	0.14 (0.19)	287.2	0.44

Model 1 is without interaction, and Model 2 is with interaction. AIC = −2(log-likelihood − number of parameters) (smaller is better). The *p*-value is for the likelihood-ratio test comparing models 1 and 2.

As a comparison, we also fitted calibration model (1.2). Typically, the parameters in model (1.2) are estimated by ordinary least squares; in order to facilitate comparison with the previous model, we estimated them by maximum likelihood based on equations (2.6) and (2.7). The results are shown in Table 2. We found no evidence of an interaction for energy or protein in males or females. For protein in men, the difference in the Akaike Information Criterion (AIC) for model 2 in Tables 1 and 2 is 4.7, indicating that the measurement error model with interaction fits better than the calibration model with interaction (the two models have the same number of parameters). For protein in women and energy in men and women, the difference in AIC is < 2.

In Section 4, we showed that using approximate calibration model (1.2) when the true measurement error model is (2.2) leads in general to biased estimation in linear risk models that include interactions. For protein intake in men, the estimated bias matrix for calibration model (1.2) is

$$B = \begin{bmatrix} 1.03 & 0 & 0 \\ -0.01 & 1 & -0.01 \\ -0.24 & 0 & 0.95 \end{bmatrix}.$$

The bias matrix is used to estimate bias in linear risk models with interactions when regression calibration is based on the approximate model. For example, true risk parameters $\delta = (1, 1, 1)^{\mathrm{T}}$ would on average be estimated as $\delta_c = B\delta = (1.03, 0.98, 0.71)^{\mathrm{T}}$. In this example, the bias is only moderate, but in other situations it could more substantial. For example, the estimated regression coefficient of the interaction term for protein in men is $\beta_3 = 1.87$, with 95% confidence interval = (0.29, 3.45); if β_3 had been larger, say $\beta_3 = 3.45$, while the other parameters remained the same, the bias matrix would have been

$$B = \begin{bmatrix} 1.09 & 0 & 0.01 \\ -0.04 & 1 & -0.01 \\ -0.49 & 0 & 0.87 \end{bmatrix},$$

and $\delta = (1, 1, 1)^{\mathrm{T}}$ would have been estimated as $\delta_c = (1.10, 0.94, 0.38)^{\mathrm{T}}$.

Table 3. *True measurement error parameters for the simulations*

Case	β_1	β_2	β_3
1	0.8	−0.5	0
2	0.8	−0.5	2
3	0.4	−0.5	4
4	0.4	−0.5	−4
5	0.4	−4	4
6	0.4	−4	−4

6. Simulation study

In Section 4, we investigated the consistency of regression calibration estimates for linear risk models when an approximate calibration model is used. In this section, we use simulations to investigate the performance of regression calibration for a nonlinear risk model under measurement error model (2.2), and compare calibration model (2.4) to simpler approximate calibration models.

6.1 *Description of simulations*

In the simulated data, Z_i is generated from a normal distribution with mean zero and standard deviation 0.25, W_i, X_i, and R_{ij} are generated from models (2.2), (2.3), and (2.5), and Y_i is a binary response that is related to X_i and Z_i. We consider two risk models,

$$\text{pr}(Y_i = 1 \mid X_i, Z_i) = H(\delta_0 + \delta_1 X_i + \delta_2 Z_i), \tag{6.1}$$

$$\text{pr}(Y_i = 1 \mid X_i, Z_i) = H(\delta_0 + \delta_1 X_i + \delta_2 Z_i + \delta_3 Z_i X_i), \tag{6.2}$$

where $H(x)$ is the logistic distribution function. Risk model (6.1) includes no interaction terms, while model (6.2) includes an interaction between X_i and Z_i. Since an interaction in the measurement error model does not imply an interaction in the risk model (or vice versa), both cases are of interest. In all simulations, we set $\delta_1 = \delta_2 = 1$, and set $\delta_0 = -2.2$, so that the overall probability $\text{pr}(Y_i = 1) \approx 0.1$. In the simulations for risk model (6.2), we set $\delta_3 = 1$.

Simulations are based on the estimated measurement error parameters for protein in men in the OPEN study. In all simulations, we set $\gamma_0 = 0$, $\gamma_1 = 0.5$, $\sigma_e = 0.5$, $\sigma_u = 0.25$, $\sigma_\xi = 0.25$, and $\beta_0 = 0$. Parameters β_1, β_2, and β_3 vary by simulation, as shown in Table 3. For each simulation, we simulate a main study of 100 000 subjects with observed covariates W_i and Z_i and binary response Y_i, and a calibration study of 1000 subjects with observed covariates W_i and Z_i and repeat measurements of unbiased reference measure R_{ij}, $j = 1, 2$. The relative sample sizes of the main and calibration studies are typical of the large prospective cohorts used in nutritional epidemiology. The Women's Health Initiative (WHI) Observational Study, for example, is a cohort of 93 000 women with a calibration study of 450 women (Zheng *and others*, 2014). The calibration study is used to estimate the parameters in a calibration model, which can then be used to predict true intake for subjects in the main study. Simulation results are based upon 1000 simulated data sets.

We compare three calibration models for use with regression calibration:
Calibration Model 1: Equation (2.4), based on true measurement error model (2.2).
Calibration Model 2: Equation (4.2), with $Q_i = (W_i, Z_i, Z_i W_i)^{\mathrm{T}}$.
Calibration Model 3: Equation (4.2), with $Q_i = (W_i, Z_i)^{\mathrm{T}}$.

6.2 *Simulation results*

We simulated cases 1–6 described in Table 3. Simulation results for risk model (6.1), which does not include an interaction, are presented in Table 4 and are summarized as follows:

- In case 1, where the true measurement error model has no interaction ($\beta_3 = 0$), the three calibration models performed very similarly, giving unbiased estimates of the risk parameters, and having nearly the same standard deviations.
- In case 2, the true measurement error model includes an interaction similar to that seen in OPEN ($\beta_3 = 2$); again, the three calibration models performed similarly, although model 3 resulted in a small bias.
- In cases 3 and 4, the interaction term in the measurement error model ($|\beta_3| = 4$) is large compared with β_1. In these cases, calibration models 1 and 2 continued to perform well, but model 3 resulted in substantial bias and large standard deviations.
- In cases 5 and 6, both β_2 and β_3 are large compared with β_1. In these cases, calibration model 2 resulted in moderate bias, while model 1 had little or no bias.
- In all six cases, the estimated parameters in measurement error model (2.2) were approximately unbiased (results not shown).

In Section 4, we showed that calibration models 2 and 3 lead to consistent estimation of risk parameters in linear risk models that do not include interactions. These simulations indicate that the same is not true for nonlinear risk models.

Simulation results for risk model (6.2), which includes an interaction, are presented in Table 5. The results for main effects δ_1 and δ_2 are qualitatively similar to those in Table 4, and we limit our remarks to the results for interaction term δ_3. For case 1, the three calibration models performed similarly, with little or no bias and similar standard deviations. For cases 2–6, calibration model 1 resulted in a small underestimation of the interaction term δ_3, with bias ranging from 5% to 8%, while calibration models 2 and 3 resulted in more substantial, sometimes severe, bias.

In these simulations, we compared regression calibration using the true calibration model to simple approximate models that approximated the conditional expectation $E(X_i \mid W_i, Z_i)$ as a linear function of W_i, Z_i, and $Z_i W_i$ (model 2), or of W_i and Z_i (model 3). Calibration model 3 was too simple and led to biased estimates if β_3 was large compared with β_1. Model 2 performed better, but led to biased estimates when both β_2 and β_3 were large or the risk model included an interaction with unobserved covariate X. A notable feature of calibration model 1 is that it appears to be as efficient as model 3 even when the true measurement error model includes no interaction terms (case 1 in Tables 4 and 5).

For simplicity, we simulated the data so that X_i, W_i, and Z_i all had mean zero. In general, one might want to center risk model (6.2) by replacing X_i and Z_i with $X_i - \mu_X$ and $Z_i - \mu_Z$. Otherwise, any bias in the estimate of δ_3 would cause a bias in the estimated regression coefficients for X_i and Z_i.

7. DISCUSSION

We have shown that measurement error models with interactions lead to calibration models that are much more complex than simple interaction models, and have extended regression calibration to this class of models. For linear risk models, we showed that regression calibration using approximate model (4.2) when the true measurement error model is (2.2) can lead to biased estimation of the risk parameters if the risk model includes interactions. For nonlinear risk models, we showed through simulations that regression calibration using (4.2) can lead to bias, even if the risk model does not include interactions. More generally,

Table 4. *Simulation results when risk model does not have an interaction term*; *simulated means and standard deviations of estimated risk model parameters*

Case	Calibration model	Parameter $\delta_1 = 1.0$		Parameter $\delta_2 = 1.0$	
		Mean (s.e.)	Std dev	Mean (s.e.)	Std dev
1	1	1.00 (0.01)	0.15	0.99 (0.01)	0.09
	2	1.00 (0.01)	0.16	0.99 (0.01)	0.09
	3	1.01 (0.01)	0.15	0.99 (0.01)	0.09
2	1	1.00 (0.01)	0.12	0.98 (0.01)	0.08
	2	0.97 (0.01)	0.14	0.98 (0.01)	0.09
	3	1.13 (0.01)	0.17	0.90 (0.01)	0.10
3	1	1.00 (0.01)	0.11	0.99 (0.01)	0.07
	2	0.93 (0.01)	0.15	1.01 (0.01)	0.09
	3	1.71 (0.02)	0.60	0.60 (0.01)	0.31
4	1	0.99 (0.01)	0.12	1.00 (0.01)	0.08
	2	0.97 (0.01)	0.17	1.00 (0.01)	0.10
	3	0.33 (0.01)	0.30	1.33 (0.01)	0.16
5	1	1.00 (0.01)	0.11	0.98 (0.01)	0.07
	2	1.27 (0.01)	0.25	0.91 (0.01)	0.13
	3	1.75 (0.02)	0.63	0.58 (0.01)	0.32
6	1	0.99 (0.01)	0.12	1.00 (0.01)	0.08
	2	0.68 (0.01)	0.19	1.12 (0.01)	0.12
	3	0.34 (0.01)	0.35	1.33 (0.01)	0.19

The risk model is (6.1). "Case" refers to settings 1–6 in Table 3. Models: 1 = measurement error model with interaction; 2 = approximate calibration model with interaction; 3 = approximate calibration model without interaction.

the simulation results show the importance of choosing an appropriate calibration model. Simple linear approximations of the conditional expectation of X given W and Z are not always appropriate.

As discussed in Section 2.1, models (2.2) and (2.3) are not identifiable unless true covariate X or an unbiased reference instrument R is observed on a subset of study subjects. In our example, we used replicate measurements of a reference instrument to fit the measurement error model. If replicate measurements are not available, one could use a single application of a reference instrument if the within-individual variance of the instrument is known from previous studies.

We assumed a relatively simple measurement error model that may itself be inadequate in some situations. Possible extensions could include interactions between two or more covariates measured with error or, more generally, polynomials of covariates measured with and without error, although we do not know how stable such models might be, or even if they would be identifiable. Alternatively, one could consider nonparametric approaches such as interaction splines (Chen, 1993) or local estimating equations (Carroll *and others*, 1998). For example, Jiang *and others* (2003) used local estimating equations to estimate the conditional expectation of true protein intake (X) given observed intake (W) as a nonparametric function of BMI (Z). Such nonparametric procedures may protect against bias due to misspecification of the measurement error model, but at the expense of added variability in the estimators.

Wang (2012) proposed an estimator that treats W as an instrumental variable and is consistent under a very general measurement error model for W. A limitation of his estimator is that it requires one to fit the risk model in the calibration study. This may not be practical in the kind of study we consider here, where the disease risk is small and the size of the calibration study is very small compared with the main study.

Table 5. *Simulation results when risk model has an interaction term*; *simulated means and standard deviations of estimated risk model parameters*

Case	Calibration model	Parameter $\delta_1 = 1.0$		Parameter $\delta_2 = 1.0$		Parameter $\delta_3 = 1.0$	
		Mean (s.e.)	Std dev	Mean (s.e.)	Std dev	Mean (s.e.)	Std dev
1	1	1.00 (0.01)	0.16	1.03 (0.01)	0.10	0.97 (0.01)	0.22
	2	1.00 (0.01)	0.16	1.03 (0.01)	0.10	0.97 (0.01)	0.24
	3	1.00 (0.01)	0.16	1.03 (0.01)	0.10	0.97 (0.01)	0.22
2	1	1.01 (0.01)	0.14	1.01 (0.01)	0.09	0.95 (0.01)	0.21
	2	1.13 (0.01)	0.16	0.96 (0.01)	0.10	0.41 (0.01)	0.21
	3	1.25 (0.01)	0.19	0.86 (0.01)	0.11	0.92 (0.01)	0.21
3	1	1.01 (0.01)	0.12	1.01 (0.01)	0.08	0.92 (0.01)	0.20
	2	1.19 (0.01)	0.21	0.96 (0.01)	0.11	0.21 (0.01)	0.20
	3	2.42 (0.03)	0.87	0.28 (0.01)	0.44	0.74 (0.01)	0.26
4	1	0.99 (0.01)	0.12	1.04 (0.01)	0.08	0.92 (0.01)	0.19
	2	0.99 (0.01)	0.17	1.05 (0.01)	0.10	0.45 (0.01)	0.19
	3	−0.64 (0.01)	0.33	1.90 (0.01)	0.17	0.17 (0.01)	0.32
5	1	1.01 (0.01)	0.12	1.01 (0.01)	0.08	0.94 (0.01)	0.20
	2	1.47 (0.01)	0.34	0.86 (0.01)	0.18	1.78 (0.01)	0.27
	3	2.48 (0.03)	0.89	0.25 (0.01)	0.45	0.73 (0.01)	0.26
6	1	0.99 (0.01)	0.12	1.04 (0.01)	0.08	0.93 (0.01)	0.18
	2	0.80 (0.01)	0.33	1.16 (0.01)	0.18	0.01 (0.01)	0.25
	3	−0.66 (0.01)	0.38	1.91 (0.01)	0.19	0.17 (0.01)	0.31

The risk model is (6.2). "Case" refers to settings $1 - 6$ in Table 3. Models: 1 = measurement error model with interaction; 2 = approximate calibration model with interaction; 3 = approximate calibration model without interaction.

The WHI Observational Study, for example, includes 93 000 women in the main study, but only 450 in the calibration study. In an analysis of energy intake and disease risk in this cohort, Zheng *and others* (2014) reported 348 incident ovarian cancers out of 65 347 women analyzed; the number of ovarian cancers in the calibration study was not reported, but is presumably around $450 \times (348/65\,347) = 2.4$.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

R.J.C.'s research was supported by a grant from the National Cancer Institute (U01-CA057030).

REFERENCES

AFSHARTOUS, D. AND PRESTON, P. A. (2011). Key results of interaction models with centering. *Journal of Statistics Education* **19**, 1–24.

Bingham, S. A. and Cummings, J. H. (1985). Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *American Journal of Clinical Nutrition* **42**, 1276–1289.

Buonaccorsi, J. P. (2010) *Measurement Error: Models, Methods and Applications*. Boca Raton, Florida: Chapman and Hall CRC Press.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006) *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. Boca Raton, FL: Chapman and Hall CRC Press.

Carroll, R. J., Ruppert, D. and Welsh, A. H. (1998). Local estimating equations. *Journal of the American Statistical Association* **93**, 214–227.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasilikelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association* **85**, 652–663.

Cole, S. R., Chu, H. and Greenland, S. (2006). Multiple-imputation for measurement-error correction. *International Journal of Epidemiology* **35**, 1074–1081.

Chen, Z. (1993). Fitting multivariate regression functions by interaction splines. *Journal of the Royal Statistical Society, Series B* **55**, 473–491.

Freedman, L. S., Fainberg, V., Kipnis, V., Midthune, D. and Carroll, R. J. (2004). A new method for dealing with measurement error in explanatory variables of regression models. *Biometrics* **60**, 172–181.

Fuller, W. A. (1987) *Measurement Error Models*. New York: John Wiley and Sons.

Huang, L. S., Wang, H. and Cox, C. (2005). Assessing interaction effects in linear measurement error models. *Journal of the Royal Statistical Society, Series C* **54**, 21–30.

Jiang, W., Kipnis, V., Midthune, D. and Carroll, R. J. (2001). Parameterization and inference for nonparametric regression problems. *Journal of the Royal Statistical Society, Series B* **63**, 583–591.

Kipnis, V., Subar, A. F., Midthune, D., Freedman, L. S., Ballard-Barbash, R., Troiano, R., Bingham, S., Schoeller, D. A., Schatzkin, A. and Carroll, R. J. (2003). The structure of dietary measurement error: results of the OPEN biomarker study. *American Journal of Epidemiology* **158**, 14–21.

Murad, H. and Freedman, L. S. (2007). Estimating and testing interactions in linear regression models when explanatory variables are subject to classical measurement error. *Statistics in Medicine* **26**, 4293–4310.

Neuhouser, M. L., Tinker, L., Shaw, P. A., Schoeller, D., Bingham, S. A., Van Horn, L., Beresford, S. A. A., Caan, B., Thompson, C., Satterfield, S. *and others* (2008). Use of recovery biomarkers to calibrate nutrient composition self-reports in the Women's Health Initiative. *American Journal of Epidemiology* **167**, 1247–1259.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331–342.

Prentice, R. L., Sugar, E., Wang, C. Y., Neuhouser, M. and Patterson, R. (2002). Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Public Health Nutrition* **5**(6A), 977–984.

Schoeller, D. A. (1988). Measurement of energy expenditure in free-living humans by using doubly labeled water. *Journal of Nutrition* **118**, 1278–1289.

Stenfanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika* **74**, 703–716.

Subar, A. F., Kipnis, V., Troiano, R. P., Midthune, D., Schoeller, D., Bingham, S., Sharbaugh, C., Trabulsi, J., Runswick, S., Ballard-Barbash, R., Sunshine, J. and Schatzkin, A. (2003). Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *American Journal of Epidemiology* **158**, 1–13.

SUGAR, E. A., WANG, C. Y. AND PRENTICE, R. L. (2007). Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics* **63**, 143–151.

WANG, C. Y. (2012). Robust best linear estimation for regression analysis using surrogate and instrumental variables. *Biostatistics* **13**, 326–340.

ZHENG, C., BERESFORD, S. A., VAN HORN, L., TINKER, L. F., THOMSON, C. A., NEUHOUSER, M. L., DI, C., MANSON, J. E., MOSSAVAR-RAHMANI, Y., SEGUIN, R. *and others* (2014). Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *American Journal of Epidemiology* **180**, 526–535.

[*Received August 13, 2014; revised October 6, 2015; accepted for publication October 7, 2015*]