Published in final edited form as:

Biom J. 2019 May ; 61(3): 558–573. doi:10.1002/bimj.201700279.

Correcting for measurement error in fractional polynomial models using Bayesian modelling and regression calibration, with an application to alcohol and mortality

Christen M. Gray¹, Raymond J. Carroll², Marleen A. H. Lentjes³, and Ruth H. Keogh¹

¹Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

²Department of Statistics, Texas A&M University, College Station, TX, USA

³Department of Public Health & Primary Care, University of Cambridge, Strangeways Research Laboratories, Cambridge, United Kingdom

Abstract

Exposure measurement error can result in a biased estimate of the association between an exposure and outcome. When the exposure–outcome relationship is linear on the appropriate scale (e.g. linear, logistic) and the measurement error is classical, that is the result of random noise, the result is attenuation of the effect. When the relationship is non-linear, measurement error distorts the true shape of the association. Regression calibration is a commonly used method for correcting for measurement error, in which each individual's unknown true exposure in the outcome regression model is replaced by its expectation conditional on the error-prone measure and any fully measured covariates. Regression calibration is simple to execute when the exposure is untransformed in the linear predictor of the outcome regression model, but less straightforward when non-linear transformations of the exposure are used. We describe a method for applying regression calibration in models in which a non-linear association is modelled by transforming the exposure using a fractional polynomial model. It is shown that taking a Bayesian estimation approach is advantageous. By use of Markov chain Monte Carlo algorithms, one can sample from the distribution of the true exposure for each individual. Transformations of the sampled values can then be performed directly and used to find the expectation of the transformed exposure required for regression calibration. A simulation study shows that the proposed approach performs well. We apply the method to investigate the relationship between usual alcohol intake and subsequent all-cause mortality using an error model that adjusts for the episodic nature of alcohol consumption.

Correspondence: Dr. Christen M. Gray, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, United Kingdom., cmgray@alum.mit.edu.

SUPPORTING INFORMATION

Additional supporting information including source code to reproduce the results may be found online in the Supporting Information section at the end of the article.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

Keywords

Bayesian modelling; episodic consumers model; fractional polynomials; measurement error; regression calibration

1 | INTRODUCTION

Measurement error in continuous exposures results in biased estimates of exposure-outcome associations and loss of power to detect associations. There is a large literature on methods for mitigating the effects of exposure measurement error in regression analyses (Buonaccorsi, 2010; Carroll, Ruppert, Stefanski, & Crainiceanu, 2006; Fuller, 1987; Gustafson, 2004), with the most commonly used method being regression calibration (Armstrong, 1985; Kuha, 1994; Rosner et al., 1990; Rosner, Willett, & Spiegelman, 1989). Regression calibration corrects for measurement error by replacing the unobserved true exposure in the outcome regression model with its expectation conditional on the observed error-prone exposure and any fully measured covariates. This method has been well described for settings in which the exposure is entered in the model as a linear term and has also been extended to when the exposure is entered in the model as a quadratic term (Carroll et al., 2006; Strawbridge, 2011). It is often of interest to explore flexible non-linear transformations of exposures, which can be done using the fractional polynomial method, a powerful approach to polynomial selection (Royston & Sauerbrei, 2008; Royston, Ambler, & Sauerbrei, 1999). However, to date, very little has been published to extend measurement error correction methods to models using fractional polynomials (Strawbridge, 2011).

The aim of this article is to describe how regression calibration can be used to correct for the effects of measurement error in a continuous exposure modelled using fractional polynomials and to apply the approach in a setting with complex measurement error.

Royston and Altman (1994) first presented the fractional polynomial method for selection of the best polynomial transformation of a continuous predictor in regression modelling. It has since become used in the field of epidemiology as an alternative to using simple polynomial transformations chosen a priori or categorisation of a continuous predictor (Abbas et al., 2008; Cao, Moineddin, Urquia, Razak, & Ray, 2014; Takagi & Umemoto, 2013). The practical application of the method was later fully laid out in a book dedicated to multi-variable model building (Royston & Sauerbrei, 2008). The fractional polynomial method selects a set of power transformations X^p for exposure X , with p usually chosen from the set $S = \{-3, -2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ ($p = 0$ denotes the log-transformation) in an effort to create more parsimonious and interpretable models. The method does so by methodically fitting models with different power transformations, also varying the number of power transformations included, and comparing the models using deviances. To apply regression calibration in models using fractional polynomials, each power transformation of the unobserved true exposure X^p must be replaced with its expectation conditional on the error-prone measure and non-error prone covariates. We show how to achieve this using a Bayesian approach.

The work is motivated by an application to investigate the association between usual alcohol intake and all-cause mortality, which has been postulated to be non-linear (Di Castelnuovo et al., 2006; Knott, Coombs, Stamatakis, & Biddulph, 2015; Stockwell et al., 2016), using data from the EPIC-Norfolk study (Day et al., 1999; Riboli et al., 2002). In this study, alcohol intake was measured in up to three 7-day diaries (7DD) taken between 0 and 6 years from enrolment into the study. The outcome is a binary representation of all-cause mortality within 15 years from the baseline measurement. The observed measures of alcohol intake are subject to classical measurement error as well as the problem of excess zero measurements due to ‘episodic consumers’ of alcohol whose intake is not captured in the short-term diet diary. We show how the regression calibration approach for fractional polynomials can be used with classical measurement error in a simulation study. We extend this in the application for use with the ‘episodic consumers (EC) model’ for measurement error (Kipnis et al., 2009, 2016, Zhang et al., 2011a; Tooze et al., 2006; b). The EC model has been fitted using both maximum likelihood (Kipnis et al., 2009, 2016) and Bayesian methods (Zhang et al., 2011a; b) but has not previously been used in regression calibration with fractional polynomials.

This paper is organised as follows. In Section 2, we outline the exposure-outcome model of interest and the fractional polynomial method. In Section 3, we describe our methods for measurement error correction. The methods are assessed using a simulation study (Section 4) and applied to study the association between alcohol and mortality in the EPIC-Norfolk study (Section 5). In Section 6, we discuss the relevance of our findings and future directions.

2 | THE EXPOSURE-OUTCOME MODEL & THE FRACTIONAL POLYNOMIAL METHOD

We begin by outlining the target model for the association between the true but unobserved usual intake of alcohol, denoted X_i ($i = 1, \dots, N$ for a study of N individuals), and a binary outcome Y_i , for example indicating mortality over a specified follow-up period, adjusted for confounding variables $\mathbf{Z}$. We assume the logistic model

$$\text{logit}(P(Y_i = 1 | X_i, Z_i)) = \beta_0 + \psi_m(X_i; \mathbf{p}, \boldsymbol{\beta}_X) + \boldsymbol{\beta}_Z \mathbf{Z}_i, \quad (1)$$

where $\psi_m(X_i; \mathbf{p}, \boldsymbol{\beta}_X)$ represents a fractional polynomial transformation of X_i of degree m and with powers $\mathbf{p}$ and associated regression coefficients $\boldsymbol{\beta}_X$ (Royston & Sauerbrei, 2008). It is assumed that X_i is positive and continuous. The methods described apply similarly for continuous outcomes modelled using linear regression.

For a first degree fractional polynomial ($m = 1$), FP1, $\mathbf{p}$ is a single power, typically selected from the set of $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, where 0 indicates the log-transformation. The FP1 transformation is defined as

$$\psi_1(X_i; p, \beta_x) = \beta_x h(X_i, p), \text{ where } h(X_i, p) = \begin{cases} X_i^p, & p \neq 0 \\ \log(X_i), & p = 0. \end{cases} \quad (2)$$

In the degree 2 case ($m = 2$), FP2, two powers are selected from the set of S and the fractional polynomial is defined as

$$\psi_2(X_i; \mathbf{p}, \beta_x) = \beta_{x,1} h(X_i, p_1) + \beta_{x,2} h'(X_i, p_1, p_2), p_1 \leq p_2 \quad (3)$$

$$\text{where } h'(X_i, p_1, p_2) = \begin{cases} h(X_i, p_2) & p_1 \neq p_2 \\ h(X_i, p_1) \log(X_i), & p_1 = p_2. \end{cases}$$

For the FP1 case, the outcome model, Equation (1), is fit to each possible FP1 model, that is for every $p \in S$. The transformation giving the lowest deviance is selected if and only if the deviance difference between that model and the linear model, and the null model is statistically significant, given a chosen α -level, when the deviance difference is compared to the χ^2 distribution with appropriate degrees of freedom as outlined by Royston and Sauerbrei (2008). Otherwise the linear model or null model is selected.

For the FP2 case, polynomial models are fitted with every combination of p_1 and p_2 ($p_1 \leq p_2$) from the set of S , resulting in 28 combinations, where equal powers and zero powers are handled as in (3). The best fitting FP2 model is defined as that with the lowest deviance and is compared with the best fitting FP1 model. The FP2 model is selected over the FP1 model if there is a statistically significant lowering of the deviance based on a chosen α -level. Typically, the degree of the FP is $m = 3$ (Royston & Sauerbrei, 2008). This combination allows for a broad range of curves while still creating a parsimonious model. In this application, we will use FP2 models.

3 | MEASUREMENT ERROR CORRECTION USING REGRESSION CALIBRATION

3.1 | Regression calibration

In this section, after defining classical measurement error, we describe how the regression calibration method for measurement error correction can be used together with fractional polynomial transformations of error-prone exposures.

Instead of observing the true exposure, X , we observe a measure of X which is subject to measurement error, R . The measurement for individual i ($i = 1, \dots, N$) on occasion j ($j = 1, \dots, J$) is denoted R_{ij} and $\mathbf{R}_i$ is the vector of all observed measurements. A common scenario is that, while all or most individuals have the first measurement R_{i1} observed, only a subset has

replicate measurements. In our simulation study $J=2$ and in the application $J=3$. Under an assumption of classical measurement error for R_{ij} we have

$$R_{ij} = X_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, N, j = 1, 2, \dots, J, \quad (4)$$

where ε denotes a random error term which is normally distributed with zero mean and constant variance, σ_ε^2 . This model assumes there is no correlation between errors in replicate measures, $\text{corr}(\varepsilon_{i1}, \varepsilon_{i2}) = 0$, and that the errors are independent of the covariates $\mathbf{Z}_i$ and the outcome Y_i (i.e. non-differential error). Under these assumptions, R_{ij} is unbiased for X_i , that is $E(X_i) = E(R_{ij})$ and the observed variance of R_{ij} would be the sum of σ_X^2 and σ_ε^2 .

Consider first a logistic outcome model with only a linear term for the main exposure of interest X_i : $\text{logit}(P(Y_i = 1 | X_i, \mathbf{Z}_i)) = \beta_0 + \beta_X X_i + \beta_Z \mathbf{Z}_i$. The naïve approach is to use the error-prone measure, or the mean of several repeated error-prone measures, in place of the true exposure. For example, letting $\bar{R}_i$ denote the mean of up to three 7DDs for each individual, the naïve analysis is based on the model $\text{logit}(P(Y_i = 1 | \mathbf{R}_i, \mathbf{Z}_i)) = \beta_0^* + \beta_X^* \bar{R}_i + \beta_Z^* \mathbf{Z}_i$. Estimates of β_X^* are biased for the parameter of interest β_X , and the other parameters in the model are also affected.

In regression calibration, we replace X_i in the outcome model with $E(X_i | \mathbf{R}_i, \mathbf{Z}_i)$, which is obtained on the basis of an assumed error model. Regression calibration uses the result that $E(Y_i | \mathbf{R}_i, \mathbf{Z}_i) = E(E(Y_i | \mathbf{R}_i, X_i, \mathbf{Z}_i) | \mathbf{R}_i, \mathbf{Z}_i) = E(E(Y_i | X_i, \mathbf{Z}_i) | \mathbf{R}_i, \mathbf{Z}_i)$, where the second step uses the assumption that error in R_{ij} is non-differential error. First consider a simple setting without any FP transformations of the exposure and where the outcome model of interest is a linear regression: $E(Y_i | X_i, \mathbf{Z}_i) = \beta_0 + \beta_X X_i + \beta_Z \mathbf{Z}_i$. Using the above results, we have $E(Y_i | \mathbf{R}_i, \mathbf{Z}_i) = \beta_0 + \beta_X E(X_i | \mathbf{R}_i, \mathbf{Z}_i) + \beta_Z \mathbf{Z}_i$, where β_X in this is the same as the β_X in the outcome model of interest. Regression calibration works exactly for linear regression but also approximately for logistic regression when the rare disease assumption is met, and either the coefficient of the true X or the ratio of the measurement error variance to the variance of X is small (Keogh & White, 2014; Rosner et al., 1989; Rosner, Spiegelman, & Willett, 1992). In the setting of simple classical error, $E(X_i | \mathbf{R}_i, \mathbf{Z}_i, \boldsymbol{\theta})$ (where $\boldsymbol{\theta}$ denotes the parameters of a specified model for the conditional expectation) can be estimated from a regression of R_{i2} on R_{i1} and $\mathbf{Z}_i$. In a more complex measurement error setting, it may be possible to estimate the expectation by maximum likelihood. Below we propose an alternative Bayesian approach.

First, we consider the extension of regression calibration to the setting in which the exposure is transformed using fractional polynomials. This requires estimation of the expectations $E\{h(X_i, p) | \mathbf{R}_i, \mathbf{Z}_i, \boldsymbol{\theta}\}$ and $E\{h(X_i, p) \log(X) | \mathbf{R}_i, \mathbf{Z}_i, \boldsymbol{\theta}\}$, where $h(X_i, p)$ denotes the fractional polynomial transformation with powers p . We use these expectations in $\psi_m(X_i; p, \beta_X)$, the

additive contribution to the linear predictor of the outcome regression as defined in (2) and (3). For example, for an FP2 model regression calibration involves fitting the model

$$\text{logit}(P(Y_i = 1 | \mathbf{R}_i, \mathbf{Z}_i)) = \beta_0 + \beta_{X,1} E(h(X_i, p_1) | \mathbf{R}_i, \mathbf{Z}_i) + \beta_{X,2} E(h'(X_i, p_1, p_2) | \mathbf{R}_i, \mathbf{Z}_i) + \beta_Z \mathbf{Z}_i. \quad (5)$$

For the classical error model, the expectations needed could be estimated by maximum likelihood under additional distributional assumptions about X , for example conditional normality. However, this is not in general straightforward and may require a separate numerical integration for each expectation (8 for FP1 and 16 for FP2), in particular for more complex types of error as found in our application in Section 5. Using a Bayesian approach offers a simpler solution as we now explain.

3.2 | A hybrid approach using Bayesian analysis and regression calibration

We first outline our proposed ‘MCMC-RC’ approach in the context of the substantive model without any FP transformations of X , before extending to the FP setting.

In the first step of this approach (the MCMC step), a Bayesian model is used for the latent exposures X_i given $\mathbf{Z}_i$ and the measurement error model. In the measurement error context, the unobserved exposures X_i are treated as parameters in a Bayesian analysis and the posterior is $f(X_i, \theta | \mathbf{R}_i, \mathbf{Z}_i) \propto f(\theta) f(\mathbf{R}_i, \mathbf{Z}_i | X_i, \theta)$ where $f(\theta)$ represents the prior distribution of the parameters (reflecting prior knowledge about the parameter values). A large number of samples, $\tilde{X}_{ik} (k = 1, \dots, K)$, is drawn from the posterior distribution $f(X_i, \theta | \mathbf{R}_i, \mathbf{Z}_i)$ using a long MCMC chain. We propose that the mean of the posterior draws, denoted $\hat{E}(\tilde{X}_i | \mathbf{R}_i, \mathbf{Z}_i, \theta)$, is then used as an estimate of $E(X_i | \mathbf{R}_i, \mathbf{Z}_i, \theta)$ in regression calibration. This is the second step of MCMC-RC. Details on priors are given in the context of the simulation study in Section 4.

This approach extends in a straightforward way to the situation in which the outcome model contains FP transformations of X . The posterior draws of $\tilde{X}_i$ can be transformed using $h(\tilde{X}_i, p)$ and $E(h(X_i, p) | \mathbf{R}_i, \mathbf{Z}_i, \theta)$ can be estimated using the mean of $h(\tilde{X}_i, p)$ over many posterior draws $\tilde{X}_{ik} (k = 1, \dots, K)$, that is $\tilde{h}(\tilde{X}_i, p) = \hat{E}(h(\tilde{X}_i, p) | \mathbf{R}_i, \mathbf{Z}_i, \theta) = (1/K) \sum_{k=1}^K h(\tilde{X}_{ik}, p)$. the expectations are obtained separately for each value of p , though the Bayesian analysis is only performed once. Having obtained the $\tilde{h}(\tilde{X}_i, p)$ for all relevant powers p , the fractional polynomial model selection procedure can be implemented as described in Section 2. However, the model selection procedure has to be implemented ‘by hand’ because the procedure for obtaining $\psi_m(\tilde{X}_i; \hat{p}, \hat{\beta}_X)$ means that automated software for fractional polynomials model selection, such as the ‘mfp’ package in R (Ambler & Benner, 2015), cannot be used.

Available software for running MCMC on Bayesian models, such as WinBUGS or JAGS (Lunn, Thomas, Best, & Spiegelhalter, 2000; Plummer, 2013) makes the application of the

Bayesian part of the MCMC-RC analysis relatively straightforward, and we provide example R and JAGS code in the Supporting Information.

4 | SIMULATION STUDY

4.1 | Data generation

We performed a simulation study to evaluate the performance of the MCMC-RC approach for measurement error correction, in comparison to analyses using the true X and the naïve analysis. Performance was assessed primarily in terms of the bias in the estimated curve of association, but we also considered variability. The metrics used to compare the methods are described in Section 4.2.

We assumed a sample size of $N = 20,000$. The latent exposure, X , was generated from a normal distribution with mean 20 and variance 100 and truncated at zero. This distribution of the exposure ensured that the fractional polynomial method could accurately detect the J-curve when using the latent X . We simulated a binary outcome, Y , using the latent X , based on a fractional polynomial model of degree 2 (FP2) for X with powers $\mathbf{p} = (-0.5, 1)$. The model powers and regression coefficients were chosen based on the corrected fit of the application to alcohol and mortality in Section 5. The model used to generate Y was

$$\text{Logit}(P(Y = 1)) = -1.8 + 0.67 \times X^{-0.5} + 0.010 \times X. \quad (6)$$

The intercept was chosen so that 20% of individuals have $Y = 1$. An error-prone observed exposure R_{11} was generated for all simulated individuals assuming classical error as in Equation (4), with measurement error variance $\sigma_e^2 = 25$ (i.e. 25% of the variance of X). A second error-prone exposure R_{12} was generated for a random set of 25% of the simulated population. One hundred simulated data sets were generated.

Source code to reproduce the results is available as the Supporting Information on the journal's webpage (<http://onlinelibrary.wiley.com/doi/xxx/supinfo>).

4.2 | Methods

For each simulation, we applied the fractional polynomial method to the, typically unobserved, latent X as the reference method. A naïve analysis was performed using $\bar{R}_i$, the mean of measures R_i , in the fractional polynomial method in place of X_i . We also apply the MCMC-RC method using $\psi_m(\tilde{X}_i; \mathbf{p}, \beta_X)$ as described in Section 3. For each of these three methods, we retained the best FP1 model and the best FP2 model as well as the p-value of the χ^2 distribution comparing their deviances to indicate the model chosen by the FP selection method.

In the Bayesian analysis, the exposure model for the latent X was the normal distribution with a mean μ_X and variance σ_X^2 , truncated at 0.0001 to avoid zero or negative values. A Gaussian prior with mean zero and variance 1,000 was stipulated for μ_X . The variance

parameters, σ_X^2 and σ_ϵ^2 , were assigned inverse Gamma priors, which take only positive values and are conjugate to the Gaussian distribution, parameterized with shape and scale parameters of 1, that is $IG(1, 1)$. To create samples from the distribution of each X_j we used 10,000 cycles to ‘burn-in’ then an additional 20,000 samples were drawn for inference keeping only 1 every 50 to reduce the burden on computer memory (referred to as thinning).

We obtained the average log odds curve for each method across all simulations, by finding the mean log odds at each exposure point across an exposure range of 1 to 60. For each simulation, the best fitting FP1 or FP2 model was selected depending on the p-value associated with the comparison between the deviances. Linear or null models were not considered in the selection process. The methods were assessed visually and compared using several metrics. The estimated association curves obtained under each method are presented graphically. The mean integrated difference between the true log odds curve (Equation (6)) and the predicted mean log odds curve for each method M is estimated as follows:

$$\hat{\delta} = \frac{1}{T} \sum_{t=1}^T |\widehat{f_{X(t)}}_M - f_{X(t)M}| \quad (7)$$

where $f_X(t)$ denotes the true log odds at exposure level $X = t$ and $\widehat{f_{X(t)}}_{t,M}$ denotes its estimate under method M , and the sum is over T support points. We used support points across an exposure range of 1 to 60 in steps of 0.01.

We also looked at the powers selected among FP1 and FP2 models and the proportion of simulated data sets in which an FP2 model was selected as the best fitting model. We also obtained the exposure value at which the log odds curve was at a minimum and the estimated OR comparing exposure $X = 1$ to exposure $X = 10$.

4.3 | Simulation results

Across 100 simulations, the average posterior mean of the sampled measurement error variance was 18.4. The average posterior mean of the expectation and variance of the latent X was 20.1 and 101.6, respectively. The Geweke diagnostic was run on the parameters for each simulation to test stationarity (Raftery & Lewis, 1995); 85 simulations had p-values of > 0.05 for all three parameters and 96 simulations had p-values of > 0.01 for all parameters.

The average log odds curve across 100 simulations is depicted in Figure 1 for each method. As expected, using the latent X recovers the true underlying curve. The average naïve curve is much flatter at both the lower exposure values and the higher exposure values. The 95% range of the empirical distribution of the mean log odds curve across simulations show that the MCMC-RC method results in much greater variability particularly in the lower range of the exposure.

The $\hat{\delta}$ comparing the mean latent X curve to the true curve is 0.013 while that from the naïve analysis is significantly larger at 0.054 (Table 1A). The mean MCMC-RC curve is much closer to the true curve with a $\hat{\delta}$ of 0.025.

The exposure value at the lowest point of the true curve, which we refer to as the nadir, is at 10.39. The nadir of the mean curve was 10.22 for the latent X and 9.85 for the naïve analysis. The nadir for the mean MCMC-RC curve was close to the true value at 10.59. The true odds ratio (OR) comparing the odds of mortality at exposure $X = 1$ to exposure $X = 10$ is 1.45. The estimated ORs from the mean curves are 1.38 from the latent X analysis, 1.05 from the naïve analysis and 1.15 from the MCMC-RC analysis.

Figure 2 shows the estimated mean curve for each method based on the best fitting fractional polynomial of degree $m = 1$, FP1, and based on the best fitting FP2 model. Using the selection method in Section 2, the FP2 model was selected using the latent X in 86 simulations, using the naïve analysis in 28 simulations, and using MCMC-RC in 20 simulations (Table 1). It is important to note that while the average best fitting FP1 curves, including that based on the latent X , have difficulty fitting the “J-shaped” curve, the average best fitting FP2 curves are all much closer to the true curve.

The $\hat{\delta}$ values based on the average best fitting FP2 curves were 0.007, 0.034 and 0.016, from the latent X , naïve, and MCMC-RC analyses, respectively (Table 1A). Of note, when assuming an FP2 model, the estimated OR comparing $X = 1$ to $X = 10$ from the mean MCMC-RC curve (1.41) is very similar to that from the analysis using the latent X (1.40).

Table 1B shows the mean and standard deviation for the mean integrated difference, the nadir and the OR across all simulations. These summary statistics capture the variability in the per simulation results. When assuming an FP2 model, the MCMC-RC method on average recovers an OR very similar to the latent X , but has far more variability as the SD of the latent X OR is 0.18 and the SD of the MCMC-RC OR is 0.42. The naïve method is on average most biased, but has an SD of only 0.11. With the FP selection criteria applied, the variability remains largely the same, but the estimated mean OR is much lower for MCMC-RC, 1.20, and slightly lower for the naïve method, 1.06, and the same for the latent X , 1.40.

While the linear and null models were not included in our selection process, it is important to note the loss of power to detect any non-linear effect using both the naïve analysis and the MCMC-RC analysis. The linear model would have been rejected in favour of a non-linear model in only 7 (of 100) simulations for MCMC-RC and in 16 for the naïve analysis. The null model would have been rejected in 88 simulations for MCMC-RC but only in 51 in the naïve analysis. In all simulations using the latent X both the linear and null models were rejected. This demonstrates the loss of power that results from exposure measurement error.

5 | APPLICATION TO ALCOHOL AND MORTALITY

5.1 | EPIC-Norfolk cohort

The EPIC-Norfolk study recruited 25,636 eligible individuals, aged 39–79 years, between 1993 and 1998 from the population of Norfolk, UK. as a part of the European Prospective Investigation into Cancer and Nutrition (Bingham et al., 2001). Alcohol intake was recorded using 7DD and food frequency questionnaires (FFQ), at a baseline health check shortly after study recruitment and, in subsets of individuals, at several follow-up times. FFQs asked participants to recall how frequently they consumed alcohol over the past year. The study

and methods of dietary assessment have been described in detail elsewhere (Bingham et al., 2001; Day et al., 1999; Welch et al., 2001). We focus on deaths from all causes observed during the first 15 years of follow-up. The following variables were considered as potential confounders and were identified from previous studies: age, sex, smoking status, social class and education level (Table 2; Ferrari et al., 2014; Knott et al., 2015; Thun et al., 1997). Ethnicity was not included as a potential confounder as less than 0.5% of the EPIC-Norfolk population were non-white. The mean of 1–2 recorded FFQ values, and the baseline body mass index (BMI) was used as predictive covariates in the measurement error model in addition to the potential confounders from the outcome model.

We excluded 3,588 individuals who had no FFQ recorded, had missing baseline values (baseline 7DD, outcome model covariates or predictive covariates), or recorded less than 7 days in their baseline 7DD (see the Supporting Information for further detail). We further excluded 2,570 individuals who reported prior serious illness, including cancer, diabetes, stroke or heart attack, which may change alcohol intake behaviour (i.e. abstinence or reduced drinking) while representing a higher risk of mortality (Ferrari et al., 2014; Khaw et al., 2008). Analyses are based on 19,478 individuals. In the first 15 years of follow-up, there were a total of 3,085 deaths.

Table 2A displays a summary of the covariates used to predict the usual alcohol intake and the potential confounders in the outcome regression model. The EPIC-Norfolk study population used for analysis had a higher proportion of women than men across an age spectrum from 39 to 79 at baseline. Only 11% were current smokers, while 41% were previous smokers. The average BMI was 26, but the range for BMI observed varied widely from 15 to 59. More than half the population achieved at least A-levels, while roughly a third had no educational attainment. Potentially due to the longer recall time of the FFQ, the percentage of alcohol consumers was higher compared to the 7DD; however, the mean and median usual alcohol intake was lower than those from the 7DD observations. This may be due to participants difficulty in estimating an average consumption of an episodic food such as alcohol which might have a wide variety in portion size per drinking occasion over a year.

The distributions of alcohol intake observed in the three 7DDs are summarized in Table 2B along with the number reporting zero intake for the week of the 7DD. At baseline, 26% of individuals reported zero intake in their 7DD with a median intake of 5.7 g/day. Only 87% identified themselves as consumers of alcohol (either by being observed to drink alcohol on any 7DD or on any FFQ); among these, 15% reported zero intake in their 7DD with a median intake of 7.9 g/day. The proportion observed not to drink was slightly higher in follow-up 7DDs 1 and 2 (31% and 29%, respectively), as was the median intake (4.3 and 4.6 g/day, respectively). These observations were preserved when exclusively describing consumers of alcohol.

5.2 | The episodic consumers model

Measures of alcohol intake from short-term assessments such as the 7-day diet diary are known to be subject to excess zero measurements due to episodic consumers, as well as other sources of measurement error. We therefore used a measurement error model which account for episodic consumers (EC). The EC model (Keogh & White, 2011; Kipnis et al.,

2009; Zhang et al., 2011a, 2011b) extends the classical model to accommodate excess zero measurements in addition to classical measurement error. The model for R_{ij} is split into two components. The first component is a model for the probability that alcohol was consumed during the week of the 7DD, and the second is a model for the amount consumed conditional on consumption having occurred. The model assumes that the true usual alcohol intake is non-zero. We model the first component using a logistic model and the second component using a normal model. The two parts of the model are linked by a vector of person-specific random effects, $\mathbf{u}_i = (u_{1i}, u_{2i})$:

$$P(R_{ij} > 0 | \mathbf{u}_i, \mathbf{C}_i) = H(\gamma_{1,0} + \gamma_{1,C} \mathbf{C}_i + u_{1i}) \quad (8)$$

$$R_{ij} | (R_{ij} > 0), \mathbf{u}_i, \mathbf{C}_i \sim N(\gamma_{2,0} + \gamma_{2,C} \mathbf{C}_i + u_{2i}, \sigma_\varepsilon^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2), \quad (9)$$

where $H(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$, γ_1 and γ_2 are vectors of regression coefficients, ε_{ij} is random error and $\mathbf{C}_i$ is the vector of covariates that are predictive of consumption. The covariates, $\mathbf{C}_i$, need not be identical for both Equations (8) and (9), but typically they are and we assume that here. $\mathbf{C}_i$ generally includes covariates in the outcome regression model as well as additional covariates related to X_i given $\mathbf{R}_i$ (Kipnis et al., 2009). It is assumed that the measurement errors, ε_{ij} , have constant variance, σ_ε^2 , are uncorrelated with each other and are uncorrelated with $\mathbf{u}_i$. It makes sense that the person-specific random effects for the probability and the amount of consumption may be correlated. We therefore model the distribution of $\mathbf{u}_i$ as a bivariate normal distribution with variances σ_{u1}^2 and σ_{u2}^2 and correlation ρ :

$$\mathbf{u}_i \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u1}^2 & \rho \sigma_{u1} \sigma_{u2} \\ \rho \sigma_{u1} \sigma_{u2} & \sigma_{u2}^2 \end{pmatrix} \right) \quad (10)$$

Under the assumption that R_{ij} is an unbiased measure of X_i , as in the classical error model, the true usual intake can then be modelled as the product of the probability of consumption and the expectation of the amount of consumption given consumption occurred:

$$X_i = E(R_{ij} | \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}) = P(R_{ij} > 0 | \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}) E(R_{ij} | R_{ij} > 0, \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}), \quad (11)$$

where $\boldsymbol{\theta} = \{\gamma_1, \gamma_2, \sigma_{u1}^2, \sigma_{u2}^2, \rho, \sigma_\varepsilon^2\}$ is the set of all model parameters. The EC model as shown in Equations (8) and (9) assumes that $R_{ij} | R_{ij} > 0$ is normally distributed; sometimes a transformation may be required to make this assumption plausible. This can be performed using a Box-Cox transformation (Box & Cox, 1964), $g(x, \lambda) = (x^\lambda - 1)/\lambda$, where $\lambda = 0$ denotes

the log transformation (Keogh & White, 2011; Kipnis et al., 2009; Tooze et al., 2006). The Box–Cox transformed non-zero measures are denoted $R_{ij}^* = g(R_{ij}, \lambda)$. Maintaining the assumption that the R_{ij} are unbiased for X_i on the original scale gives the modified model for X_i of

$$X_i = E(R_{ij} | u_i, C_i; \theta) = P(R_{ij} > 0 | u_i, C_i; \theta) E\{g^{-1}(R_{ij}^*, \hat{\lambda}) | R_{ij} > 0, u_i, C_i; \theta\}, \quad (12)$$

where $g^{-1}(R_{ij}^*, \hat{\lambda})$ indicates the inverse Box–Cox transformation for the selected value of $\hat{\lambda}$ used for the transformation in the first place. The fitted values based on model (11) or (12) are $\hat{X}_i(\theta) = E(X_i | R_i, C_i, \theta)$, and these form the basis of the measurement error correction methods to be described in the next section. Kipnis et al. (2009) and Keogh and White (2011) used maximum likelihood estimation to estimate the parameters of the EC model and hence to obtain $\hat{X}_i(\hat{\theta})$. The maximum likelihood-based method has been found to suffer problems with convergence and can therefore be unreliable in finding solutions for all data sets meeting the stated criteria (Bhadra et al., 2017) due to lack of empirical identifiability in a finite sample set.

The EC model can alternatively be fitted using a Bayesian approach (Bhadra et al., 2017; Zhang et al., 2011a). We used JAGS (Plummer, 2013) via R package ‘rjags’ (Plummer, 2015) to implement the MCMC sampling from the EC model; code for the model can be found in the Supporting Information. Previous implementations have used MATLAB (MATLAB and Statistics Toolbox, The MathWorks, Inc; Bhadra et al., 2017; Zhang et al., 2011a).

5.3 | Fitting the EC model

The EC model was fitted to the 7DD measures of alcohol intake, using up to three repeated 7DD measures per individual using a Bayesian approach. All included individuals had a baseline measure of alcohol intake.

As the non-zero observed measures are skewed, a Box–Cox transformation was used to transform them to be approximately normally distributed (Box & Cox, 1964). The best transformation parameter, $\hat{\lambda}$, for the Box–Cox transformation was found by finding the value of λ that maximizes the profile log-likelihood (using the ‘boxcox’ function from the MASS package in R). The estimated transformation parameter for the observed measures of alcohol intake was $\hat{\lambda} = 0.27$. The Box-Cox transformed measures were then scaled by their standard deviation ($SD = 2.23$).

The continuous EC model covariates (age, baseline BMI and mean FFQ), where deemed necessary, were also transformed using a Box-Cox transformation ($\hat{\lambda}_{FFQ} = 0.08$, $\hat{\lambda}_{age} = 0.45$, and $\hat{\lambda}_{BMI} = 0$), centered and then scaled in order to improve efficiency in the Metropolis–Hastings algorithm (Kruschke, 2010). We used the same priors as used by Zhang et al. (2011a) and Bhadra et al. (2017); further detail can be found in the appendices to those papers. All γ coefficients in the EC model were given normal priors

with mean zero and variance 100. Due to the scaling of R_{ij} , the standard deviation σ_ε was given a uniform prior between 0 and 3. This differs from the inverse Gamma prior used for the variances in the simulation study; the inverse gamma prior is commonly used for variances and allows for an infinite positive range of values unlike the uniform prior, hence we selected to use that in the simulation study where our focus was not on the EC model (see Section 4.2).

The covariance matrix for the person-specific random effect, u_i , was given an inverse Wishart prior, which is a multivariate generalization of the scaled χ^2 -distribution and is conjugate to the multivariate normal distribution (Alvarez, Niemi, & Simpson, 2014). The inverse Wishart prior is parameterized by the degrees of freedom, df , and an expectation matrix. Specifying the df as greater than more than the matrix order (i.e. for an $m \times m$ matrix, $df > m + 1$) results in a prior concentration around zero for the correlation between the random effects (i.e. no correlation). The higher the df specified, the stronger the belief in independent variances. Specifying $df = m$ results in a prior concentration around $\rho = 1$ and -1 . In this application, we specified five degrees of freedom (for a matrix of order 2), and for the expectation matrix, we specified an expectation of the variances on the diagonal equal to 1 and the correlation equal to 0.5 (Bhadra et al., 2017; Zhang et al., 2011a).

Approximately 250,000 burn-in samples were required before the MCMC sampling reached the stationary distribution as determined by the Geweke diagnostic (Raftery & Lewis, 1995). We collected an additional 500,000 samples after burn-in on three chains for inference, keeping only every 200th sample to reduce burden on computer memory.

The posterior expected values of the EC model parameters γ_1 and γ_2 can be found in Table 3. Additionally, the posterior means for ρ , σ_ε^2 , $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ within the EC model are 0.937, 0.435, 7.426 and 0.175, respectively. As would be expected, an increase in the mean FFQ (reported in gram/day) was associated with a large increase in probability of consumption and with a notably higher amount consumed. An increase in BMI or age at baseline was both associated with decreased probability of consumption. However, an increase in BMI at baseline was associated with a slight increase in amount consumed while an increase in age at baseline was associated with a slight decrease in amount consumed. Those who had never smoked had the lowest probability of drinking and lower consumption when they do drink when compared to either former or current smokers. Former smokers eclipsed current smokers in likelihood of drinking, but had a lower amount consumed when they did drink. Greater educational attainment as well as more skilled work was both associated with a higher probability of drinking and greater amounts consumed.

The sampled densities of X_i for the first nine individuals, that is $i = 1$ to 9, are shown in the Supporting Information. The densities of the posterior means of the sampled values of $\tilde{X}_i$, compared to the naive distribution of the mean observed alcohol intake ($\bar{R}_i$), across this study population can be found in the Supporting Information, and summary values of the distributions are presented in Table 2. The densities of posterior means of the samples appear to be quite distinct different from the naive distribution. The peak of the distribution of

$\hat{E}(\tilde{X}_i | R_i, C_i, \theta)$ was observed to be higher than for $\bar{R}_i$, which is expected because individuals with low usual consumption are likely to have reported zero consumption in the 7DD. The minimum $\hat{E}(\tilde{X}_i | R_i, C_i, \theta)$ for any individual was 0.9 g/day, while the maximum was 142.8 g/day. Due to this minimum, a product of assuming all members of our population have some probability of drinking, we can make inference down to only approximately 1 g/day.

5.4 | Fractional polynomial analysis using regression calibration

We performed two analyses. In the first, the adjustment covariates used in the outcome model were limited to sex and age (the ‘minimally adjusted outcome model’). In the second, we adjusted for additional potential confounders, smoking status, social class and educational attainment (i.e. the ‘fully adjusted outcome model’). The naïve analysis and MCMC-RC were both applied.

It is a known property of the fractional polynomial method that outlier covariate values at either extreme can be high leverage points (Royston & Sauerbrei, 2008). Therefore, 16 individuals with posterior expected values over 90 g/day were excluded from the primary analysis (both naïve and MCMC-RC). Results without these exclusions are shown in the Supporting Information.

For the naïve analysis, the fractional polynomial method was performed using the ‘mfp’ package available in R (Ambler & Benner, 2015; R Core Team, 2016). The naïve measure, $\bar{R}_i$, includes zeros, and the fractional polynomial method cannot directly accommodate non-positive data points; therefore, the method adds the constant 0.1 to every individual’s measure in order to make all points positive (Royston & Sauerbrei, 2008). Use of this method of accommodating zeros is generally unsatisfactory as the choice of the constant added can have large influence on the subsequent choice of transformations and therefore the shape of the curve. This is not an issue in the MCMC-RC analysis, because all of the posterior sampled values $\tilde{X}_i$ are positive.

Both methods found that the best-fitting FP1 curve had power $\hat{p} = -2$ for both minimally adjusted and fully adjusted models and provided a better fit than both the linear and null models ($p\text{-value} < 10^{-8}$). Both methods also found that the best-fitting FP2 model provided a better fit than the best-fitting FP1 model, but they differed on the powers selected in the FP2 model (Table 4). For both the minimally adjusted and fully adjusted models, the naïve method selected $\hat{p} = \{0, 0.5\}$ and MCMC-RC selected $\hat{p} = \{0, 0.5\}$ for the minimally adjusted and $\hat{p} = \{-0.5, 1\}$ for the fully adjusted curve. The resulting curves are shown in Figure 3. It is possible for different FP2 polynomial transformations to represent nearly identical curves, and all the curves have in common the characteristic J-shape of a risk which first falls sharply then rises again more slowly. The confidence intervals associated with the selected model fit by each method are included in the Supporting Information. These confidence intervals do not incorporate the uncertainty due to measurement error or model uncertainty.

The naïve curve is the flatter across the range of alcohol intake (Figure 3). This is what we expect based on the simulation study results. The higher risk associated with lower alcohol

consumption is tempered in the naïve model compared to the MCMC-RC measurement error corrected curve. The higher risk of mortality as associated with higher alcohol intake is also modest compared to the MCMC-RC curve. The nadir at which the risk is lowest is 6.2 g/day for the naïve minimally adjusted model and 6.8 g/day for the fully adjusted model while the nadirs for the MCMC-RC curves are 9.6 g/day and 10.1 g/day, respectively. All had odds ratios (ORs) with elevated risk for those with usual alcohol intake of 1 g/day in comparison to those with 8 g/day (approximately equivalent to 1 unit of alcohol). But these ORs were much more exaggerated after correction for measurement error with MCMC-RC (Table 4).

There were no great differences in the estimated coefficients for the covariates between the naïve and MCMC-RC methods selected models (Table 3).

6 | DISCUSSION

The fractional polynomial method is becoming a popular tool in epidemiological and other studies in which measurement error in exposures is common. We therefore identified a need for methods to correct for this error when exposures are modelled using fractional polynomials. The STRATOS initiative (STREngthening Analytical Thinking for Observational Studies) in 2013 identified the need not just to develop methods for measurement error correction or selection of correct functional forms, but for the successful integration of these areas (Sauerbrei et al., 2014). We believe this is a step towards such integration.

In this paper, we have described an approach for correcting for measurement error by combining regression calibration, the standard method for measurement error correction, with the power of Bayesian MCMC. The method uses a two-stage approach: In the first stage, a measurement error model is fitted and used to obtain estimates of the expectations of fractional polynomial transformations, that is $E(h(X_i, p) | \mathbf{R}_i, \mathbf{Z}_i, \boldsymbol{\theta})$ in the second stage the expectations are used in a standard fractional polynomial model selection procedure. The method was described for use with classical measurement error and the more complex episodic consumers model. However, it could be applied for other measurement error models, for example systematic error, provided that the information needed to identify the measurement error model parameters is available, for example via a validation study. The maximum likelihood method is arguably more labour intensive, particularly in the context of more complex measurement error because it requires a series of separate numerical integrations to obtain the expectations needed for regression calibration. The Bayesian regression calibration method relies on the same basic principles as standard regression calibration, that is replacing the latent variable for each individual with its expectation (Equation (5)). Therefore, when extended to logistic regression and the Cox proportional hazards model, it is subject to the same assumptions as standard regression calibration.

The simulation study demonstrated that for a J-shaped association between the continuous exposure and outcome, the proposed approach recovers, on average, an estimated curve which is much closer to the true curve than that obtained if the measurement error is ignored. When an FP2 model was assumed, the average MCMC-RC curve is nearly identical to the simulated latent X curve generated by the standard fractional polynomial method.

When assuming an FP2 model, the OR between an exposure value of 1 and 10 was shown to be nearly the same (1.41 for MCMC-RC and 1.40 for the latent X) whereas the naive OR was biased towards the null (1.15). The MCMC-RC curve fits resulted in much greater variability than the latent X or naïve $\bar{R}$ curve fits particularly near zero, suggesting the need for a more informative data set as for very low exposure values.

Unfortunately, in the setting explored in our simulation study where relatively few replicate measures are available non-linearity had to be assumed as the power to detect a non-linear model was not recovered. In their book on the topic, Royston and Sauerbrei list insufficient power and sensitivity to extreme values as the greatest disadvantage of the fractional polynomial method (Royston & Sauerbrei, 2008). Measurement error also inherently decreases power, and while bias can be corrected, power cannot be recovered (Carroll et al., 2006). In the face of such uncertainty, very high sample sizes are required. Use of less stringent criteria for choosing between the FP models of different degrees may be appropriate when there is measurement error and is an area for further investigation.

The MCMC-RC method was motivated by an application to investigate the association between alcohol intake and all-cause mortality in the EPIC-Norfolk cohort, in which 7DD measures of alcohol intake are subject to measurement error due to episodic consumers. However, the method may be extended to any non-linear association with a defined model for the measurement error in the exposure. The application of the proposed method recovered a stronger and more non-linear association between alcohol intake and all-cause mortality, compared with the naive approach.

MCMC-RC has potential due to its flexibility, but given the loss of power and lack of efficiency inherent in regression calibration, other methods may build upon this work. Additionally, the expectations used within the models will not retain the original variable relationship, that is for $p = \{1, 2\}$, $E(X^2) \neq E(X)^2$, which may impact model selection. A fully Bayesian model which combines both measurement error correction and model selection will be poorly identifiable in the settings outlined in this paper; this remains an area for further research. Cut point models as presented by Plummer may have the potential to accommodate measurement error and model selection (Plummer, 2014).

In the EPIC-Norfolk data set, most of the observed exposures were at the lower end of the range of alcohol consumption. It is therefore difficult to make inference about the association between drinking large quantities of alcohol and risk of 15-year mortality. Because our model did not allow for a usual consumption of zero, we also cannot make inference about non-drinkers. We hoped to understand better the relationship between very low to moderately high alcohol consumption and where the minimum risk lies. Because of the skewed distribution of the exposure, providing more data around the lower extreme of the association curve, the application had more power to reject linear models than did our simulations. The level at which risk is lowest was found to be just under 1 unit per day (where the standard UK unit of alcohol is 8 g/day) by the naïve model and just over 1 unit per day by the MCMC-RC model. Di Castelnuovo et al. (2006) found that the range of one to two drinks per day in women and two to four drinks per day in men had an inverse association with all-cause mortality. Bergmann et al. examined alcohol consumption in

relation to cause of death and found that those drinking light to moderate amounts had the lowest risk of death from cardiovascular disease, more so than very light users (Bergmann et al., 2013). Although the J-shaped association between alcohol and all-cause mortality remains a topic of much debate (Fillmore, Stockwell, Chikritzhs, Bostrom, & Kerr, 2007). We adjusted for a number of potential confounders in the analysis, including socio-economic indicators; it is possible that there is residual confounding that may in part account for the raised risk of all cause-mortality with very low alcohol intake.

A limitation with regard to the application was the exclusion of missing data; further work is needed to handle measurement error correction and missing data simultaneously in this context. An extension to survival analysis would also be of great interest. Finally, in this paper, the final model does not take into account either the measurement error correction or the model selection procedure. A more accurate variance could in theory be generated by bootstrapping, but in practice would be unwieldy in combination with the MCMC and model selection procedure. In practice, the additional variability due to estimation of the measurement error model is typically small relative to the errors of estimation in the outcome model. In the application presented in this paper, the model selection and the outcome logistic regression are the greatest source of variance. Our simulation study was based around the application. In further work, our investigations may be extended to consider further simulation scenarios.

An extension to the EC model has been developed to include non-drinkers (Bhadra et al., 2017; Keogh & White, 2011). The EC model assumes that every participant has some non-zero usual intake. Depending upon the population, this assumption may be too restrictive. In this application, a high proportion (87%) of participants either said they drank alcohol in one or more FFQ or were observed to drink alcohol in one or more 7DDs. A comparison using the extended model incorporating non-drinkers would be of interest, though in previous applications identifying never consumers has proved difficult (Jayasekara, English, Room, & MacInnis, 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

CMG was supported by a Economic and Social Research Council doctoral studentship. RHK was supported by a Medical Research Council Fellowship (MR/M014827/1). RJC was supported by the U.S. National Cancer Institute (U01-CA057030). The authors would like to acknowledge the contribution of the staff and participants of the EPIC-Norfolk Study. EPIC-Norfolk is supported by the Medical Research Council programme grants (G0401527, G1000143) and Cancer Research UK programme grants (C864/A8257, C864/A14136).

Funding information Economic and Social Research Council; National Cancer Institute, Grant/Award Number: U01-CA057030; Medical Research Council, Grant/Award Numbers: G0401527, G1000143, MR/M014827/1; Cancer Research UK, Grant/Award Numbers: C864/A8257, C864/A14136

REFERENCES

- Abbas S, Linseisen J, Slinger T, Kropp S, Mutschelknauss EJ, Flesch-Janys D, & Chang-Claude J (2008). Serum 25-hydroxyvitamin D and risk of post-menopausal breast cancer—Results of a large case-control study. *Carcinogenesis*, 29, 93–99. [PubMed: 17974532]
- Alvarez I, Niemi J, & Simpson M (2014). Bayesian inference for a covariance matrix. *Annual Conference on Applied Statistics in Agriculture*, 26, 71–82.
- Ambler G, & Benner A (2015). mfp: Multivariable fractional polynomials. R package version 1.5.2, 1–10.
- Armstrong B (1985). Measurement error in the generalised linear model. *Communications in Statistics—Simulation and Computation*, 14, 529–544.
- Bergmann MM, Rehm J, Klipstein-Grobusch K, Boeing H, Schütze M, Drogan D, ... Ferrari P (2013). The association of pattern of lifetime alcohol use and cause of death in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *International Journal of Epidemiology*, 42, 1772–1790. [PubMed: 24415611]
- Bhadra A, Wei R, Lilly E, Keogh R, Buckman DW, & Carroll RJ (2017). Measurement Error Models with Zero Inflation and Hard Zeros, with Applications to Never-Consumers in Nutrition.
- Bingham S, Welch A, McTaggart A, Mulligan A, Runswick S, Luben R, ... Day NE (2001). Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutrition*, 4, 847. [PubMed: 11415493]
- Box GEP, & Cox DR (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological)*, 26, 211–252.
- Buonaccorsi JP (2010). *Measurement error: Models, methods, and applications*. Boca Raton, FL: Chapman & Hall/CRC.
- Cao S, Moineddin R, Urquia ML, Razak F, & Ray JG (2014). J-shapedness: An often missed, often miscalculated relation: The example of weight and mortality. *Journal of Epidemiology and Community Health*, 68, 683–690. [PubMed: 24683176]
- Carroll RJ, Ruppert D, Stefanski LA, & Crainiceanu CM (2006). *Measurement Error in Nonlinear Models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Di Castelnuovo A, Costanzo S, Bagnardi V, Donati MB, Iacoviello L, & de Gaetano G (2006). Alcohol dosing and total mortality in men and women. *Archives of Internal Medicine*, 166, 2437–2445. [PubMed: 17159008]
- Day N, Oakes S, Luben R, Khaw KT, Bingham S. a, Welch A, & Wareham N (1999). EPIC-Norfolk: Study design and characteristics of the cohort. *European Prospective Investigation of Cancer. British Journal of Cancer*, 80(Suppl 1), 95–103. [PubMed: 10466767]
- Ferrari P, Licaj I, Muller DC, Kragh Andersen P, Johansson M, Boeing H, ... Romieu I (2014). Lifetime alcohol use and overall and cause-specific mortality in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *BMJ Open*, 4, e005245.
- Fillmore KM, Stockwell T, Chikritzhs T, Bostrom A, & Kerr W (2007). Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies and new hypotheses. *Annals of Epidemiology*, 17, S16–S23. [PubMed: 17478320]
- Fuller WA (1987). *Measurement error models*. New York, NY: Wiley.
- Gustafson P (2004). *Measurement error and misclassification in statistics and epidemiology: Impacts and Bayesian adjustments*. Boca Raton, FL: Chapman & Hall/CRC.
- Jayasekara H, English DR, Room R, & MacInnis RJ (2014). Alcohol consumption over time and risk of death: A systematic review and meta-analysis. *American Journal of Epidemiology*, 179, 1049–1059. [PubMed: 24670372]
- Keogh RH, & White IR (2011). Allowing for never and episodic consumers when correcting for error in food record measurements of dietary intake. *Biostatistics*, 12, 624–636. [PubMed: 21378386]
- Keogh RH, & White IR (2014). A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Statistics in Medicine*, 33, 1–21.

- Khaw K-T, Wareham N, Bingham S, Welch A, Luben R, & Day N (2008). Combined impact of health behaviours and mortality in men and women: The EPIC-Norfolk prospective population study. *PLoS Medicine*, 5, e12. [PubMed: 18184033]
- Kipnis V, Freedman LS, Carroll RJ, & Midthune D (2016). A bivariate measurement error model for semicontinuous and continuous variables: Application to nutritional epidemiology. *Biometrics*, 72, 106–115. [PubMed: 26332011]
- Kipnis V, Midthune D, Buckman DW, Dodd KW, Patricia M, Krebs-smith SM, ... Freedman LS (2009). Modeling data with excess zeros and measurement error: Application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*, 65, 1003–1010. [PubMed: 19302405]
- Knott CS, Coombs N, Stamatakis E, & Biddulph JP (2015). All cause mortality and the case for age specific alcohol consumption guidelines: Pooled analyses of up to 10 population based cohorts. *BMJ*, 350, h384. [PubMed: 25670624]
- Kruschke JK (2010). *Doing Bayesian data analysis: A tutorial with R and BUGS*. New York, NY: Academic Press.
- Kuha J (1994). Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Statistics in Medicine*, 13, 1135–1148.
- Lunn DJ, Thomas A, Best N, & Spiegelhalter D (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, 325–337.
- Plummer M (2013). Just another Gibbs sampler (JAGS), version 3.4.0.
- Plummer M (2014). Cuts in Bayesian graphical models. *Statistics and Computing*, 25, 37–43.
- Plummer M (2015). rjags: Bayesian graphical models using MCMC. R package version 4–6.
- R Core Team (R Foundation for Statistical Computing) (2016). *R: A language and environment for statistical computing*.
- Raftery AE, & Lewis SM (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. *Practical Markov Chain Monte Carlo*, 7, 763–773.
- Riboli E, Hunt K, Slimani N, Ferrari P, Norat T, Fahey M, ... Saracci R. (2002). European Prospective Investigation into Cancer and Nutrition (EPIC): Study populations and data collection. *Public Health Nutrition*, 5, 1113–1124. [PubMed: 12639222]
- Rosner B, Spiegelman D, & Willett WC (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American Journal of Epidemiology*, 136, 1400–1413. [PubMed: 1488967]
- Rosner B, Spiegelman D, & Willett WC (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *American Journal of Epidemiology*, 132, 734–745. [PubMed: 2403114]
- Rosner B, Willett WC, & Spiegelman D (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, 8, 1051–1069; discussion 1071–1073. [PubMed: 2799131]
- Royston P & Altman DG (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat*, 43, 429–467.
- Royston P, Ambler G, & Sauerbrei W (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology*, 28, 964–974.
- Royston P, & Sauerbrei W (2008). *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. New York, NY: Wiley.
- Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S, Carpenter J, Abrahamowicz M,... Woodward M. (2014). STRENGTHENING analytical thinking for observational studies: The STRATOS initiative. *Statistics in Medicine*, 33, 5413–5432. [PubMed: 25074480]
- Stockwell T, Zhao J, Panwar S, Roemer A, Naimi T, & Chikritzhs T (2016). Do “moderate” drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. *Journal of Studies on Alcohol and Drugs*, 77, 185–198. [PubMed: 26997174]
- Strawbridge AD (2011). *Modelling Non-Linear Exposure-disease Relationships in a Large Individual Participant Meta-Analysis Allowing for the Effects of Exposure Measurement Error*. University of Cambridge.

- Takagi H, & Umemoto T (2013). Limit to benefits of large reductions in low-density lipoprotein cholesterol levels: Use of fractional polynomials to assess the effect of low-density lipoprotein cholesterol level reduction in metaregression of large statin randomized trials. *JAMA Internal Medicine*, 173, 1028–1029. [PubMed: 23700132]
- Thun MJ, Peto R, Lopez AD, Monaco JH, Henley SJ, Heath CW, Doll R (1997). Alcohol consumption and mortality among middle-aged and elderly U.S. adults. *The New England Journal of Medicine*, 337, 1705–1714. [PubMed: 9392695]
- Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, ... Kipnis V. (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Dietetic Association*, 106, 1575–1587. [PubMed: 17000190]
- Welch AA, McTaggart A, Mulligan AA, Luben R, Walker N, Khaw KT, ... Bingham SA. (2001). DINER (Data Into Nutrients for Epidemiological Research) - a new data-entry program for nutritional analysis in the EPIC-Norfolk cohort and the 7-day diary method. *Public Health Nutrition*, 4, 1253–1265. [PubMed: 11796089]
- Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V, ... Carroll RJ. (2011a). Fitting abivariate measurement error model for episodically consumed dietary components. *The International Journal of Biostatistics*, 7, 1–32.
- Zhang S, Midthune D, Guenther PM, Krebs-Smith SM, Kipnis V, Dodd KW, ... Carroll RJ. (2011b). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *Annals of Applied Statistics*, 5, 1456–1487. [PubMed: 21804910]

FIGURE 1.

Average log odds curves from 100 simulated data sets (solid lines) with 95% range of the empirical distribution across simulations (dashed lines). The gray curve included in each plot is the true curve used to generate the data

FIGURE 2.

Average log odds curves from 100 simulated data sets given different selection criteria

FIGURE 3.

Estimated association between usual alcohol intake and 15-year mortality in the EPIC-Norfolk data based on the best fitting model selected by the FP method. Left hand plot: adjusted for sex and age (depicted for male, age 50 years). Right hand plots: adjusted for sex, age, smoking status, social class and education level (depicted for male, age 50 years, never smoked, in non-skilled work and no education). The top figures depict the log-odds curve, and the dashed lines depict the 95% confidence interval bounds based on the final model chosen in each case. The bottom figures depict the OR compared with 8 g/day (1 unit of alcohol)

TABLE 1

Summary statistics for 100 simulated data sets with classical measurement error where the FP method was applied using the latent X , the naïve analysis and the MCMC-RC error-correction method. Mean-integrated difference compares the curve from each method to true curve for values of X from 1 to 60 in steps of 0.01. Nadir is the $\hat{X}$ value at the most minimal predicted logodds value. OR compares $X = 1$ to $X = 10$. SD = standard deviation

	Latent X	Naïve $\bar{R}$	MCMC-RC
A. Description of the mean curve			
FP selection criteria applied			
Mean integrated difference (X range 1 – 60)	0.013	0.054	0.025
Nadir	10.2	9.9	10.6
True value = 10.39			
Odds ratio	1.38	1.05	1.15
True value = 1.45			
FP2 models only			
Mean integrated difference (X range 1 – 60)	0.007	0.034	0.016
Nadir	10.1	13.7	11.8
True value = 10.39			
Odds ratio	1.40	1.15	1.41
True value = 1.45			
B. Summary statistics across all simulations			
FP selection criteria applied			
# FP2 models selected	86	28	20
Average mean integrated difference (X range 1 – 60)	0.053 (0.035)	0.068 (0.031)	0.065 (0.37)
Mean nadir (SD), True value = 10.39	17.7(18.1)	11.9(19.0)	10.4(19.1)
Mean odds ratio (SD) True value = 1.45	1.40(0.18)	1.06 (0.13)	1.20 (0.42)
FP2 models only			
Average mean integrated difference (X range 1 – 60)	0.045 (0.027)	0.054 (0.034)	0.057 (0.035)
Mean nadir (SD), True value = 10.39	11.8 (9.0)	19.0 (17.6)	16.1 (16.6)
Mean odds ratio (SD) comparing $X = 1$ to $X = 10$ True value = 1.45	1.41 (0.18)	1.15 (0.11)	1.46 (0.42)

TABLE 2

Descriptive statistics for A) the covariates in the EPIC-Norfolk population; and B) the error-prone exposures among those identified as usual consumers of alcohol and in the whole study population as well as the summary statistics for the measures used in the naive and MCMC-RC methods

A) Covariates (C)	Number	Percentage of total				
Sex						
Women	10,851					55.7
Men	8,627					44.3
Smoking status						
Current smoker	2,145					11.0
Previous smoker	8,019					41.2
Never smoked	9,314					47.8
Social class						
Non-skilled work	672					3.5
Semi-skilled work	2,586					13.3
Skilled work (manual or non manual)	7,738					39.7
Manager	7,134					36.6
Professional	1,348					6.9
Educational attainment						
None	6,930					35.6
O-levels	2,066					10.6
A-levels	7,956					40.8
Bachelors degree or higher	2,526					13.0
	Median	Mean	IQR	Range		
Mean alcohol intake FFQ (3,345 zeros, 17%)	4.7	8.7	0.8–10.9	0–139.6		
Baseline age (years)	58	59	51–66	39–79		
Baseline BMI (kg/m ²)	26	26	24–28	15–59		
B) Error-prone measures and estimates of usual alcohol intake	No. observed (% of total)	No. not consuming alcohol in 7DD (% of those observed)	Median (g/day)	Mean (g/day)	IQR (g/day)	Range (g/day)
Consumers Only						
Baseline 7DD	16,902 (87%)	2561 (15%)	7.9	13.9	2.0–19.6	0–247.4
Follow-up 1 7DD	1975 (87%)	408 (21%)	6.1	11.3	0.9–15.9	0–114.0
Follow-up 2 7DD	2219 (89%)	459 (21%)	6.2	11.3	0.8–16.0	0–123.3
ALL						
Baseline 7DD	19,478	5,137 (26%)	5.7	12.1	0–17.2	0–247.4
Follow-up 1 7DD	2,280	713 (31%)	4.3	9.8	0–13.9	0–114.0
Follow-up 2 7DD	2,489	729 (29%)	4.6	10.1	0–14.4	0–123.3
Naive estimate ($\bar{R}$)	19,478	4,756 (24%)	5.8	12.1	0–17.1	0–247.4
Fitted $E(\tilde{X} \mathbf{R}, \mathbf{C}, \boldsymbol{\theta})$	19,478	NA	6.1	11.2	1.3–16.0	0.9–142.8

TABLE 3

Regression coefficients for (1) the predictive covariates within the fitted EC model (posterior means) and (2) the non-error prone covariates in the selected minimally adjusted and fully adjusted outcome models fitted using the naïve analysis and the MCMC-RC method

		EC model				Minimally adjusted outcome model		Fully adjusted outcome model	
		Logistic regression for probability of consumption		Logistic regression for amount consumed		Naïve	MCMC-RC	Naïve	MCMC-RC
Intercept		$\gamma_{1,0}$	3.598	$\gamma_{2,0}$	1.166	β_0	-4.471	-4.351	-3.848
Age (Age ³ for the outcome models)		$\gamma_{1,3}$	-0.237	$\gamma_{2,3}$	-0.043	β_{Z1}	0.114 ***	0.114 ***	0.116 ***
Sex	Male	ref		ref		ref			
	Female	$\gamma_{1,4}$	0.077	$\gamma_{2,4}$	-0.051	β_{Z2}	-0.570 ***	-0.582 ***	-0.527 ***
Smoking status	Current smoker	ref		ref					
	Former smoker	$\gamma_{1,5}$	0.152	$\gamma_{2,5}$	-0.064	β_{Z3}		-0.711 ***	-0.710 ***
	Never smoker	$\gamma_{1,6}$	-0.123	$\gamma_{2,6}$	-0.133	β_{Z4}		-0.965 ***	-0.973 ***
Education	No qualifications	ref		ref					
	O-levels	$\gamma_{1,7}$	0.202	$\gamma_{2,7}$	0.017	β_{Z5}		-0.013	-0.009
	A-levels	$\gamma_{1,8}$	0.314	$\gamma_{2,8}$	0.036	β_{Z6}		-0.067	-0.065
	Bachelors degree or higher	$\gamma_{1,9}$	0.283	$\gamma_{2,9}$	0.048	β_{Z7}		-0.131	-0.127
Social class	Non-skilled work	ref		ref					
	Semi-skilled work	$\gamma_{1,10}$	0.116	$\gamma_{2,10}$	-0.001	β_{Z8}		-0.253 *	-0.253 *
	Skilled work (manual or non-manual)	$\gamma_{1,11}$	0.283	$\gamma_{2,11}$	0.003	β_{Z9}		-0.181	-0.180
	Manager	$\gamma_{1,12}$	0.596	$\gamma_{2,12}$	0.074	β_{Z10}		-0.236 *	-0.235 *
	Professional	$\gamma_{1,13}$	0.541	$\gamma_{2,13}$	0.057	β_{Z11}		-0.494 ***	-0.488 **
Mean FFQ		$\gamma_{1,1}$	4.499	$\gamma_{2,1}$	0.873				
Baseline BMI		$\gamma_{1,2}$	-0.048	$\gamma_{2,2}$	0.013				

*** indicates p-value < 0.001

** p-value < 0.01

* p-value < 0.05.

TABLE 4

Results on the association between alcohol intake and all-cause mortality in EPIC-Norfolk. Best-fitting curves: the p-value comparing the deviance of the best-fitting FP1 model to the best-fitting FP2 model, the values of $\hat{p}$ defining the chosen transformations, the coefficients of the transformed latent true usual alcohol intake values, the exposure value at the nadir of the curve and the odds ratio (OR) comparing 1 g/day with 8 g/day (1 unit of alcohol)

		FP2 vs FP1 p-value	$\hat{p}_1$	$\hat{p}_2$	$\hat{\beta}_0$	$\hat{\beta}_{X,1}$	$\hat{\beta}_{X,2}$	Nadir (g/day)	OR
Naive model, $\bar{R}$	Minimally adjusted	0.0019	0	0.5	-4.84	-0.16	0.13	6.3	1.11
	Fully adjusted	0.0192	0	0.5	-3.85	-0.15	0.11	6.9	1.11
MCMC-RC, $\psi_m(\tilde{X}_i; \hat{p}, \hat{\beta}_X)$	Minimally adjusted	0.0003	0	0.5	-4.35	-0.42	0.27	9.6	1.46
	Fully adjusted	0.0085	-0.5	1	-3.83	0.67	0.010	10.1	1.43