

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



Gray, CM (2018) Use of the Bayesian family of methods to correct for effects of exposure measurement error in polynomial regression models. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: <https://doi.org/10.17037/PUBS.04649757>

Downloaded from: <http://researchonline.lshtm.ac.uk/4649757/>

DOI: [10.17037/PUBS.04649757](https://doi.org/10.17037/PUBS.04649757)

#### Usage Guidelines

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

LONDON  
SCHOOL *of*  
HYGIENE  
& TROPICAL  
MEDICINE



# Use of the Bayesian family of methods to correct for effects of exposure measurement error in polynomial regression models

Christen M. Gray

Thesis submitted in accordance with the requirements for the degree of Doctor of  
Philosophy of the University of London

July 2018

**Department of Medical Statistics  
Faculty of Epidemiology and Population Health  
London School of Hygiene & Tropical Medicine**

Funded by the Economic and Social Research Council

I, Christen Michelle Gray, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

A handwritten signature in black ink, appearing to read "Christen Michelle Gray". The signature is written in a cursive style with a large initial 'C' and a checkmark-like flourish at the end.

## ABSTRACT

---

Measurement error in a continuous exposure, if ignored, may cause bias in the estimation of the relationship between exposure and outcome. This presents a significant challenge for understanding exposure-outcome associations in many areas of research, including economic, social, medical and epidemiological research. The presence of classical, i.e. random, measurement error in a continuous exposure has been shown to lead to underestimation of a simple linear relationship. When the functional form of the exposure within a regression model is not linear, i.e. when transformations of the exposure are included, measurement error obscures the true shape of the relationship by making the association appear more linear. Bias in this case will be unknown in direction and vary by exposure level.

The most commonly used method for measurement error correction is regression calibration, but this requires an approximation for logistic and survival regression models and does not extend easily to more complex error models. This work investigates three methods for measurement error correction from the Bayesian family of methods: Bayesian analysis using Markov chain Monte Carlo (MCMC), integrated nested Laplace approximations (INLA), and multiple imputation (MI). These have been proposed for measurement error correction but have not been extensively compared, extended for use in several important scenarios, or applied to flexible parametric models. The focus on Bayesian methods was motivated by their flexibility to accommodate complex measurement error models and non-linear exposure-outcome associations.

Polynomial regression models are widely used and are often the most interpretable models. In order for measurement error correction methods to be widely implemented, they should be able to accommodate known polynomial transformations as well as model selection procedures when the functional form of the error-prone exposure is unknown. Therefore, in this thesis, correction methods are integrated with the fractional polynomial method, a flexible polynomial model-building procedure for positive continuous variables.

In this thesis, I perform a large simulation study comparing proposed methods for measurement error correction from the Bayesian family (i.e. MCMC, INLA, and MI) to the most common method of measurement error correction. Extensions of INLA and MI are presented in order to accommodate both a validation study setting wherein the error-free exposure is measured in a subgroup as well as a replicate study setting wherein there are multiple measures of the error-prone exposure. In order to accommodate unknown polynomial transformations of the error-prone variable, two approaches not used before in this context are proposed and explored in simulation studies alongside more standard methods. The first approach uses Bayesian posterior means in lieu of maximum likelihood estimates within regression calibration. The second approach adapts

methods of Bayesian variable selection to the selection of the best polynomial transformation of the error-prone exposure while accommodating measurement error. Successful methods are applied to a motivating example, fitting the non-linear association between alcohol intake and all-cause mortality.

By combining measurement error correction adaptable to complex error models with polynomial regression inclusive of model-selection, this work fills a niche which will facilitate wider use of measurement error correction techniques.

# TABLE OF CONTENTS

---

Abstract.....	3
Acknowledgements.....	11
Abbreviations.....	12
List of Tables .....	14
List of Figures.....	16
1 Introduction.....	20
1.1 Motivation.....	20
1.2 Chapter aims .....	21
1.3 The substantive model .....	21
1.3.1 Generalized linear models.....	21
1.3.2 Generalized additive models.....	22
1.3.3 The fractional polynomial method.....	22
1.4 The measurement error model .....	24
1.4.1 Classical error model .....	24
1.5 The effects of measurement error .....	25
1.5.1 Illustration of classical measurement error in an untransformed predictor.....	26
1.5.2 Illustration of effect of measurement error in a transformed exposure.....	27
1.6 Validation and replicate studies.....	30
1.7 Thesis aims.....	31
1.8 Literature overview .....	32
1.9 Thesis contributions .....	34
1.10 Illustrative examples .....	35
1.11 Overview of the thesis.....	36
2 Methods for measurement error correction.....	37
2.1 Aims.....	37
2.2 Regression Calibration (RC).....	37
2.2.1 Estimation of variance .....	39

2.2.2	Extensions to non-linear functional forms of the predictor.....	40
2.2.3	Software .....	40
2.3	Bayesian analysis - overview .....	40
2.3.1	Exposure model.....	42
2.4	Markov chain Monte Carlo (MCMC) sampling .....	42
2.4.1	Sampling and convergence diagnostics .....	44
2.4.2	Estimation of variance .....	45
2.4.3	Extensions to non-linear functional forms of the predictor.....	45
2.4.4	Software .....	46
2.5	Integrated nested Laplace approximations (INLA).....	46
2.5.1	Estimation of variance .....	47
2.5.2	Software .....	48
2.6	Multiple imputation (MI) .....	49
2.6.1	Application of multiple imputation to missing data.....	49
2.6.2	Application of multiple imputation to measurement error.....	51
2.7	Summary .....	54
3	Measurement error correction in generalized linear models when the exposure is untransformed .....	55
3.1	Aims & Overview .....	55
3.2	Simulation study design .....	55
3.3	Criteria for evaluating fit.....	56
3.4	Implementation of methods .....	57
3.5	Simulation study results .....	58
3.5.1	Naïve analysis .....	58
3.5.2	Validation study .....	58
3.5.3	Replicate study.....	64
3.5.4	Computing time.....	70
3.6	Application in the Framingham Heart Study .....	72
3.7	Summary .....	73

4	Measurement error correction for a quadratic transformation of the error-prone exposure	76
4.1	Aims & Overview	76
4.2	Quadratic shapes	76
4.3	Naïve analysis and the impact of measurement error	77
4.4	Correction methods for a quadratic model	79
4.4.1	Regression calibration (RC)	79
4.4.2	Bayesian analysis using MCMC	80
4.4.3	INLA	80
4.4.4	Bayesian regression calibration	81
4.4.5	Multiple imputation	82
4.5	Simulation study with a continuous outcome	84
4.5.1	Simulation study design	84
4.5.2	Criteria for evaluating quadratic fit	84
4.5.3	Method implementation	86
4.5.4	Fitting the latent $X$ and naïve models	87
4.5.5	Results: Model fit after the application of measurement error correction methods	95
4.6	Simulation study with a binary outcome	96
4.6.1	Simulation study design and evaluation criteria	96
4.6.2	Implementation of methods	97
4.6.3	Fitting the latent $X$ and naïve models	97
4.6.4	Results: Model fit after application of measurement error correction methods	98
4.7	Summary	99
5	Measurement error correction for selection of a quadratic term	101
5.1	Aims & overview	101
5.2	Hypothesis testing	101
5.3	Measures of predictive accuracy (information criteria)	102
5.4	Bayesian transformation selection (BTS)	103
5.4.1	Kuo and Mallick method applied to Bayesian variable selection	104



5.4.2	Kuo and Mallick method applied to selection of the quadratic model (Bayesian transformation selection).....	105
5.5	Simulation study extension to model selection for a continuous outcome .....	106
5.5.1	Simulation study design .....	106
5.5.2	Criteria for evaluating power and type I error .....	107
5.5.3	Method implementation .....	108
5.5.4	Model selection using the latent $\mathbf{X}$ .....	109
5.5.5	Model selection using the naïve analysis .....	109
5.5.6	Results: Model selection using hypothesis testing.....	110
5.5.7	Results: Model selection using AIC/DIC.....	112
5.5.8	Results: Model selection using BTS .....	113
5.6	Simulation study extension to model selection for a binary outcome.....	115
5.6.1	Simulation study design and evaluation criteria .....	115
5.6.2	Method implementation .....	116
5.6.3	Model selection using the latent $\mathbf{X}$ .....	116
5.6.4	Model selection using the naïve analysis .....	117
5.6.5	Results: Model selection using hypothesis testing.....	117
5.6.6	Results: Model selection using AIC/DIC.....	118
5.6.7	Results: Model selection using BTS .....	118
5.7	Summary .....	119
6	Measurement error correction for polynomial model selection using the fractional polynomial method .....	122
6.1	Aims & overview .....	122
6.2	Fractional polynomial method and shapes of association.....	123
6.3	Naïve analysis and the effect of measurement error .....	124
6.4	Correction methods in the context of fractional polynomials.....	125
6.4.1	Regression calibration (RC).....	125
6.4.2	Bayesian regression calibration .....	126
6.4.3	Bayesian transformation selection (BTS) .....	127
6.5	Simulations study.....	130

6.5.1	Simulation study set up .....	130
6.5.2	Criteria for evaluating fractional polynomial curve fits.....	131
6.5.3	Method implementation .....	132
6.5.4	Model selection using the latent $X$ .....	133
6.5.5	Model selection using the naïve analysis .....	144
6.5.6	Results: Model selection using RC and Bayesian RC .....	145
6.5.7	Results: Lognormal error sensitivity analysis .....	147
6.5.8	Results: Model selection using BTS .....	148
6.6	Logistic regression .....	153
6.6.1	Simulation study set up.....	153
6.6.2	Implementation of methods .....	154
6.6.3	Fitting the latent $X$ and naïve models.....	154
6.6.4	Results after application of correction methods.....	155
6.7	Summary .....	156
7	Application to alcohol and mortality in the EPIC-Norfolk study .....	158
7.1	Aims & Overview .....	158
7.2	EPIC-Norfolk cohort.....	158
7.3	The episodic consumers model .....	160
7.4	Fitting the EC model.....	162
7.5	Results: application .....	165
7.6	Summary .....	168
8	Discussion.....	170
8.1	Aims & Overview .....	170
8.2	Summary of results .....	171
8.2.1	Reference method: regression calibration.....	171
8.2.2	Use of multiple imputation for measurement error correction.....	172
8.2.3	Use of INLA for measurement error correction.....	173
8.2.4	Bayesian regression calibration .....	173

8.2.5	Use of Bayesian MCMC modelling for a known functional form of the exposure	175
8.2.6	Use of Bayesian MCMC modelling for an unknown functional form of the exposure	175
8.3	Bayesian “feedback”	177
8.3.1	Bayesian “feedback” in the measurement error correction methods presented	178
8.3.2	Future directions: Potential methods for addressing Bayesian “feedback”	179
8.4	Further extensions	180
8.4.1	Extensions to multiple covariates measured with error	180
8.4.2	Extensions to survival analysis	181
8.5	Limitations	181
8.6	Future directions	182
8.7	Conclusions	183
	Bibliography	184
A.	Code for measurement error correction methods in generalized linear models when the exposure is untransformed	196
B.	Code for measurement error correction methods with a quadratic transformation of the error-prone exposure	207
C.	Code for measurement error correction in the context of the fractional polynomial method	215
D.	Code of the Bayesian implementation of the episodic consumers model in JAGS and INLA software	227
E.	Manuscript prepared for submission to Statistics in Medicine	231
F.	Manuscript in review for The Biometrical Journal	232
G.	Additional plots of Chapter 3 simulations	233
H.	Sensitivity analysis for regression calibration with model selection included in the bootstrapped 95% confidence intervals	239

## ACKNOWLEDGEMENTS

---

First and foremost, I would like to thank my supervisor Ruth Keogh for her patience and dedication, for being so generous with her time, and for being my touchstone to reality throughout the learning process. I sincerely couldn't have asked for a better supervisor. I would also like to thank Jonathan Bartlett and Karla DiazOrdaz for their wonderful guidance and insight.

Thank you to the Economic and Social Research Council for not just funding my work but also providing an excellent support structure to accomplish it.

I am also grateful to have had so much love and support from my family and friends. There was always a kind word when I needed it, a listening ear, and understanding when I went silent and just needed to get this done. Throughout my doctoral degree, I was never alone.

I am particularly indebted to Rachel Allgaier and Jackie Britton for their help in proofreading. To Fred Smith for being my number one cheerleader. To my Dad for always being my biggest fan. And to my Mom, who is not here to see the finish line, but who has been with me all the way.

## ABBREVIATIONS

---

7DD – seven day diet diary

AIC – Akaike information criterion

BF – Bayes factor

BMI – body mass index

BTS – Bayesian transformation selection

CHD – coronary heart disease

CI – confidence interval

CrI – credible interval

DIC – deviance information criterion

EC model – episodic consumers model

ESS – effective sample size

FFQ – food frequency questionnaire

FP1 – first degree fractional polynomial

FP2 – second degree fractional polynomial

FP $m$  –  $m$  degree fractional polynomial

GAM – generalized additive model

GLM – generalized linear model

GMRF – Gaussian Markov random field

HPD – highest posterior density

INLA – integrated nested Laplace approximations

INLA-RC – Bayesian implementation of RC using INLA

IQR – interquartile range

JAV – just another variable

logOR – logarithm of the odds ratio

MCE – Monte Carlo error

MCMC – Markov chain Monte Carlo

MCMC-RC – Bayesian implementation of RC using MCMC

MI – multiple imputation

MISE – mean integrated square error

OR – odds ratio

RC – regression calibration

RDR – regression dilution ratio

RMSE – root mean square error

SBP – systolic blood pressure

SD – standard deviation

SE – standard error

SMC-FCS – substantive model compatible fully conditional specification

SSVS – stochastic search variable selection

## LIST OF TABLES

---

1.1 Models for exposure-outcome association shapes.....	27
1.2 P-values from a likelihood ratio test.....	30
3.1 Values used for simulation studies.....	56
3.2 Non-convergence rate of MCMC sampling after maximum burn-in for logistic regression.....	59
3.3 Time per single simulated data set as an average of 100 simulations (minutes) in each setting for each method.....	71
3.4 Estimate of the association between underlying systolic blood pressure (SBP) and coronary heart disease (CHD).....	73
4.1 Models for exposure-outcome association shapes with a continuous outcome.....	78
4.2 Simulations with a validation study and measurement error variance $\frac{1}{4}$ the variance of X...	89
4.3 Simulations with a validation study and measurement error variance equal to the variance of X.....	91
4.4 Simulations with a replicate study and measurement error variance equal to the variance of X.....	93
4.5 Models for exposure-outcome association shapes with a binary outcome.....	97
4.6 Simulations with a validation study and a binary outcome and measurement error variance $\frac{1}{4}$ the variance of X.....	98
5.1 Simulations to determine the power and type I error rate .....	110
5.2 Quadratic model fit from Bayesian transformation selection (BTS) samples.....	112
5.3 Bayes Factor (BF) evidence for the quadratic model using the Bayesian transformation selection (BTS) approach.....	114
5.4 Selection of the quadratic model with a weak quadratic or linear association with a binary outcome.....	117
5.5 Quadratic model fit with a binary outcome from Bayesian transformation selection (BTS) samples.....	118

5.6 Bayes Factor (BF) evidence for the quadratic model with a binary outcome using the Bayesian transformation selection (BTS) approach.....	119
6.1 Example of a results table summarizing the frequency of each model sampled using Bayesian transformation selection (BTS).....	129
6.2 Simulation model selection and fit for the J-shaped association.....	140
6.3 Simulation model selection and fit for the asymptotic association.....	141
6.4 Simulation model selection and fit for the weak quadratic association.....	142
6.5 Simulation model selection and fit for the linear association.....	143
6.6 Simulation model selection and fit when the error is lognormally related to the exposure...	148
6.7 Simulation model selection using Bayesian transformation selection (BTS) results including only fractional polynomial terms $h(X, p)$ (“FP1”).....	149
6.8 Simulation model selection using Bayesian transformation selection (BTS) results including fractional polynomial terms $h(X, p)$ and $h(X, p)\log(X)$ (“FP2”).....	150
6.9 Simulation model selection and fit for logistic regression models.....	155
7.1 Descriptive statistics for the covariates in the EPIC-Norfolk population.....	159
7.2 Descriptive statistics for the error-prone exposures in the EPIC-Norfolk population.....	160
7.3 The fitted EC model using either MCMC sampling or INLA approximation.....	163
7.4 Results on the association between alcohol intake and all-cause mortality.....	165
7.5 Regression coefficients for the non-error prone covariates.....	167



## LIST OF FIGURES

---

Box 1.1 Notation.....	21
Box 1.2 Simulated data.....	26
Figure 1.1 Simulated data demonstrating effect of classical measurement error.....	28
Figure 1.2 Simulated data demonstrating effect of classical measurement error.....	29
Box 2.1 Three-part conditional independence structure.....	41
Box 2.2 Metropolis algorithm.....	43
Box 2.3 SMC-FCS proposal distribution.....	51
Box 2.4 SMC-FCS algorithm for measurement error correction.....	52
Box 3.1 Simulation study design.....	55
Figure 3.1 Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure $X$ is observed for 30% ( $P = 0.3$ ) of the study participants.....	60
Figure 3.2 Simulation study results in the linear regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure $X$ is observed for 10% ( $P = 0.1$ ) of the study participants.....	61
Figure 3.3 Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure $X$ is observed for 30% ( $P = 0.3$ ) of the study participants.....	62
Figure 3.4 Simulation study results in the logistic regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure $X$ is observed for 10% ( $P = 0.1$ ) of the study participants.....	63
Figure 3.5 Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ).....	66
Figure 3.6 Simulation study results in the linear regression setting with sample size ( $N$ ) of 500 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ).....	67

Figure 3.7 Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ).....	68
Figure 3.8 Simulation study results in the logistic regression setting with sample size ( $N$ ) of 500 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ).....	69
Figure 4.1 Plots of exposure-outcome associations fit to the quadratic model.....	77
Figure 4.2 Demonstration of the effect on turning point bias.....	79
Box 4.1 Simulation study design.....	84
Figure 4.3 Example of simulated data for each quadratic association shape.....	85
Figure 4.4 Latent $X$ fit in simulation: the quadratic model fit to the latent $X$ .....	88
Figure 4.5 Simulations with a validation study and measurement error variance $\frac{1}{4}$ the variance of $X$ .....	90
Figure 4.6 Simulations with a validation study and measurement error variance equal to the variance of $X$ .....	92
Figure 4.7 Simulations with a replicate study and measurement error variance equal to the variance of $X$ .....	94
Box 4.2 Simulation study summary.....	99
Box 5.1 Bayes Factor (BF).....	105
Box 5.2 Simulation study design.....	107
Figure 5.1 Histograms of the posterior mean of $I_{X_2}$ .....	113
Box 5.3 Simulation study design.....	116
Figure 6.1 The chosen curve fits from the application of the fractional polynomial method with the latent $X$ and naïve analysis.....	134
Figure 6.2 Heatmaps for second degree fractional polynomial model selection for the J-shaped association.....	135
Figure 6.3 Barplots for first degree fractional polynomial model selection.....	136
Figure 6.4 The chosen curve fits from the fractional polynomial method with a validation study and classical measurement error variance $\frac{1}{4}$ the variance of $X$ .....	137

Figure 6.5 The chosen curve fits from the fractional polynomial method with a validation study and classical measurement error variance equal to the variance of $X$ .....	138
Figure 6.6 The chosen curve fits from the fractional polynomial method with a replicate study and classical measurement error variance equal to the variance of $X$ .....	139
Figure 6.7 The curve fits after application of Bayesian transformation selection (BTS) including only fractional polynomial terms $h(X, p)$ (“FP1”).....	151
Figure 6.8 Heatmap for model selection for the J-shaped association applying Bayesian transformation selection.....	152
Box 6.1 Simulation study design.....	153
Figure 7.1 Kernel density plot of empirical distribution of usual alcohol intake.....	164
Figure 7.2 Estimated association between usual alcohol intake and 15-year mortality.....	166
Figure G.1. Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure $X$ is observed for 10% ( $P = 0.1$ ) of the study participants.....	233
Figure G.2. Simulation study results in the linear regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure $X$ is observed for 30% ( $P = 0.3$ ) of the study participants.....	234
Figure G.3. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure $X$ is observed for 10% ( $P = 0.1$ ) of the study participants.....	235
Figure G.4. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure $X$ is observed for 30% ( $P = 0.3$ ) of the study participants.....	236
Figure G.5. Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for 30% ( $P = 0.3$ ) of the study participants.....	237
Figure G.6. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for 30% ( $P = 0.3$ ) of the study participants.....	238

Figure H.1 The chosen curve fits after regression calibration (RC) applied in the context of the fractional polynomial method where the 95% confidence bounds were estimated by bootstrapping either inclusive or non-inclusive of bootstrapping.....239

# 1 INTRODUCTION

---

## 1.1 MOTIVATION

Measurement error in a continuous exposure or predictor has been shown to cause bias in the estimation of exposure-outcome associations. This presents a significant challenge to estimation in many areas of research, including economic, social, medical and epidemiological research. Often the measurement error stems from measuring a fluctuating exposure at only a single time point when true interest lies in the relationship of a ‘long term average’ or ‘usual’ exposure with the outcome. Other sources of error include instrument error and self-reporting error. Examples of exposures subject to error-prone measurements are blood pressure, dietary intake, household expenditure, radiation exposure, and even biomarkers obtained from blood or urine samples.

Measurement error subject only to the addition of random error, termed classical measurement error, has been shown to lead to underestimation or attenuation of a linear relationship between a single exposure and outcome [1–3]. When measurement error is present in one or more covariates, the direction of bias in any of the regression coefficient estimates can be in either direction [4].

When the exposure has a non-linear functional form in relation to the outcome, i.e. in the linear predictor of a regression model, the effect of classical measurement error is to make the association appear more linear than it really is [1,5,6]. The result is that the risk or association effect estimate may be overestimated for some values of the exposure and underestimated for others.

It is often of interest to test for the inclusion of specific transformations of the error-prone exposure during regression model-building. Furthermore, in many applications researchers may want to explore flexible non-linear transformations of exposures, which can be done using the fractional polynomial method, a powerful approach to the selection of the best polynomial regression model [7,8]. The impact of classical measurement error within this model selection process will be to obfuscate the true shape of the association and reduce the power to detect any non-linear model [1,5,6].

The exposure of interest may be subject to more complex exposure measurement error, including systemic bias and excess zero measurements, which must be accommodated for an unbiased estimate of the association [5].

A recent systematic review found that 44% of original research articles in the top twelve medical and epidemiology journals published in 2016 reported on measurement error with 35% reporting measurement error in the primary exposure [9]. However, only 3% investigated or corrected for measurement error. The STRATOS (STRengthening Analytical Thinking for Observational

Studies) Initiative [10] attributed this phenomenon, in measurement error and other important areas, at least in part to the lack of methods or guidance on methods for integrating multiple complex aspects of analysis such as measurement error correction and model selection. In this thesis, I have proposed and explored candidate methods for the integration of these two analytical concepts.

## 1.2 CHAPTER AIMS

In this chapter, I aim to introduce the components of the problem; that is, the substantive model (i.e. the exposure-outcome model), polynomial model selection, the measurement error correction model, and additional data collection required for correction techniques.

In Section 1.3, I provide an overview of the forms used for the substantive model: generalized linear models, generalized additive models, and the fractional polynomial method, an efficient method of polynomial model selection based on generalized additive models. The classical measurement error model is detailed in Section 1.4, and its impact on the exposure-outcome association is illustrated in Section 1.5 for a single error-prone exposure with a continuous outcome. The two common methods of collecting additional data necessary for measurement error correction are introduced in Section 0.

The chapter concludes with an overview of the thesis aims, a brief literature review, thesis contributions, and the examples to be used to illustrate the methods presented in this thesis.

## 1.3 THE SUBSTANTIVE MODEL

The focus in this work is on estimating the association between a continuous exposure of interest,  $X$ , whether transformed as  $f(X)$  or untransformed, which cannot be measured directly or accurately on all study participants and an outcome  $Y$ . The association between the exposure and the outcome is assumed to be modelled using a generalized linear model (GLM) or generalized additive model (GAM), conditional on one or more covariates  $\mathbf{Z}$  which are not subject to measurement error.

### Box 1.1 Notation

$X_i$ : the exposure of interest

$Y_i$ : the outcome of interest

$W_i$ : an error-prone measure of the exposure of interest

$\mathbf{Z}_i$ : one or more accurately measured covariates in the exposure-outcome model

### 1.3.1 Generalized linear models

In Chapters 3-5, the model of interest is assumed to be a GLM with a continuous or binary outcome. In the simplest case, the exposure,  $X$ , and any accurately measured covariates,  $\mathbf{Z}$ , are assumed to have a linear effect on the expectation of the response variable via the link function  $g(\cdot)$ :

$$1.1 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_X X_i + \boldsymbol{\beta}_Z^T \mathbf{Z}_i, \quad i = 1, 2, \dots, n.$$

The model parameters include the regression parameters  $\beta_0$ ,  $\beta_X$ , and  $\boldsymbol{\beta}_Z$  and, for linear regression, the residual variance,  $\sigma_{Y|XZ}^2$ . The primary focus is on estimation of  $\beta_X$ . For linear regression this quantifies the effect of one unit of change in the exposure on the outcome, conditional on  $\mathbf{Z}$ . In the case of logistic regression,  $\beta_X$  is a log odds ratio (logOR); typically in epidemiology, the odds ratio (OR), i.e.  $\exp(\beta_X)$ , is reported.

### 1.3.2 Generalized additive models

If the model allows for a transformation of  $X$  in the linear predictor, a GAM provides more flexibility [11,12]:

$$1.2 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + f(X_i) + \boldsymbol{\beta}_Z^T \mathbf{Z}_i \quad i = 1, 2, \dots, n,$$

where  $f(X_i)$  represents the transformation of  $X_i$ . In a traditional GAM,  $f(X_i)$  may represent any smoothing function. Multiple exposures subject to smoothing functions may be included, but in this thesis, I focus on one error-prone exposure subject to smoothing. It is assumed that the functional forms of any accurately measured covariates are known, but these, too, may in theory be subject to smoothing functions. For the purposes of this thesis, the functional forms of  $X$  are restricted to polynomials. In which case,  $f(X_i)$  may represent one or more polynomial transformations of  $X_i$  combined with one or more regression coefficients equal to the degree  $m$  of the polynomial:

$$1.3 \quad f(X_i) = \sum_{k=1}^m \beta_{X_k} f_k(X).$$

For example, in a quadratic model,  $f(X_i) = \beta_{X_1} X_i + \beta_{X_2} X_i^2$ . Replacing  $f(X_i)$  in Equation 1.2, the quadratic model can then be expressed as a GLM:

$$1.4 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_{X_1} X_i + \beta_{X_2} X_i^2 + \boldsymbol{\beta}_Z^T \mathbf{Z}_i \quad i = 1, 2, \dots, n.$$

This thesis investigates cases in which the form of  $f(X_i)$  is pre-specified and cases in which the appropriate functional form is unknown while restricting the functional forms of  $X$  to polynomials. In the latter case, Equation 1.2 may be used as the basis for an algorithm for finding the optimal polynomial for  $f(X_i)$ .

### 1.3.3 The fractional polynomial method

Royston and Altman first presented the fractional polynomial method for selection of the best polynomial transformation of a continuous predictor in regression modelling [13]. It has since become widely used as an alternative to simple polynomial transformations chosen *a priori*, use of splines, or simple categorization of a continuous predictor [14–16]. The practical application of the method was later more fully detailed in a book dedicated to multivariable model-building

[8]. The fractional polynomial method selects a set of power transformations  $X^p$  for the exposure, with  $p$  usually chosen from the set  $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$  ( $p = 0$  denotes the log transformation) in an effort to create parsimonious and interpretable models. These transformations require that the exposure be limited to only positive values.

Starting from the definition of a GAM in Equations 1.2 and 1.3, a first degree fractional polynomial (FP1; i.e.  $m = 1$ ) has a single polynomial transformation of the exposure which is parameterized by  $p$ :

$$1.5 \quad f(X_i) = \beta_X h(X_i, p), \text{ where } h(X_i, p) = \begin{cases} X_i^p, & p \neq 0 \\ \log(X_i), & p = 0 \end{cases}$$

Using the standard set of  $S$  for power transformations, this represents eight possible polynomial models including the linear model (i.e.  $p = 1$ ) (Equation 1.1).

Expansion to a second degree fractional polynomial (FP2; i.e.  $m = 2$ ) follows by selecting two powers from the set of  $S$ ,  $\mathbf{p} = \{p_1, p_2\}$ , applying Equation 1.5 in a successive manner and multiplying by  $\log(X_i)$  in the second transformation if the two powers are equal:

$$1.6 \quad f(X_i) = \beta_{X_1} h(X_i, p_1) + \beta_{X_2} h'(X_i, p_1, p_2), \quad p_1 \leq p_2,$$

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

The quadratic model (Equation 1.4) can then be expressed as an FP2 model where  $\mathbf{p} = \{1, 2\}$ . For higher degree fractional polynomials (FP $m$ ), this definition is expanded in a like manner.

For the FP1 case, the substantive model, Equation 1.2, is fit using each possible FP1 transformation, i.e. for every  $p \in S$ . The fitted model with the least deviance, i.e. twice the negative log-likelihood, is selected as the best fitting FP1 model. To ascertain whether the exposure under consideration should be in the model at all, the deviance difference between the best FP1 model and the null model is compared to the  $\chi^2$ -distribution with two degrees of freedom; if the difference is statistically significant based on a chosen  $\alpha$ -level (typically  $\alpha = 0.05$ ), the term is retained. If the best FP1 model is not the linear model ( $p = 1$ ), then the deviance difference between the best FP1 model and the linear model is then compared to the  $\chi^2$ -distribution with one degree of freedom. If the deviance difference is statistically significant then the FP1 model is accepted as the chosen model (or “best” model) otherwise the linear model is considered the chosen model.

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 36 combinations, where equal powers and zero powers are handled as in Equations 1.5 and 1.6. The best fitting FP2 model is the FP2 model with the lowest deviance.



The deviance difference between the best fitting FP2 model and the best fitting FP1 model is then compared to the  $\chi^2$ -distribution with two degrees of freedom. The FP2 model is selected as the chosen model (or “best” model) over the FP1 model if the resulting test statistic is statistically significant. If chosen, the FP2 model is compared to the linear and null model as described above with three and four degrees of freedom, respectively.

Typically, the degree of the fractional polynomial is specified *a priori* as  $m \leq 3$  [8]. This combination allows for a broad range of curves while still creating a parsimonious model. In this thesis, only  $m \leq 1$  and  $m \leq 2$  models are applied.

Royston and Sauerbrei note that the biggest disadvantages of the fractional polynomial method are insufficient power to detect a nonlinear function and sensitivity to extreme values of the covariate of interest [8].

## 1.4 THE MEASUREMENT ERROR MODEL

The measurement error model relates the observed error-prone measure,  $W$ , to the latent true value. Choice of the appropriate error model depends on subject-matter knowledge, but suitability of certain model assumptions may be evaluated to some extent given the data [1].

### 1.4.1 Classical error model

A classical error model assumes that the source of the error is random (i.e. no systematic bias). It further assumes that the error is not related in any way to the outcome after conditioning on  $X$  and  $\mathbf{Z}$  ( $W \perp\!\!\!\perp Y \mid X, \mathbf{Z}$ ), i.e. non-differential [1,3,5]. In this thesis, I focus on the classical additive error model in which the error-prone measure,  $W$ , is the true predictor,  $X$ , plus some random error,  $U$ :

$$1.7 \quad W_i = X_i + U_i, \quad i = 1, 2, \dots, n.$$

This random error is assumed to be normally distributed about a mean of zero with a constant variance (homoscedastic),  $\sigma_U^2$ .  $\mathbf{W}_i$  represents a vector of the observed measures for an individual  $\{W_{i1}, \dots, W_{ij}\}$  for up to  $J$  measures. The classical error model further assumes there is no correlation between errors in replicate measures,  $\rho = \text{corr}(U_{i1}, U_{i2}) = 0$  [1,3,5]. The unknown parameters in the classical measurement error model are the mean of  $X$ ,  $\mu_X$ , the variance of  $X$ ,  $\sigma_X^2$ , and the measurement error variance,  $\sigma_U^2$ .

In Chapter 7, I use a model for episodically consumed food and drink, which extends the classical error model for application to dietary intake to accommodate the probability of observing a zero measurement [17–19].

## 1.5 THE EFFECTS OF MEASUREMENT ERROR

Use of the error prone exposure,  $W$ , in place of  $X$  in the substantive model produces a naïve estimate of the association,  $\beta_W$ :

$$1.8 \quad g(E[Y_i|W_{i1}, \mathbf{Z}_i]) = \beta_0^* + \beta_W W + \boldsymbol{\beta}_Z^{*T} \mathbf{Z}_i, \quad i = 1, 2, \dots, n.$$

Regression coefficients in the naïve model, as indicated by  $*$ , are not equal to the regression estimates from the model fit with the true exposure  $X$ . Estimates of  $\boldsymbol{\beta}_Z^*$  from the naïve model may be biased in either direction, depending on the relationship between  $X$  and  $Z$ . While the expectations of  $W$  and  $X$  are equal under the classical error model, the variance of  $W$  will always be greater. The effect of the classical error model in the presence of only accurately measured covariates is to dilute the relationship between  $X$  and  $Y$  which further impacts the power to detect the association. For linear regression when the predictor is untransformed, it can be shown that

$$1.9 \quad \beta_W = \frac{\text{cov}(Y, W|Z)}{\text{var}(W|Z)} = \frac{\text{cov}(Y, X|Z)}{\text{var}(X|Z)} \times \frac{\sigma_{X|Z}^2}{(\sigma_{X|Z}^2 + \sigma_U^2)} = \beta_X \lambda, \quad \lambda = \frac{\sigma_{X|Z}^2}{(\sigma_{X|Z}^2 + \sigma_U^2)},$$

where  $\lambda$  is referred to as the attenuation factor or regression dilution ratio (RDR). Illustration of the effect of this attenuation for linear regression can be found in Section 1.5.1.

For logistic regression, the association between  $\beta_X$  and  $\beta_W$  given in Equation 1.9 is an approximation which performs well when the disease is rare, when the OR is not large, or when the measurement error variance relative to the variance of  $X$  given  $\mathbf{Z}$  is not high [20,21]. A finer approximation which also illustrates the limitation of Equation 1.9 can be constructed by replacing the logistic function with the probit function,  $\Phi\{k(\beta_0 + \beta_X X + \boldsymbol{\beta}_Z \mathbf{Z})\}$ , where  $k \approx 0.588$ . These are approximately equivalent as long as the rare disease assumption is met [5,22]. With this approximation, the attenuation factor can be redefined as:

$$1.10 \quad \lambda \approx \frac{\beta_W}{\beta_X} + \beta_W k \sigma_{X|WZ}, \quad \sigma_{X|WZ}^2 = \sigma_{X|Z}^2 - \frac{\text{cov}(X, W|Z)^2}{\sigma_{W|Z}^2}$$

where  $\sigma_{X|WZ}^2$  is the residual variance in the model when  $X$  is regressed on  $W$  and  $\mathbf{Z}$ . When  $\sigma_{X|WZ}$  or  $\beta_X$  are small in logistic regression and the rare disease assumption is true, Equation 1.9 becomes approximately true.

When the functional form of the error-prone exposure is non-linear, the effect of measurement error is much more unpredictable. The primary effect is to mask the true shape of the association curve and to make the relationship appear more linear [1,23]. This reduces the power to detect non-linearities in the association and can reduce the power to detect the association at all. Risk may be both under- or overestimated depending on the exposure level [6,23]. Turning points in the curve, which may be of particular importance in drawing inference about what exposure is

associated with the least or greatest risk, may be obfuscated [6]. The associated outcome value at the turning point will be lower for a concave turning point thereby underestimating the maximum risk, or higher for a convex turning point thereby overestimating the minimum risk (illustrated in Section 1.5). When the functional form of the exposure is non-linear, there is no concise expression of the effect such as the attenuation factor (Section 4.3).

Keogh, Strawbridge, and White explored the effect of classical measurement error on six shapes deemed of interest in epidemiology in combination with the Cox proportional hazards model [23]. Grouped exposure analysis, the fractional polynomial method, and p-splines were used to fit the models. Bias in either direction (i.e. away from or toward the null) was observed as well as a shift in the turning point for some shapes. The loss of power to detect non-linearities as measurement error was increased, was demonstrated, and was particularly severe when employing the fractional polynomial method.

### 1.5.1 Illustration of classical measurement error in an untransformed predictor

#### Box 1.2 Simulated data

Sample size:  $N = 250$

True exposure:  $X_i \sim N(10, 1)$

Classical error model:

$$W_i = X_i + U_i, U_i \sim N(0, \sigma_U^2)$$

Measurement error variance:

$$\sigma_U^2 = \{0.25, 1, 4\}$$

Substantive model:

$$Y_i = \beta_0 + \beta_X X_i + \varepsilon_i$$

Residual errors:  $\varepsilon_i \sim N(0, 1)$

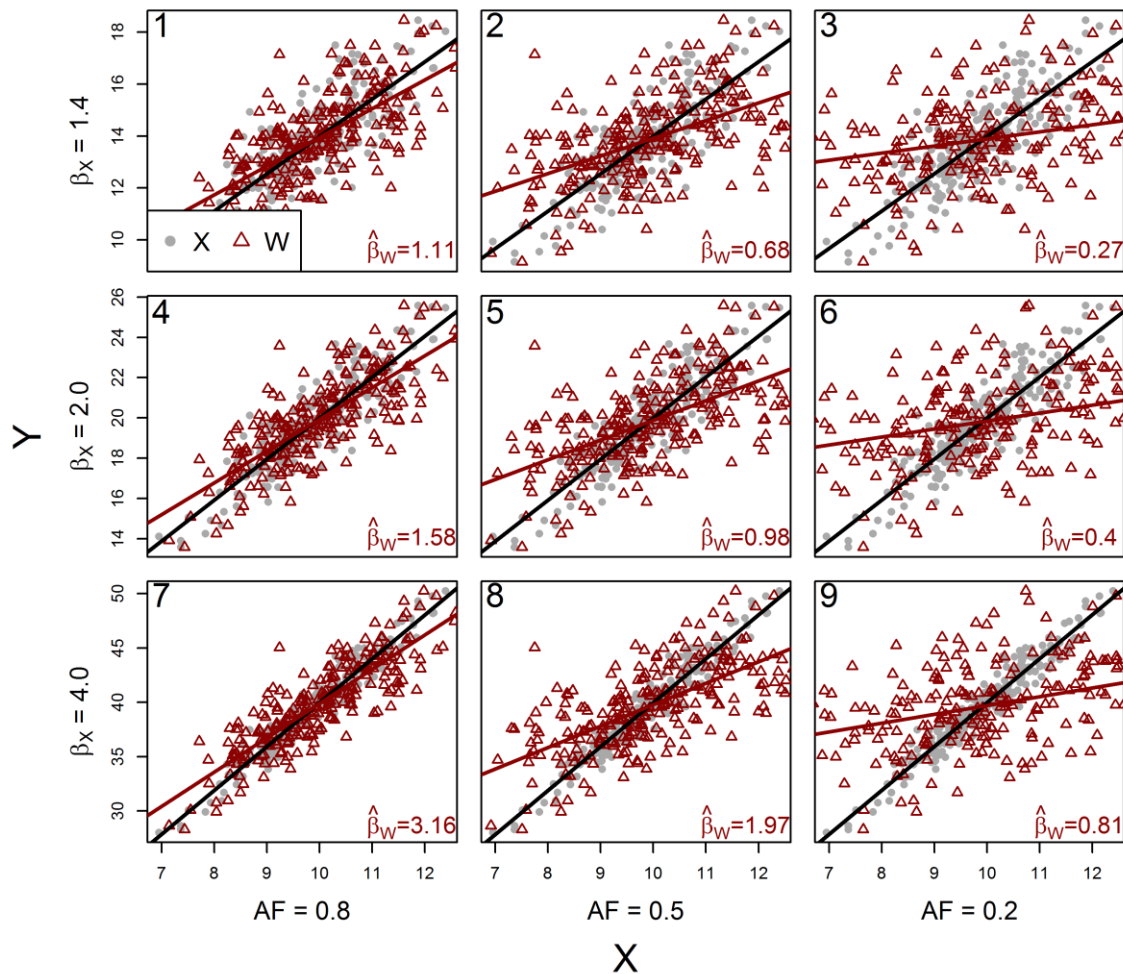
Effect size:  $\beta_X = \{1.4, 2.0, 4.0\}$

An illustrative data set was created with 250 simulated participants for nine scenarios with varying measurement error variance and effect size as detailed in Box 1.2. The measurement error variance values,  $\sigma_U^2 = \{0.25, 1, 4\}$ , are equal to  $\frac{1}{4}$  the variance of  $X$  or an attenuation factor of 0.8; equal to the variance of  $X$  or an attenuation factor of 0.5; or equal to 4-fold the variance of  $X$  or an attenuation factor of 0.2. The intercept,  $\beta_0$ , was placed at zero in all cases and no covariates were included.

In Figure 1.1, the resulting  $X$  and  $Y$  values are plotted in gray alongside the resulting  $W$  and  $Y$  values in red. One can see that the  $Y$  values remain fixed while the error-prone measures spread out across the horizontal axis to

either side of the  $X$  values. This spread, indicating the increased variance in  $W$  relative to  $X$ , creates the attenuation of the association with  $Y$ . Figure 1.1 also shows the fitted regression lines from regressions of  $Y$  on  $X$  (in black) and  $Y$  on  $W$  (in red), demonstrating the result from Equation 1.9 that the effect of measurement error is proportional to the true effect size. In absolute terms, underestimation of the exposure-outcome association due to measurement error is greater for larger effect sizes.

A larger sample size will not reduce the bias in  $\beta_W$  but can only estimate the biased parameter more precisely (i.e. estimates will become closer to  $\lambda\beta_X$ ; Equation 1.9). However, classical



*Figure 1.1 Simulated data demonstrating effect of classical measurement error on the association between a linear predictor,  $X$  (horizontal axis), and the outcome,  $Y$  (vertical axis).  $Y$ -values are plotted against  $X$ -values as gray dots and against  $W$ -values (the error-prone measure of  $X$ ) as red triangles. The attenuation factor ( $AF$ ) is varied across panels from left to right while the effect size ( $\beta_X$ ) is varied across panels from top to bottom. The black line indicates the linear regression of  $Y$  on  $X$  and the red line indicates the linear regression of  $Y$  on  $W$ . The naïve regression coefficient representing the association between  $W$  and  $Y$ ,  $\hat{\beta}_W$ , is shown in the bottom right of each panel.*

measurement error also reduces the power to detect a significant effect, resulting in a higher type II error rate, falsely accepting the null hypothesis. Larger sample size can overcome the measurement error in this case to detect an association.

### 1.5.2 Illustration of effect of measurement error in a transformed exposure

The effect of classical measurement error when the functional form of the error-prone exposure is non-linear is demonstrated using several common quadratic association shapes. These include three of the six shapes illustrated by Keogh, Strawbridge, and White, e.g. a U-shaped, J-shaped and increasing quadratic association, as well as a “weak quadratic” association [23]. These are illustrated in Figure 1.2. A U-shaped curve is relevant when there is very high relative risk at both extremes. A J-shaped curve implies greatly increased risk at one extreme of the exposure scale and a more moderate increase in risk at the other extreme. The association between alcohol and mortality has sometimes been found to be J-shaped, where non-consumers have a slightly

*Table 1.1 Models for exposure-outcome association shapes with a continuous outcome where the exposure has a non-linear functional form.*

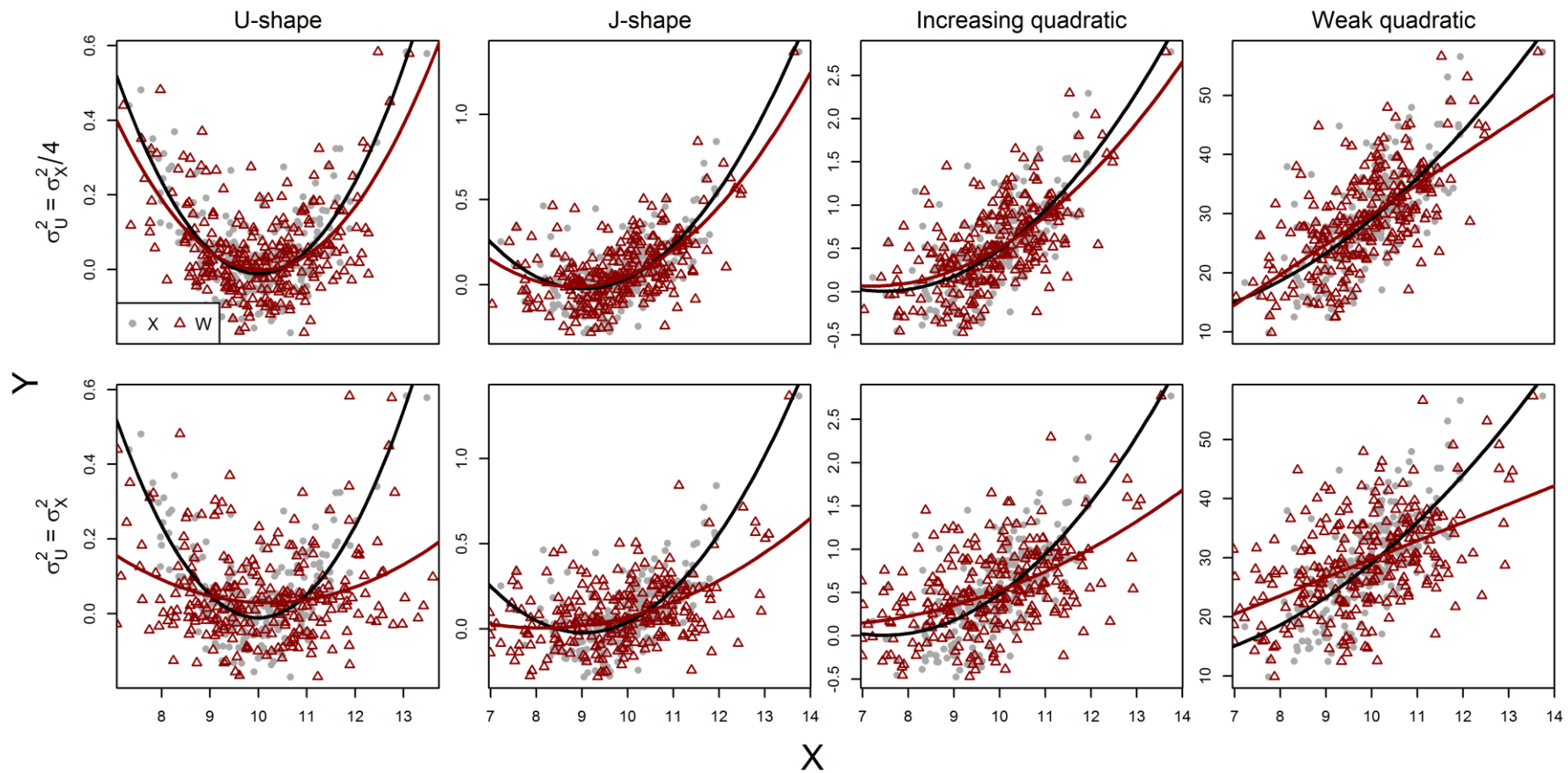
Shape	Linear predictor	Residual variance ( $\sigma_{Y X}^2$ )
U-shape	$0.06 * (X - 10)^2$	0.082
J-shape	$0.06 * (X - 9)^2$	0.14
Increasing quadratic	$0.06 * (X - 7)^2$	0.36
Weak quadratic	$0.3 * X^2$	6.0

increased risk over moderate consumers and risk increases sharply for heavy consumers [24–26]. The increasing quadratic shape contains no optima over the range of exposures modelled. The ‘weak quadratic’ shape illustrated here is another example of an increasing quadratic shape where the quadratic term in the linear regression model is relatively small. The weak quadratic shape is included in order to demonstrate the effect of measurement on power to detect non-linearities.

The true exposure and error-prone measure values were generated as described in the previous simulation (Box 1.2). Measurement error variance was implemented at two values,  $\sigma_U^2 = \{0.25, 1\}$ ; that is, measurement error variance was  $\frac{1}{4}$  the variance of  $X$  or equal to the variance of  $X$ . The models used to generate the outcome for each association shape can be found in Table 1.1. The value of  $\sigma_{Y|X}^2$  was chosen for each model to generate an  $R^2$  near to 0.5. As before, no covariates were included. The simulated data were then fit to a quadratic model (Equation 1.4).

Figure 1.2 illustrates the impact of measurement error for the four shapes. When the measurement error variance is just  $\frac{1}{4}$  the variance of  $X$  (Figure 1.2, top row), the shape of the error-prone curve is much closer to the shape of the true curve. The measurement error mostly leads to underestimation of the risk in groups with very high exposure for all shapes and groups with very low exposure for the U- and J-shapes. For the weak quadratic shape, one can see that even this small amount of measurement error makes the regression line appear very linear. When measurement error is equal to the variance of  $X$ , the naïve curve is a poor fit for all groups. One can readily see that the curve appears more linear for all shapes.

In Table 1.2, the p-value associated with the likelihood ratio comparing the quadratic model to the linear model can be found for each shape and measurement error. For the weak quadratic shape, the p-value from the model fitted using the true exposure  $X$  is 0.029. Using  $W$  in place of  $X$  increased the p-value to 0.073 for measurement error variance  $\frac{1}{4}$  the variance of  $X$  and to 0.258 for measurement error variance equal to the variance of  $X$ . For either measurement error variance, application of the likelihood ratio test in the naïve analysis would lead one to conclude that the association is linear rather than quadratic, assuming a type I error rate of 5%. For the increasing quadratic association, even with a p-value associated with selection of the quadratic for the true model of  $<0.0001$ , in the presence of measurement error with variance equal to the variance of  $X$ , the p-value became 0.031. For the U-shape and J-shape, although the quadratic model would still



**Figure 1.2** Simulated data demonstrating effect of classical measurement error on the association between a quadratic predictor,  $X$  (horizontal axis), and the outcome,  $Y$  (vertical axis).  $Y$ -values are plotted against  $X$ -values as gray dots and against  $W$ -values (the error-prone measure of  $X$ ) as red triangles. Measurement error variance ( $\sigma_U^2$ ) is either  $1/4$  the variance of  $X$  ( $\sigma_X^2$ ; top row) or equal to the variance of  $X$  (bottom row). The black line indicates the fitted curve from a regression of  $Y$  on  $X$  and  $X^2$  and the red line indicates the fitted curve from a regression of  $Y$  on  $W$  and  $W^2$ .

**Table 1.2 P-values from a likelihood ratio test comparing the quadratic model regressing  $Y$  on  $X$  and  $X^2$  and the linear model regressing  $Y$  on  $X$ , and corresponding results when  $W$  is used in place of  $X$ . Results are shown for four quadratic shapes. Measurement error variance ( $\sigma_U^2$ ) is either  $1/4$  the variance of  $X$  ( $\sigma_X^2$ ; top row) or equal to the variance of  $X$  (bottom row).**

		P-value from the likelihood ratio test			
		U-shaped association	J-shaped association	Increasing quadratic association	Weak quadratic association
$\sigma_U^2 = \sigma_X^2/4$	Using X	$7.3 \times 10^{-52}$	$6.6 \times 10^{-24}$	$1.2 \times 10^{-06}$	0.029
	Using W	$2.6 \times 10^{-24}$	$2.5 \times 10^{-14}$	$1.8 \times 10^{-04}$	0.073
$\sigma_U^2 = \sigma_X^2$	using X	$7.3 \times 10^{-52}$	$6.6 \times 10^{-24}$	$1.2 \times 10^{-06}$	0.029
	Using W	$4.7 \times 10^{-05}$	$3.0 \times 10^{-05}$	$3.1 \times 10^{-02}$	0.258

be selected, the significance associated with the likelihood ratio is reduced by many orders of magnitude.

## 1.6 VALIDATION AND REPLICATE STUDIES

The classical error model is a relatively simple error model and has only three unknown parameters,  $\mu_X$ ,  $\sigma_X^2$ , and  $\sigma_U^2$ . While  $\mu_X$  may be estimated from a single error prone measure gathered on each individual if one assumes the classical error model (i.e.  $E[W_i|\mathbf{Z}_i] = E[X_i|\mathbf{Z}_i]$ ), the values of  $\sigma_X^2$  and  $\sigma_U^2$  cannot. Further information is required to obtain a model which is identifiable [27].

In some cases, there may be external estimates of either  $\sigma_X^2$  or  $\sigma_U^2$  or other information about the measurement error that can be used to make inference (i.e. an external validation study). However, for the external information to be valid, one must have good evidence that the population in which the external validation study was performed was similar to the primary study population with respect to the exposure-outcome association. Additional uncertainty may need to be quantified and incorporated into a model which uses external validation information.

In this thesis, I will focus exclusively on settings in which either an internal validation study or a replicate study are present. Both rely on additional data collection within the study population.

In an internal validation study (hereafter denoted simply “validation study”) the true predictor is observed in a subset of the study population. In this case,  $\sigma_X^2$  and  $\mu_X$  may be estimated from the observed  $X_i$  values with no further assumptions required. Given the classical error model,  $\sigma_U^2$  may be estimated as the difference between the variance observed in  $W$  and  $\sigma_X^2$ . A validation study provides the best opportunity for checking model assumptions about the error directly from the data including whether there is systematic bias, heteroscedasticity, or differential error [5]. To provide reliable estimates of the parameters describing the measurement error, a validation study must be of sufficient size to capture the variability in the population.

Often it is too time-consuming, costly, or simply impossible to measure the true predictor. In these cases, one is limited to the use of a replicate study, where multiple error-prone exposure measurements are taken on all or a subset of individuals. Assuming the classical error model, the  $\mu_X$  may be estimated as the expected value of the error-prone measures and  $\sigma_X^2$  may be estimated as the covariance of the repeated error-prone measures, i.e.  $\text{cov}(W_{i1}, W_{i2})$ . As above,  $\sigma_U^2$  can then be estimated as the difference between the variance observed in  $W$  and  $\sigma_X^2$ . More complex models such as those with systematic bias are not usually identifiable with only replicate data [5].

Furthermore, even though a model may be identifiable asymptotically, for finite sample sizes a model must also have sufficient information from the likelihood to estimate the specific (i.e. singular) values of the parameters in the model. Congdon defines empirical identifiability in the book *Bayesian Statistical Modelling* as having data sufficient to precisely identify a complex model involving many parameters [28]. This concept will be further discussed as it applies to the Bayesian context in Chapter 2, Section 2.3. As will be shown in Chapter 3, even for classical measurement error when the model is identifiable asymptotically, finite sample sizes (i.e. small study size or replicates performed on a small proportion of the study population) may result in empirically non-identifiable variance parameters particularly when the attenuation factor is very low.

## 1.7 THESIS AIMS

The overarching aim of this thesis was to propose and explore methods that integrate exposure measurement error correction with either assumed polynomial models or polynomial model selection.

(Aim 1) In order to do so, it was first necessary to investigate the performance of the proposed measurement error correction methods in exposure-outcome relationships where the error-prone exposure is untransformed using simulation studies. This investigation covered both common scenarios (e.g. high signal-to-noise ratio, moderately large sample sizes) to establish common performance and more challenging scenarios (e.g. low signal-to-noise ratio, small sample sizes) to assess performance when little information is available from the likelihood. Methods were applied to a subset of the Framingham Heart Study data relating usual blood pressure and coronary heart disease for illustration where replicate measures of blood pressure are available [1,29].

(Aim 2) Promising methods identified from the first aim were extended as necessary and adapted for use in a polynomial model which incorporates the addition of a squared term for the error-prone exposure (Equation 1.4). Methods were evaluated using simulation studies for several different association shapes to assess whether bias in the curve was minimized after correction. Furthermore, I evaluated the power to detect the squared term after application of correction



methods as well as the type I error rate in selecting the squared term when the association was linear.

(Aim 3) The most promising methods from the second aim were extended for use with polynomial model selection using the fractional polynomial method. The power and type I error for each method were evaluated within the measurement error corrected selection framework. The ability to recover a shape with similar features to the original curve or at least of the same degree as the original curve was evaluated.

(Aim 4) Methods that can accommodate complex error models and the fractional polynomial method with only data from a replicate study were applied to the motivating example of usual alcohol intake and all-cause mortality. The association between usual alcohol intake and all-cause mortality has been postulated to be “J-shaped” [24–26]. As a non-linear relationship between alcohol and many outcomes has been observed, the fractional polynomial method is frequently employed in alcohol studies to determine the best polynomial model for modelling the dose-response relationship [30–32]. Estimation of usual alcohol intake, given unbiased observations of intake during brief time periods, requires a special error model which can accommodate the excess number of zeros recorded as people consume alcohol episodically. This error model is called the “episodic consumers model” or EC model and is described in greater detail in Chapter 7.

## **1.8 LITERATURE OVERVIEW**

Many methods have been developed for correction of the effects of measurement error in continuous variables and there are several books on the topic [1–3,27]. Regression calibration (RC) [33] is the simplest and most commonly used method. While RC is a powerful tool in many scenarios, it requires an approximation for use in non-linear exposure-outcome models (e.g. logistic regression or survival analysis) which produces an increasingly biased effect measure (e.g. the OR or hazard ratio) as the effect size and/or measurement error variance increases [20,34]. A recent systematic review [9] found that only 3% of original research articles in the top twelve medical and epidemiology journals published in 2016 corrected for measurement error and 11% (2) of these used RC.

Bayesian methods have been suggested as possible alternatives to RC for the correction of exposure measurement error [1,27,35–37]. Richardson and Gilks described Bayesian methods using Markov chain Monte Carlo (MCMC) sampling algorithms for exposure measurement error correction using conditional independence models [35,36]. Detailed overviews are given in the books by Carroll et al (Chapter 9) [1] and Gustafson [27]. There have been a few simulation studies published assessing the methods in terms of error model misspecification [38–40], and

fewer which directly compare the Bayesian approach to other methods of measurement error correction [37,40].

Use of integrated nested Laplace approximations (INLA) in place of MCMC sampling methods has recently been presented as a fast alternative in Bayesian analyses [41–43]. Muff et al adapted this method to classical measurement error problems where the error variances are known or where a replicate study is available [29]. However, the performance of INLA has not previously been investigated using simulation studies nor has it been adapted for validation studies. To my knowledge, this method has not yet been used in applied publications.

Finally, multiple imputation (MI), a method originally developed as a way of handling missing data, is derived from Bayesian principles [44–46]. MI has been suggested as a method for handling measurement error [5,47,48] but it has not been extensively assessed in simulation studies or, to my knowledge, used in practice. MI is now widely used for missing data and software for this purpose is well-developed. Additionally, MI is designed to produce estimates in the frequentist framework, which may be more comfortable for some researchers.

RC can be adapted easily when a quadratic transformation of the error-prone variable is assumed [1,49]. Strawbridge [49] further demonstrated the use of RC in the context of the fractional polynomial model using specific distributional assumptions about the error-prone exposure and a survival outcome. RC has also been extended for use with specific complex error models [50,51], including the EC model used in the EPIC-Norfolk alcohol and mortality application [17,18]. However, these models depend upon numerical integration techniques which have been found to suffer problems of convergence [52] for some data sets.

Carroll et al demonstrated measurement error correction in Bayesian analysis for specific functional forms of the exposure within the substantive model [1]. While some work has been done in applying the Bayesian approach to measurement error correction in combination with splines or nonparametric models [1,53,54], no publications have specifically addressed polynomial model selection in the Bayesian context in the presence of measurement error. Bové Sabanés proposed a method of Bayesian fractional polynomials for high dimensional models [55], but this method does not adapt easily to more widely available software and does not incorporate measurement error correction. More complex error models are easily adapted for use in the Bayesian context and many examples are available in the literature [39,56–58] including implementation of the EC model [19,52,59].

No publications have explicitly explored either INLA or MI for measurement error correction with a transformation of the error-prone variable in the substantive model or a more complex error model. The “substantive model compatible” approach to MI described by Bartlett et al is thought

to be able to accommodate a substantive model (i.e. the exposure-outcome model) wherein the error-prone exposure is transformed as has been described for missing data [46].

## **1.9 THESIS CONTRIBUTIONS**

The contributions to the field of measurement error correction contained in this thesis are summarized below.

I perform a large simulation study comparing proposed methods for measurement error correction from the Bayesian family, namely MCMC, INLA, and MI to the most common method of measurement error correction (i.e. RC) under a variety of conditions for a single continuous exposure when that exposure is untransformed. This is performed for the setting with a validation study and the setting with only a replicate study. These methods have not been previously compared.

I extended the method proposed by Muff et al using INLA to correct for classical measurement error in the replicate study setting for use in the validation study setting [29]. Additionally, I present for the first time an extension of the ‘substantive model compatible’ method of Bartlett et al [46] for missing data to the measurement error context accommodating either a validation study or a replicate study making use of code and earlier investigation by Jonathan Bartlett (unpublished). Example code for each of these methods is available in Appendix A.

In order to accommodate more complex measurement error and unknown polynomial transformations of the error-prone variable, two novel measurement error correction approaches are proposed and explored in simulation studies alongside more standard methods. The first is a hybrid method combining RC with Bayesian sampling via either MCMC or INLA (example code in Appendix B). This hybrid method is particularly efficient for use with the fractional polynomial method. Unlike standard RC, it avoids the need for explicit derivation of the expectation of each transformation of the error-prone exposure. More complex measurement error models may be accommodated and prior knowledge incorporated. This hybrid method is also much faster to run than a fully Bayesian model incorporating measurement error correction, fitting of the substantive model, and model selection.

The second novel approach is a fully Bayesian approach which adapts methods of Bayesian variable selection to selection of the best polynomial transformation(s) in a manner similar to the fractional polynomial method while incorporating measurement error correction. Two specific methods of Bayesian variable selection methods are proposed and evaluated [60,61]; code for each is included in Appendix B and C.

The EC model has been previously described using maximum likelihood methods [17,18] as well as Bayesian MCMC with a bespoke sampler in MATLAB (MATLAB and Statistics Toolbox, The MathWorks, Inc) [19,52]. In order to apply promising methods to the motivating example of usual alcohol intake and all-cause mortality, it was required that I adapt the EC model for use in the more user-friendly JAGS program [62] and for use in the INLA software ([www.r-inla.org](http://www.r-inla.org)) via R packages (code provided in Appendix D).

Finally, I apply the most promising methods to estimation of the association curve in the motivating example of usual alcohol intake and all-cause mortality.

I have prepared a manuscript prepared for submission to Statistics in Medicine including a large simulation study comparing RC, MCMC, INLA, and MI for an untransformed error-prone exposure; the abstract is included in Appendix E. A manuscript describing one of the novel approaches and its application to the motivating example has been submitted to The Biometrical Journal and is in the second round of review (abstract in Appendix F).

## **1.10 ILLUSTRATIVE EXAMPLES**

In this thesis I will illustrate the methods described using two examples of important exposure used in epidemiological studies that are subject to measurement error. The first example is the relationship between usual systolic blood pressure (SBP) and coronary heart disease (CHD). Usual blood pressure, i.e. the long term average, is subject to classical measurement error as a person's blood pressure fluctuates from day to day [63]. It is not possible to get a measure of the true usual blood pressure ( $X$ ), therefore replicate measures of the error-prone measure ( $W$ ) must be used. The second example is the association between usual alcohol intake and all-cause mortality. Even an unbiased measure of alcohol intake over a brief period of time is subject to classical measurement error; additionally, measures of alcohol intake are subject to excess zeros due to the episodic nature of consumption [17,19,59,64,65]. Furthermore, the relationship between usual alcohol intake and all-cause mortality is non-linear requiring either a first or second degree polynomial.

The blood pressure and CHD example is used to illustrate the measurement error correction methods when the error-prone exposure is untransformed and a replicate study has been performed. I make use of a subset of the Framingham Heart Study data which was used by Carroll et al [1] and Muff et al [29] for illustrating measurement error correction methods. The Framingham Heart Study was initiated in 1948 to identify common factors leading to CHD [66]. The study initially enrolled 4,641 individuals aged 40–69 (44% male) [67]. The subset included only men aged 45 or over at the second exam with a serum cholesterol between 200 and 300 at the third exam ( $N = 641$ ). Replicate measures of SBP two and four years after baseline were

available. CHD was assessed ten years after baseline and is included as a binary outcome. Further details can be found in Chapter 3, Section 3.6.

The second example, usual alcohol intake and all-cause mortality, will make use of data from the EPIC-Norfolk study [68,69]. The EPIC-Norfolk study recruited 25,639 eligible individuals aged 39 to 79 years between 1993 and 1998 from the population of Norfolk, UK as a part of the European Prospective Investigation into Cancer and Nutrition [70]. Measures of alcohol intake over one week were obtained using 7-day food diaries, considered to be an unbiased measure for the time observed. Up to three such measures taken over six years were available for each individual. The study and methods of dietary assessment have been described in detail elsewhere [68,70,71]. Death due to any cause was assessed for the entire cohort until 31 March 2015. All-cause mortality up to 15 years after enrollment was used as a binary outcome. Potential confounders such as sex, age, smoking status, body mass index ( $\text{kg}/\text{m}^2$ ), educational attainment, and social class were established at enrollment. While ethnicity details for each participant were included, the homogeneity of the data set precludes its use in the model. Further details can be found in Chapter 7.

## **1.11 OVERVIEW OF THE THESIS**

In this first chapter, I have detailed the structure of the problem and demonstrated the effect of classical measurement error on linear and quadratic associations. Chapter 2 presents the proposed methods for measurement error correction when the error-prone exposure takes a linear function form within the substantive model. In Chapter 3, these methods are compared using extensive simulation studies for continuous and binary outcome models. Measurement error correction methods are extended, where possible, for use when a squared term of the error-prone exposure is included in the model in Chapter 4. Methods for selection of the correct model, linear or quadratic, are explored in Chapter 5. In Chapter 6, methods are extended for use with the fractional polynomial method wherein the best polynomial functional form of the error-prone exposure is determined. In Chapter 7, methods capable of correcting bias from exposure measurement error using a complex error model in the context of the fractional polynomial method are demonstrated by modelling the association of alcohol and mortality. The thesis concludes with a discussion and directions for further work.

## 2 METHODS FOR MEASUREMENT ERROR CORRECTION

---

### 2.1 AIMS

In this chapter, I aim to outline four methods for measurement error correction: regression calibration (RC), a Bayesian approach using Markov chain Monte Carlo (MCMC), a Bayesian approach using integrated nested Laplace approximations (INLA), and multiple imputation (MI). The focus will be on the implementation of each method with an untransformed error-prone exposure  $X$ . Specific extensions for adaptation to non-linear transformations of the error-prone exposure in the substantive model will be detailed in further chapters.

RC is used throughout as a reference method, but, as discussed below, has a number of limitations.

Bayesian approaches were felt to be the most flexible for adaptation to more complex scenarios, such as the setting with a transformed exposure or complex measurement error, and are therefore the focus of this work. Carroll et al referred to Bayesian analysis as fundamentally modular and therefore more able to incorporate complex models with one another [1]. Bartlett and Keogh demonstrated this feature in combining measurement error correction and the handling of missing data into a single Bayesian model [37]. Modern software has made Bayesian methods more accessible; it is no longer required that the researcher be able to derive the likelihood for the joint model or to program their own sampler for each application.

Also introduced in this chapter is the software used to apply the methods, as user-friendly software is key to the widespread use of Bayesian analysis.

### 2.2 REGRESSION CALIBRATION (RC)

RC is the most commonly applied method for correction of classical measurement error. It was first introduced by Armstrong in 1985 in the context of generalized linear models [33]. RC has since been extended to logistic regression by Rosner et al (1990; 1992) and Spiegelman et al (1997) and to survival models by Hughes (1993) [21,34,72,73].

The substantive model of interest is  $g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_X X_i + \boldsymbol{\beta}_Z^T \mathbf{Z}_i$  (Equation 1.1). In RC, the true but unobserved exposure  $X_i$  is replaced in the substantive model by its expectation conditional on the observed error-prone measures,  $\mathbf{W}_i$ , and any accurately measured covariates,  $\mathbf{Z}$  (i.e.  $E(X_i|\mathbf{W}_i, \mathbf{Z})$ ), to obtain an (approximately) unbiased estimate of  $\beta_X$  [1,20].

RC uses the result that  $E[Y_i|\mathbf{W}_i, \mathbf{Z}_i] = E[E[Y_i|\mathbf{W}_i, X_i, \mathbf{Z}_i]|\mathbf{W}_i, \mathbf{Z}_i] = E[E[Y_i|X_i, \mathbf{Z}_i]|\mathbf{W}_i, \mathbf{Z}_i]$ , where the second step uses the assumption that error in  $X$  is non-differential error. When the outcome is continuous, i.e. the substantive model is linear regression, we have:

$$\begin{aligned}
2.1 \quad E[Y_i | \mathbf{W}_i, \mathbf{Z}_i] &= E[E[Y_i | \mathbf{W}_i, X_i, \mathbf{Z}_i] | \mathbf{W}_i, \mathbf{Z}_i] \\
&= E[E[Y_i | X_i, \mathbf{Z}_i] | \mathbf{W}_i, \mathbf{Z}_i] \\
&= E[(\beta_0 + \beta_X X_i + \boldsymbol{\beta}_Z^T \mathbf{Z}_i) | \mathbf{W}_i, \mathbf{Z}_i] \\
&= \beta_0 + \beta_X E[X_i | \mathbf{W}_i, \mathbf{Z}_i] + \boldsymbol{\beta}_Z^T \mathbf{Z}_i.
\end{aligned}$$

When a validation study is available, the expectation  $E[X_i | \mathbf{W}_i, \mathbf{Z}_i]$  can be obtained simply by regressing  $X_i$  on  $W_{i1}$  and  $\mathbf{Z}_i$ :

$$2.2 \quad X_i = \lambda_0 + \lambda W_{i1} + \boldsymbol{\lambda}_Z^T \mathbf{Z}_i + e_i.$$

In the case of a replicate study, an estimate of  $E[X_i | \mathbf{W}_i, \mathbf{Z}_i]$  can be obtained from a regression of  $W_{i2}$  on  $W_{i1}$  and  $\mathbf{Z}_i$  [21]:

$$2.3 \quad W_{i2} = \lambda_0 + \lambda W_{i1} + \boldsymbol{\lambda}_Z^T \mathbf{Z}_i + e_i^*.$$

The regression coefficient  $\lambda$  in the RC models 2.2 and 2.3 is the same as the attenuation factor defined in Equation 1.9. Using  $E[X_i | \mathbf{W}_i, \mathbf{Z}_i]$  in place of  $X_i$  in the substantive model is equivalent to applying  $\hat{\lambda}^{-1}$  to  $\hat{\beta}_W$  as a correction factor (i.e.  $\beta_X = \beta_W \lambda^{-1}$ ) (Section 1.4.1).

More efficient methods of RC have been proposed for use in the setting of a replicate study. As described above, the standard approach to estimating  $\lambda$  in a replicate study is to regress  $W_2$  on  $W_1$  and  $Z$  arriving at  $\hat{\beta}_{X,RC1}$ . Alternatively, one could regress  $W_1$  on  $W_2$  and  $Z$ , giving an alternative estimate  $\hat{\lambda}_{ALT}$  to arrive at  $\hat{\beta}_{X,RC2}$ . Spiegelman et al proposed ‘efficient RC’ which uses a weighted combination of  $\hat{\beta}_{X,RC1}$  and  $\hat{\beta}_{X,RC2}$  [74]. The weights are the inverses of the estimated variances. Alternatively, assuming  $f(X_i, \mathbf{W}_i | \mathbf{Z}_i)$  is multivariate normal, it can be shown that the distribution of  $X_i$  given  $\mathbf{W}_i$  and  $\mathbf{Z}_i$  (with probability density function  $f(X_i | \mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\gamma})$ ) is a normal distribution with mean  $\mu_{X_i}$  and variance  $\sigma_{X_i | WZ_i}^2$  given by [1,5,37]:

$$2.4 \quad \mu_{X_i} = E[X_i | W_{i1}, W_{i2}, \mathbf{Z}_i] = E[X_i | \mathbf{Z}_i] + \frac{\sigma_{X_i | Z}^2}{2\sigma_{X_i | Z}^2 + \sigma_U^2} (W_{i1} + W_{i2} - 2E[X_i | \mathbf{Z}_i])$$

$$2.5 \quad \sigma_{X_i | WZ_i}^2 = \text{var}(X_i | W_{i1}, W_{i2}, \mathbf{Z}_i) = \frac{\sigma_{X_i | Z}^2 \sigma_U^2}{2\sigma_{X_i | Z}^2 + \sigma_U^2}.$$

Parameters are estimated via maximum likelihood as a random-intercepts mixed model for  $\mathbf{W}_i$  conditional on  $\mathbf{Z}_i$ .

RC can be applied in the same way as described above for logistic regression and Cox regression, though approximations are required for these non-linear models [21,33,34,72]. For logistic regression, the approximation requires that the outcome be rare and the log odds ratio (logOR) or

the measurement error variance be small [6,20]. For the Cox proportional hazards model the event rate must be low or the measurement error variance small [72].

Alternatively, a finer approximation for logistic regression may be applied using the higher order approximation in Equation 1.10 to replace  $\lambda$  and  $\sigma_{X|WZ_i}^2$  in either standard RC or the more efficient RC implementations.

For simplicity in this thesis, in extension to the models where the functional form of the predictor is non-linear, I focus exclusively on standard RC implementations.

### 2.2.1 Estimation of variance

Fitting  $E[X_i|W_i, Z_i]$  in place of  $X_i$  in Equation 1.1, one may reduce the bias in the point estimate of  $\beta_X$ , but the variance of  $\beta_X$  within the regression model will be underestimated using standard methods from the regression model as it does not incorporate uncertainty due to estimation of the parameters in the measurement error model [1].

Using the naïve estimate of  $\beta_X$ ,  $\hat{\beta}_W$  from Equation 1.8, and the estimated attenuation factor,  $\hat{\lambda}$ , a second-order Taylor approximation, i.e. the delta method, may be used to estimate the variance of  $\hat{\beta}_X$  [5]:

$$2.6 \quad \text{var}(\hat{\beta}_X) \approx \frac{\text{var}(\hat{\beta}_W)}{\hat{\lambda}^2} + \left(\frac{\hat{\beta}_W}{\hat{\lambda}^2}\right)^2 \text{var}(\hat{\lambda}),$$

where  $\text{var}(\hat{\beta}_W)$  and  $\text{var}(\hat{\lambda})$  are estimated using standard methods from the regression model in Equation 1.8 and the linear regression models in Equations 2.2 or 2.3, respectively. The above approximation assumes that estimates of  $\beta_W$  and  $\lambda$  are independent random variables. In Chapter 3, RC is implemented using this method to find the variance of  $\hat{\beta}_X$ .

Alternatively, bootstrapping may be used to obtain estimated standard errors and construct 95% confidence intervals (CIs) [1,75–77]. Bootstrapping is a re-sampling method which assumes that each sample is independent and identically distributed and that the empirical distribution of a parameter is similar to the true distribution of the parameter. To perform bootstrapping, individuals are sampled with replacement from the study data set to re-create a data set of the same sample size. This is repeated many times, typically 1000-10,000. The method of interest is then applied to each bootstrapped data set and an estimate of the parameter of interest, in our case  $\hat{\beta}_X$ , is obtained. The variance of the parameter estimates over the bootstrapped data sets provides an estimate of the variance of the parameter(s) of interest. CIs can be obtained from bootstrapped data in several ways, but in this work they are obtained using the non-Studentized pivotal method, also referred to as the “basic” method [77]. Where a validation study is performed or a replicate study is performed on only part of the study population, bootstrapping can be performed by



stratifying (i.e. individuals within the validation study are sampled separately from those outside of the validation study) to preserve the assumption of independent and identically distributed samples [1,3].

### 2.2.2 Extensions to non-linear functional forms of the predictor

Extension of RC, as described by Carroll et al [1] and Strawbridge [49], for use when the error-prone predictor is squared in the substantive model is detailed in Chapter 4. RC was extended for use with the fractional polynomial method for polynomial regression model selection by Strawbridge [49]; this extension is detailed in Chapter 6.

### 2.2.3 Software

RC does not require specialized software as it is based on two standard regressions. There is a Stata package, “rcal”, which simplifies estimation of standard errors [78]. To my knowledge, there is no package available in R.

In this thesis, the delta method as described is implemented in R and bootstrapping is achieved using the “boot” package in R.

## 2.3 BAYESIAN ANALYSIS - OVERVIEW

In Bayesian analysis, estimation is based on the posterior distribution of the parameters, which is a function of the prior distribution (reflecting prior knowledge or belief about the parameter values) and the likelihood of the parameters given the data. To find the posterior distribution directly, Bayes’ Rule may be applied; that is, the posterior density of the parameters given the data (and implicitly the model) is equal to the likelihood multiplied by the prior density, represented by  $f(\theta)$ , over the probability density of the data. Using the law of total probability, the probability of the data is re-framed as the integration of the likelihood and the prior densities over all possible values of the parameters. In the measurement error context, the latent  $X$  is treated as a parameter and the joint posterior is as follows:

$$2.7 \quad f(X, \theta | \mathbf{W}, Y, \mathbf{Z}) = \frac{f(\theta)f(\mathbf{W}, Y, \mathbf{Z}|X, \theta)}{\int f(\theta)f(\mathbf{W}, Y, \mathbf{Z}|X, \theta)d\theta}$$

The posterior distribution of the parameters of interest frequently involves intractable integrals. For very simple models, this can be avoided by selecting prior probability distributions which are conjugate to the likelihood, i.e. distributions which result in a posterior which belongs to the same family of distributions. MCMC sampling methods, INLA, and Hamiltonian Monte Carlo sampling methods have now made it possible to perform Bayesian analyses without explicitly solving the integrals involved [41,79–83]. I will describe the first two methods in greater detail in Sections 2.4 and 2.5 for implementation in this thesis.

Use of a three-part conditional independence structure to frame the problem of measurement error correction using a Bayesian analysis was proposed by Stephens and Dellaportas in 1992 [84] and put in the specific context of epidemiology by Richardson and Gilks in 1993 [35,36]. The three-part structure involves the models defined in Box 2.1.

<b>Box 2.1 Three-part conditional independence structure</b>		
<b>Substantive model:</b>	Defines the relationship between the outcome $Y$ and the true exposure $X$ conditional on any accurately measured covariates $\mathbf{Z}$ .	$f(Y_i X_i, \mathbf{Z}_i; \boldsymbol{\beta})$
<b>Measurement error model:</b>	Relates $X$ to the error-prone measures $\mathbf{W}$ .	$f(\mathbf{W}_i X_i; \boldsymbol{\pi})$
<b>Exposure model:</b>	Specifies the distribution of the $X$ in the population dependent on other covariates $\mathbf{Z}$ .	$f(X_i \mathbf{Z}_i; \boldsymbol{\alpha})$

$\boldsymbol{\beta}$ ,  $\boldsymbol{\pi}$ , and  $\boldsymbol{\alpha}$  are vectors of parameters in the respective models and are denoted collectively by  $\boldsymbol{\theta}$ . It can be seen by factorization of the likelihood that these three models together form the joint distribution of the data given the model:

$$\begin{aligned}
 2.8 \quad L(X, \boldsymbol{\theta}|\mathbf{W}, Y, \mathbf{Z}) &= \prod_{i=1}^N f(\mathbf{W}_i, Y_i, \mathbf{Z}_i|X_i, \boldsymbol{\theta}) \\
 &= \prod_{i=1}^N f(Y_i|X_i, \mathbf{Z}_i; \boldsymbol{\beta})f(X_i|\mathbf{Z}_i; \boldsymbol{\alpha})f(\mathbf{W}_i|X_i; \boldsymbol{\pi}).
 \end{aligned}$$

In Chapter 1, specific forms for the substantive model (e.g. Equation 1.1,  $g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_X X_i + \boldsymbol{\beta}_Z^T \mathbf{Z}_i$ ) and the measurement error model (e.g. Equation 1.7,  $W_i = X_i + U_i$ ) were introduced. A form for the exposure model is described below (Section 2.3.1). The advantage of the three-part structure is that Bayesian analysis can be easily re-formulated with different forms for the three component models.

In Section 1.6, I introduced the concept of empirical identifiability; that is, when there is insufficient information in the likelihood alone given the data to identify the parameters of a model [28]. In Bayesian analysis, as one is estimating a posterior distribution instead of point estimates, the model can be weakly or poorly identifiable, i.e. having a very broad and uninformative posterior distribution, or strongly identifiable, i.e. having a narrow posterior distribution with most of the probability density over a narrow range of values. If the data is not empirically identifiable, then a very flat or uninformative prior will lead to poor identifiability in the posterior. Specifying stronger prior distributions, i.e. with greater probability density over the expected values, will result in a more informative posterior.

It is typically assumed that the priors for each of the parameters are independent of one another so that the prior density of all parameters is equal to the product of the marginal densities, i.e.  $f(\boldsymbol{\theta}) = f(\boldsymbol{\beta})f(\boldsymbol{\alpha})f(\boldsymbol{\pi})$ . Regression coefficients are typically given Gaussian priors with means of zero and large variances, making the priors “vaguely informative” or having little impact on

the posterior relative to the likelihood. Variance parameters are often given inverse Gamma distribution priors which only support positive values [1,27,85]. The parameters describing the prior distributions are often called hyperparameters to distinguish them from the parameters of the model. One may go further and assign hyperpriors, i.e. prior distributions for the hyperparameters, within the model to further indicate uncertainty. Standardization of variables can be used in order to more easily specify a prior distribution [28,85]; for example, if a value is scaled by its standard deviation, then the standard deviation, variance, and precision estimates can all be expected to be centered near 1.

### 2.3.1 Exposure model

The exposure model describes the distribution of  $X$  in the population conditional on any accurately measured covariates  $\mathbf{Z}$ ,  $f(X|\mathbf{Z})$ .  $X$  is generally assumed to be normally distributed about the mean depending linearly on  $\mathbf{Z}$ :

$$2.9 \quad X_i = \alpha_0 + \boldsymbol{\alpha}_Z^T \mathbf{Z}_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_{X|Z}^2),$$

where  $\sigma_{X|Z}^2$  represents the variance of  $X$  conditional on  $\mathbf{Z}$  and  $\alpha_0$  represents the expectation of  $X_i$  for baseline values of  $\mathbf{Z}$ . If there are no covariates in the model, this is simply the marginal distribution of  $X$ .

When the distribution of  $X|Z$  is skewed, a suitable exposure model may instead be a lognormal distribution, for example.

## 2.4 MARKOV CHAIN MONTE CARLO (MCMC) SAMPLING

A simple example of an MCMC algorithm is the Metropolis algorithm, first introduced in 1953 [79]. From some starting value, the Metropolis algorithm requires a proposal distribution,  $P(\zeta)$ , that is proportional to the target distribution,  $T(\zeta)$ , and that can be evaluated for some parameter or set of parameters  $\zeta$ . In Bayesian analysis,  $P(\zeta)$  is the product of the likelihood and the prior distribution (i.e. the numerator of Equation 2.7) and  $T(\zeta)$  is the posterior distribution. A symmetric distribution,  $g(\zeta|\zeta_0)$ , is used to generate a proposed move from the starting value  $\zeta_0$ . In the simplest case,  $g(\zeta|\zeta_0)$  may be a normal distribution where the mean is the initial starting position  $\zeta_0$  and a standard deviation is specified. The ratio of  $P(\zeta)/P(\zeta_0)$  is compared to a draw from the uniform distribution between zero and one,  $v$ ; if  $P(\zeta)/P(\zeta_0)$  is greater than or equal to  $v$ , then the proposed move is accepted and the value(s) of  $\zeta$  is saved as a sample in the MCMC chain. If  $P(\zeta)/P(\zeta_0)$  is less than  $v$ , the position is rejected and a new proposed move is generated

from  $g(\zeta|\zeta_0)$ . A complete iteration or cycle of the algorithm concludes when a proposed move is accepted. An iteration of the Metropolis algorithm is summarized in Box 2.2.

**Box 2.2 Metropolis algorithm**

1. Generate a proposed move from the initial values,  $\zeta_0$  to  $\zeta$  via a symmetric distribution  $g(\zeta|\zeta_0)$ .
2. Calculate  $\alpha = P(\zeta)/P(\zeta_0)$ .
3. Draw a value,  $U$ , from the uniform distribution between 0 and 1.
4. If  $\alpha \geq U$ , accept the proposed move to  $\zeta$  as a sample otherwise return to Step 1.

Proposed moves in the Metropolis algorithm may be in one dimension or multiple dimensions. The Metropolis algorithm cycles through the above process until the probability of being at any given position  $\zeta$  resembles the probability of that position in the target distribution,  $T(\zeta)$ . This final equilibrium is termed the stationary distribution. The cycles before reaching the equilibrium are termed the “burn-in”. Once the algorithm has reached the stationary distribution, additional cycles of the algorithm are used to collect samples for inference.

If no stationary distribution exists, the algorithm will not reach an equilibrium. A stationary distribution will not exist if the model is not identifiable; this often occurs in discrete mixture models where “label-switching” occurs between iterations of the algorithm [28].

Hastings expanded on the Metropolis algorithm in 1970 to the more general case where the condition of symmetry is not required for  $g(\zeta|\zeta_0)$  resulting in the Metropolis-Hastings algorithm [80,85]. The Metropolis-Hastings algorithm remains a powerful algorithm today as it can be employed to find the posterior distribution of virtually any defined model. This sampler is frequently used with logistic regression models.

The average ratio of accepted moves to proposed moves constitutes the acceptance rate and is largely responsible for determining whether an algorithm is efficient. When the acceptance rate is very low, as is often the case for more challenging models, this algorithm converges to the stationary distribution very slowly. A second important aspect of efficiency is whether the MCMC chain has good “mixing” through the various states of the model or parameter values; that is, that parameter values change states over the full range of the posterior distribution. Poor mixing can occur when there is multimodality in the posterior distribution [82].

When a closed-form solution is available for the conditional distribution of a single parameter  $\zeta_j$  conditional on all other parameters  $\zeta_{-j}$  and the data  $D$ ,  $f(\zeta_j|\zeta_{-j}, D)$ ,  $g(\zeta|\zeta_0)$  is replaced with this conditional distribution. The proposed move is necessarily in only one dimension while all other values  $\zeta_{-j}$  are held constant, but it is always accepted because the proposal distribution is

equivalent to the marginal posterior distribution of that parameter. This algorithm is termed a Gibbs sampler, and it is very efficient because the acceptance rate is one [86]. Linear regression typically uses a Gibbs sampler.

An advantage of MCMC algorithms is that they may be applied one parameter at a time or to a block of parameters so that the most suitable sampler is applied. This contributes to the modularity of Bayesian analysis. Modern software (Section 2.4.4) automates the selection of the best sampler for each parameter or block of parameters. Most models, all models implemented in this thesis, use multiple sampling algorithms together including those described above in addition to a variety of other fit-for-purpose samplers.

### 2.4.1 Sampling and convergence diagnostics

There are many methods which assist in determining whether an MCMC chain has converged to the stationary distribution. However, these tests are limited in that they can only detect lack of convergence rather than directly prove that convergence has occurred.

MCMC sampling methods can be sensitive to starting positions of the parameters. Additionally, for a single chain, a local optima may be encountered before the true stationary distribution is reached. Therefore, multiple chains are typically run from different starting positions (usually random) [28,82].

Gelman and Rubin's convergence diagnostic,  $\hat{R}$ , can be used to assess whether the chains have converged to the same distribution, presumably the stationary distribution [87].  $\hat{R}$  is an estimate of the factor by which the scale of the current distribution of the parameters may be reduced if sampling were continued indefinitely. For a single parameter, the marginal posterior variance may be estimated by a combination of the within-chain variance  $A$  and between-chain variance  $B$  weighted by the number of posterior samples  $s$ ,  $\widehat{\text{var}}(\theta) = \left(\frac{s-1}{s}\right)A + \frac{1}{s}B$ .  $\hat{R}$  is then estimated as  $\hat{R} = \sqrt{\left(\left(\frac{s-1}{s}\right)A + \frac{1}{s}B\right)/A}$ . A multivariate version was proposed in 1998 [88] and is applied in this thesis.

The Geweke diagnostic [89],  $\hat{G}$ , is used to assess within a single chain whether the chain has reached a stationary distribution for a single parameter. The Geweke diagnostic finds the difference in the posterior mean of the samples in the first 10% of the chain to the mean in the latter 50% of the chain and divides this difference in sample means by the estimated standard error to arrive at a Z-score.

Finally, the effective sample size (ESS) can be assessed to ensure sufficient samples are collected for inference [87]. This metric attempts to account for the correlation between MCMC samples within a chain for a given parameter. Where  $c$  is the number of chains and  $s$  is the number of

samples, the total number of samples  $cs$  are not independent (this is called “autocorrelation”) and must be weighted by an estimate of the correlations between samples:

$$2.10 \quad \text{ESS} = \frac{cs}{1+2\sum_{t=1}^T \left(1 - \left(\frac{V_t}{2\text{var}(\theta)}\right)\right)}, \quad V_t = \frac{1}{c(s-t)} \sum_{b=1}^c \sum_{a=t+1}^s (\theta_{a,b} - \theta_{a-t,b})^2$$

where  $V_t$  is the variogram at each updating of the algorithm  $t$ .

Scaling of regression covariates, and the outcome if continuous, enables better estimation of the priors for the regression coefficients and improves the speed of convergence of the MCMC chains. Where there is high collinearity between covariates in regression modelling, MCMC will converge slowly; this may be addressed by centering the covariates at zero [28].

#### 2.4.2 Estimation of variance

In this thesis, Bayesian inference is performed using posterior means and 95% credible intervals (CrIs) instead of traditional CIs to quantify uncertainty. The 95% CrI expresses the range over which 95% of the probability distribution lies, or having a 95% probability the true value lies therein. In contrast, the frequentist CI expresses the range wherein, were the test repeated many times, it is anticipated that the true value would lie between the values in 95% of results. An equi-tailed 95% CrI can be obtained from the 2.5 and 97.5 percentiles of the posterior distribution. Alternatively, a highest posterior density (HPD) 95% CrI can be obtained from the shortest interval for which the posterior probability for that region is 0.95 [90]. In this thesis, HPD 95% CrIs are obtained from MCMC sampling estimates.

CrIs are more intuitive in their meaning than CIs and are better able to describe non-symmetric distributions. However, frequentist properties of the Bayesian estimates can be determined as well [91].

#### 2.4.3 Extensions to non-linear functional forms of the predictor

Bayesian analysis using MCMC can be extended easily to models wherein the functional form of the error-prone predictor is non-linear. Examples where the predictor takes a polynomial form or even a smoothing function have been published elsewhere [1,53,92]. Bayesian modelling using MCMC when the substantive model includes a squared transformation of the error-prone exposure is included in Chapter 4. Also in Chapters 4-6, a hybrid method combining Bayesian modelling and properties of RC will be presented for use with polynomial transformations of the error-prone exposure in the substantive model. In Chapter 5, I will also introduce a novel application of Bayesian variable selection methods to the selection of the best polynomial transformation of the error-prone exposure.

#### 2.4.4 Software

Bayesian analysis with MCMC sampling can be performed without advanced knowledge of the algorithms or derivation of the full likelihood distribution using software packages such as JAGS [62], OpenBUGS [93], and WinBUGS [94]. Each can be implemented in stand-alone software; alternatively, one can use R packages such as “rjags”, “BRugs” or “R2WinBUGS” to call each software, respectively [95,96]. In this thesis, I use “rjags” to call JAGS for all implementations using MCMC and the “coda” package to perform convergence diagnostics [97]. Once the model and prior distributions are specified in the JAGS language, the JAGS software automatically selects the appropriate samplers for each parameter or block of parameters.

### 2.5 INTEGRATED NESTED LAPLACE APPROXIMATIONS (INLA)

INLA was introduced in 2009 by Rue et al as an approximate method of Bayesian inference for latent Gaussian models [41]. INLA relies on approximations of the densities involved rather than sampling resulting in a much faster method than MCMC. Modifications have been proposed and integrated into the available software in 2013 by Martins et al, improving the speed of the methods [81]. Muff et al extended INLA for use in correcting for classical measurement error in an exposure within a generalized linear model when a replicate study is available [29]. They illustrated the method using data from the Framingham Heart Study and made a direct comparison with MCMC [1]. However, simulation studies were not performed and the method was not described for use with validation studies. Here I outline the INLA method using a replicates study, as described by Muff et al, and also explain how it can be extended for use when there is a validation study.

If one assumes  $Y$  and any other observed data to be conditionally independent given a set of latent Gaussian parameters  $\mathbf{v}$  and some additional hyperparameters  $\boldsymbol{\psi}_1$ , then one may assume  $\mathbf{v}$  is a Gaussian Markov random field (GMRF) dependent on further hyperparameters,  $\boldsymbol{\psi}_2$ . A GMRF is a Gaussian field wherein any two values (or set of values)  $v_a$  and  $v_b$  are conditionally independent given the remaining elements,  $\mathbf{v}_{-ab}$ .

Assuming the three-part conditional independence model as specified in Section 2.3, those parameters that would be assigned a Gaussian prior under the standard Bayesian approach,  $\beta_0, \boldsymbol{\beta}_Z, \alpha_0, \boldsymbol{\alpha}_Z$ , and  $X_i$ , are assigned to  $\mathbf{v}$  and the non-Gaussian parameters  $\boldsymbol{\psi} = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2\}$  include  $\beta_X, \sigma_U^2, \sigma_X^2|Z, \sigma_Y^2|XZ$  [29]. Of note,  $\beta_X$  is considered by INLA to be a non-Gaussian scale parameter, even though a Gaussian prior may be specified. This is because the regression term  $\beta_X X_i$  is the product of two parameters with latent Gaussian distributions so cannot itself also be Gaussian [41]. The posterior distribution from Equation 2.7 is then rearranged as:

$$2.11 \quad f(\mathbf{v}, \boldsymbol{\psi} | \mathbf{W}, Y, \mathbf{Z}) \propto f(\boldsymbol{\psi}) f(\mathbf{v} | \boldsymbol{\psi}) \prod_{i=1}^N f(\mathbf{W}_i, Y_i, \mathbf{Z}_i | \mathbf{v}_i, \boldsymbol{\psi}).$$

If a validation study is available, Equation 2.11 is extended to include the observed values of  $X$ ,  $X_O$ , as part of the observed data on which the posterior is conditioned,  $f(\mathbf{v}, \boldsymbol{\psi} | \mathbf{W}, Y, \mathbf{Z}, X_O)$ . The  $X_i$  in  $\mathbf{v}$  continues to include all values of  $X_i$  including those observed as a part of the validation study. Specifics of the implementation can be found in Section 2.5.2.

Rue et al demonstrated that the marginal posterior densities for each element of  $\mathbf{v}$  can be derived from the joint posterior density in terms of two nested densities,  $f(v_j | \boldsymbol{\psi}, \mathbf{W}, Y, \mathbf{Z})$  and  $f(\boldsymbol{\psi} | \mathbf{W}, Y, \mathbf{Z})$ , where  $v_j$  is the  $j^{\text{th}}$  element of  $\mathbf{v}$  [41]. The marginal posterior densities for each element of  $\boldsymbol{\psi}$ , where  $\psi_k$  is the  $k^{\text{th}}$  element, can be found from integrating out all the other components  $\psi_{-k}$  from  $f(\boldsymbol{\psi} | \mathbf{W}, Y, \mathbf{Z})$ . For  $f(v_j | \boldsymbol{\psi}, \mathbf{W}, Y, \mathbf{Z})$ , Rue et al described the process of using either a Gaussian or a Laplace approximation [41]. A Gaussian approximation was not considered to achieve enough accuracy and a Laplace approximation was too computationally intensive; therefore, Rue et al introduced a simplified Laplace approximation, based on the Taylor's series expansion up to the third order. This simplified Laplace approximation is less computationally intensive than a Laplace approximation but more accurate than a simple Gaussian approximation.

In Rue et al, the joint posterior density of the non-Gaussian parameters,  $f(\boldsymbol{\psi} | \mathbf{W}, Y, \mathbf{Z})$ , was evaluated on a regular grid and the resulting values were used to numerically compute the integrals required to obtain the marginal posterior densities for  $\psi_k$  [41]. I refer the interested reader to Martins et al, for the technical details [81]. For this to be computationally feasible, the number  $q$  of elements of  $\boldsymbol{\psi}$  had to be quite small ( $q \leq 6$ ). More recently, applications of INLA have estimated the marginal posterior of the hyperparameters by approximating it with a skewness-corrected Gaussian, as described by Martins et al. This improved method may accommodate up to 20 non-Gaussian parameters [43].

Further details of the approximation methods are detailed by Rue et al [41], and greater detail on the application of INLA to classical measurement error correction can be found in Muff et al [29]. The accuracy of the approximations will depend on whether  $f(\mathbf{v} | \boldsymbol{\psi}, \mathbf{W}, Y, \mathbf{Z})$  is sufficiently close to a GMRF to merit the simplified Laplace approximations. In the case of linear regression for the substantive model in our example, the method is exact.

In contrast to the MCMC approach, estimates obtained from INLA are not based on random sampling and hence estimates are identical each time the method is run.

### 2.5.1 Estimation of variance

As in MCMC methods, estimation of parameters is based on posterior means and posterior quantiles. The equi-tailed 95% CrI is determined from the approximated posterior distribution.



Unlike typical MCMC methods, INLA is also fast enough that bootstrapping may be used to estimate variance and confidence intervals.

### 2.5.2 Software

Rue et al have provided software for implementation on INLA. The INLA software is publicly available from [www.r-inla.org](http://www.r-inla.org) and can be accessed from an R package. Application of the methods within the software involves modular specification of the substantive model, the exposure model, and the error model. Prior distributions are also specified by the user. In these ways the use of the INLA software is similar to the use of JAGS.

The INLA software applies the simplified Laplace approximation described by Rue et al [41] to each Gaussian parameter,  $f(v_j|\boldsymbol{\psi}, \mathbf{W}, Y, \mathbf{Z})$ , and the skewness-corrected Gaussian described by Martins et al [81] to the non-Gaussian hyperparameters. An additional function (“hyperpar”) is available in the INLA R package to refine estimates of the posterior marginals of the hyperparameters via numerical integration over a fine grid; the grid spacing may be specified. This function was implemented by Muff et al [14]. However, this function requires significant additional time (longer than MCMC in some cases; see Chapter 3.5.4) and was not observed to improve estimates in the settings laid out in Chapter 3.2. Therefore, this function was not used here.

Code for the setting of classical measurement error with repeated measures was provided by Muff et al [14]. I extended the method from Muff et al to the setting where a validation study is available instead of a replicate study. R code detailing implementation of this extension is included in Appendix A. The Muff et al implementation with a replicate study and classical measurement error relies on stipulating the latent  $X$  as a random effect and describing the joint model as:

$$\begin{bmatrix} Y_1 & NA & NA \\ \vdots & \vdots & \vdots \\ Y_N & NA & NA \\ NA & 0 & NA \\ \vdots & \vdots & \vdots \\ NA & 0 & NA \\ NA & NA & W_{1,1} \\ \vdots & \vdots & \vdots \\ NA & NA & W_{1,N} \\ NA & NA & W_{2,1} \\ \vdots & \vdots & \vdots \\ NA & NA & W_{2,N} \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \beta_X \begin{bmatrix} 1 \\ \vdots \\ N \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \begin{bmatrix} NA \\ \vdots \\ NA \\ -1 \\ \vdots \\ -N \\ 1 \\ \vdots \\ N \\ 1 \\ \vdots \\ N \end{bmatrix} + \beta_Z \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \alpha_0 \begin{bmatrix} NA \\ \vdots \\ NA \\ 1 \\ \vdots \\ 1 \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \alpha_Z \begin{bmatrix} NA \\ \vdots \\ NA \\ \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix}$$

Each column of the outcome matrix represents a model that can be factored from the joint model (Equation 2.8). The model family can be specified in the INLA software for each column/model. Furthermore, the outcome matrix holds only observed data. The exposure model, assumed to be Gaussian as shown in Equation 2.9, is represented as  $0 = -X + \alpha_0 + \boldsymbol{\alpha}_Z \mathbf{Z}$  because the

unobserved latent  $X$  cannot be included as an outcome. The third column of the linear predictor is the latent  $X$  which is given labels from 1 to  $N$ . The “copy” function within the INLA software allows these labelled latent  $X$  values to be used alongside  $\beta_X$  for the substantive model.

In order to extend this implementation of INLA for measurement error correction for use with a validation study, a fourth column is added to the outcome matrix:

$$\begin{bmatrix} Y_1 & NA & NA & NA \\ \vdots & \vdots & \vdots & \vdots \\ Y_N & NA & NA & NA \\ NA & 0 & NA & NA \\ \vdots & \vdots & \vdots & \vdots \\ NA & 0 & NA & NA \\ NA & NA & W_{1,1} & NA \\ \vdots & \vdots & \vdots & \vdots \\ NA & NA & W_{1,N} & NA \\ NA & NA & NA & X_{OBS,1} \\ \vdots & \vdots & \vdots & \vdots \\ NA & NA & NA & X_{OBS,N1} \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \beta_X \begin{bmatrix} 1 \\ \vdots \\ N \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \begin{bmatrix} NA \\ \vdots \\ NA \\ -1 \\ \vdots \\ -N \\ 1 \\ \vdots \\ N \\ 1 \\ \vdots \\ N1 \end{bmatrix} + \beta_Z \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \alpha_0 \begin{bmatrix} NA \\ \vdots \\ NA \\ 1 \\ \vdots \\ 1 \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix} + \alpha_Z \begin{bmatrix} NA \\ \vdots \\ NA \\ Z_1 \\ \vdots \\ Z_N \\ NA \\ \vdots \\ NA \\ NA \\ \vdots \\ NA \end{bmatrix}$$

The data must be sorted before entry into the model to have the observed values of  $X$ ,  $X_{OBS}$ , at the start of the data set so that 1 to  $N1$ , where  $N1$  is the number of observed values of  $X$ , correspond to the first  $N1$  values in the data set. The “copy” function is used again to map the values of  $X$  which have observations to this model/column. Finally, the model for the  $X_{OBS}$  column is specified as “fixed=TRUE”.

## 2.6 MULTIPLE IMPUTATION (MI)

MI was first introduced by Rubin in 1987 as a method for handling missing data [44]. It has since been further developed and is well summarized in a book on the topic by Carpenter and Kenward [98]. Its increased popularity in the last decade is in part due to improvements in computing power and the availability of software to implement it in common statistical packages [99–101]. In the past few years, MI has also been suggested as a method for handling measurement error [5,47,48]. MI is derived from Bayesian principles, but is designed to produce inferences with good frequentist properties [37,102].

### 2.6.1 Application of multiple imputation to missing data

In a missing data context, MI is based on drawing imputed values of missing observations from the posterior distribution of the missing data conditional on the observed data including the outcome (the “imputation model”) [103]. By contrast, fully Bayesian methods use the entire joint posterior of the unknown parameters and unobserved values rather than just the conditional posterior of the missing observations. After making a number of such imputations, creating several complete data sets numbered  $m = 1, 2, \dots, M$ , the substantive model is fit to each imputed

data set estimating the parameter of interest,  $\hat{\beta}^{(m)}$ . Pooled parameter estimates,  $\hat{\beta}_{MI}$ , and their corresponding variance-covariance matrix, are obtained using Rubin's Rules [44,98]. Rubin's Rules for pooling parameter estimates are as follows:

$$2.12 \quad \hat{\beta}_{MI} = \frac{\sum_{m=1}^M \hat{\beta}^{(m)}}{M}$$

$$\widehat{var}(\hat{\beta}_{MI}) = \left[ \frac{1}{M} \sum_{m=1}^M \widehat{var}(\hat{\beta}^{(m)}) \right] + \left( 1 + \frac{1}{M} \right) \left[ \left( \frac{1}{M-1} \right) \sum_{m=1}^M (\hat{\beta}^{(m)} - \hat{\beta}_{MI})^2 \right].$$

The imputation model specified must be compatible or at least semi-compatible with the substantive model [46]. To be compatible, there must in theory exist a joint model for the two. The imputation model and the substantive model are semi-compatible if they can be made compatible by setting parameters in the imputation model to zero [104]. Therefore, one can include predictors in the imputation model that do not appear in the substantive model, but the imputation model must include all terms that appear in the substantive model. Finding a compatible imputation model when the substantive model contains non-linear effects or interactions of the partially missing variable can be quite challenging [46].

The imputation model is typically a generalized linear model (the dependent variable being the partially missing variable) including in the linear predictor the substantive model covariates, the outcome, and any additional predictive auxiliary values from the data [46]. A non-informative prior distribution  $f(\Omega)$  is usually specified to ensure frequentist properties in the posterior draws from the imputation model. Multiple partially missing variables can be imputed by performing the imputation step in a sequential manner, called multiple imputation by chained equations, or more recently, fully conditional specification [98,105].

In their 2013 book on MI [98], Carpenter and Kenward advise using either a Metropolis-Hastings sampling step or rejection sampling to draw imputed values from a model that is compatible with the substantive model when the functional form of the exposure in the substantive model is non-linear. Estimates from MI with rejection sampling would be expected to be asymptotically the same as those from any MCMC sampler. Bartlett et al presented a "substantive model compatible" method of MI for missing data using rejecting sampling in detail [46]. This method can be combined with fully conditional specification for handling multiple variables with missing values and is therefore termed substantive model compatible fully conditional specification (SMC-FCS). SMC-FCS further obviates the need for approximations for non-linear substantive models such as logistic regression or the Cox proportional hazards model. SMC-FCS will be described in detail for measurement error in the Section 2.6.2.

### 2.6.2 Application of multiple imputation to measurement error

Cole et al first proposed the use of MI to correct for measurement error [47]. Cole et al applied MI to a hypothetical cohort as a method of measurement error correction with a Cox proportional hazards model as the substantive model and a validation study available. Freedman et al published a simulation study with a main measurement subject to systematic error and two calibration measures subject to classical error to compare a similar MI implementation to RC and moment reconstruction when the substantive model is logistic or linear regression [48]. Keogh and White described an imputation model for measurement error accommodating a replicate study [5]. The properties of MI for measurement error correction have not been extensively investigated in simulation studies. To my knowledge, no publications have used MI for measurement error correction with real-world applications.

Standard methods of MI, including SMC-FCS, can be used directly when a validation study is present. When a validation study is available, the error-prone variable  $X$  can simply be treated as missing for those without an observed value using the error-prone measure  $W$  as a predictive auxiliary variable in the imputation model. When only a replicate study is available, all values of  $X$  are “missing” and the imputation model must be derived as a joint model of the measurement error model and the exposure model (Section 2.3).

In this thesis, I focus on the application of the SMC-FCS imputation for use in measurement error correction and present an extension for use with a replicate study.

#### **Box 2.3 SMC-FCS proposal distribution**

##### **Validation study:**

Assuming classical measurement error, if a validation study is available, the model for the proposal distribution for  $X_i$ ,  $f(X_i|\mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\gamma})$ , can be a linear regression with normally distributed errors, and parameters can be estimated by maximum likelihood.

##### **Replicate study:**

For a replicate study with exactly two replicates for all individuals, we assume  $f(X_i, \mathbf{W}_i|\mathbf{Z}_i)$  is multivariate normal. Under this assumption, it can be shown that the distribution of  $X_i$  given  $\mathbf{W}_i$  and  $\mathbf{Z}_i$  (with pdf  $f(X_i|\mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\gamma})$ ) is a normal distribution with mean  $\mu_{X_i}$  and variance  $\sigma_{X_i}^2$  given by Equations 2.4 and 2.5. Maximum likelihood estimates for  $E[X_i|\mathbf{Z}_i]$  and  $\sigma_{X_i|Z}^2$  (in Equations 2.4 and 2.5) are found by linear regression, using  $W_{i1}$  in the place of  $X_i$  during the first iteration and the accepted  $X_i^{t-1}$  values of the previous iteration thereafter. In the first iteration,  $\sigma_U^2$  is estimated as  $\frac{1}{2n} \sum_{i=1}^n (W_{i1} - \bar{W}_i)^2 + (W_{i2} - \bar{W}_i)^2$  where  $\bar{W}_i$  is the mean of  $W_{i1}$  and  $W_{i2}$ . In following iterations, the  $\bar{W}_i$  is replaced with  $X_i^{t-1}$  in this formula.

The target distribution from which we want to draw each  $X_i$ , that is, a substantive model compatible imputation model, is:

$$2.13 \quad f(X_i | \mathbf{W}_i, Y_i, \mathbf{Z}_i) = \frac{f(\mathbf{W}_i, X_i, Y_i, \mathbf{Z}_i)}{f(\mathbf{W}_i, Y_i, \mathbf{Z}_i)} \\ \propto f(Y_i | X_i, \mathbf{Z}_i; \boldsymbol{\beta}) f(X_i | \mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\gamma}),$$

where  $f(Y_i | X_i, \mathbf{Z}_i; \boldsymbol{\beta})$  is the substantive model and  $f(X_i | \mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\gamma})$  is our “proposal distribution” (Box 2.3) composed of the exposure model and measurement error model. Appropriate prior distributions must also be specified for  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}$ . The specification of the proposal distribution for a validation and replicate study can be found in Box 2.3. The SMC-FCS algorithm for use in the context of measurement error correction is outlined in Box 2.4. This entire procedure is then repeated to create as many imputed data sets as desired [106]. The substantive model is then fitted to each imputed data set and Rubin’s Rules are applied (Equation 2.12) [44].

#### Box 2.4 SMC-FCS algorithm for measurement error correction

For each iteration of the algorithm,  $t$ :

- 1) Draw parameter values for the proposal model,  $\boldsymbol{\gamma}^t$ .
- 2) Draw possible values of  $X_i, X_i^*$ , from the proposal distribution using  $\boldsymbol{\gamma}^t$  for all individuals for whom  $X$  is unobserved.
- 3) Draw parameters of the substantive model,  $\boldsymbol{\beta}^t$  (i.e. Equation 1.1), from their approximate joint posterior using maximum likelihood estimation based on the appropriate regression model and the specified prior distributions.
- 4) Determine whether the proposed value  $X^*$  for a given individual is a value from the target posterior distribution,  $f(X | \mathbf{W}, Y, \mathbf{Z}; \boldsymbol{\beta}, \boldsymbol{\gamma})$ , by drawing from the uniform distribution and satisfying the following rejection rule:

$$U(0,1) \leq \frac{f(Y_i | X_i^*, \mathbf{Z}_i; \boldsymbol{\beta}^t)}{c(Y_i, \mathbf{Z}_i; \boldsymbol{\beta}^t)}$$

**Logistic regression:**  $c(Y_i, \mathbf{Z}_i; \boldsymbol{\beta}^t)$  equals one and  $f(Y_i | X_i^*, \mathbf{Z}_i; \boldsymbol{\beta}^t)$  represents the predicted probability  $\Pr(Y_i = 1 | X_i^*, \mathbf{Z}_i; \boldsymbol{\beta}^t)$ .

**Linear regression:**  $c(Y_i, \mathbf{Z}_i; \boldsymbol{\beta}^t) = (2\pi\sigma_{Y|XZ}^{2,t})^{-0.5}$ , where  $\sigma_{Y|XZ}^{2,t}$  is the variance of the residual error at iteration  $t$ , and  $f(Y_i | X_i^*, \mathbf{Z}_i; \boldsymbol{\beta}^t)$  represents the deviance in the predicted  $Y_i^*$  value from the observed  $Y_i$ . Refer to Bartlett et al for derivations [18].

- 5) If accepted,  $X_i^*$  is retained as  $X_i^t$ . If  $X_i^*$  is rejected, a new  $X_i^*$  is drawn from the proposal distribution (Step 3); this is repeated until a value  $X_i^t$  is found for all  $X_i$  where a value is unobserved.

Repeat Steps 2-6 above until the drawn parameters ( $\boldsymbol{\gamma}^t$  and  $\boldsymbol{\beta}^t$ ) and values of  $X_i^t$  reach a stationary distribution. Impute final values of  $X_i^t$  into an imputed data set  $m$  as  $X_i^m$ ,  $m = 1, 2, \dots, M$ .

The proposal distribution is a posterior distribution and as such is based on both a prior distribution and the likelihood. Jeffrey’s prior, a conventional non-informative but improper prior (i.e. does not integrate to one), is used for the variances in the missing data setting. As discussed by Gustafson, use of improper priors for variance parameters in the measurement error setting, here  $\{\sigma_{X|Z}^2, \sigma_U^2, \sigma_{Y|XZ}^2\}$ , can lead to an improper posterior distribution [27]. In our implementation of MI, because the algorithm for a replicate study must infer  $\sigma_{X|Z}^2$  and  $\sigma_U^2$  from the model and there is a direct trade-off in the values of each, there is potential for a prior draw of zero resulting in an improper posterior distribution. Therefore, in a replicate study setting, proper inverse-Gamma priors are used for the variances in formulating the proposal distribution, just as is done in the MCMC approach.

As recommended by Gustafson [27], regression coefficients are formulated with locally uniform improper priors defined as multivariate normal distributions around a zero mean vector with a variance of infinity. While these do not lead to improper posterior distributions, in Chapter 3 it will be shown that with insufficient information from the likelihood such vague priors lead to biased posterior estimates with high variance due to lack of empirical identifiability in such a finite sample set. In the absence of such information from the data, the priors play a much stronger role in the posterior [27].

An MCMC sampler could, in theory, be used to generate sample values of  $X$  with proper priors specified for all parameters [37]. In that case, MI with MCMC would be expected to produce estimates of  $\beta_X$  similar to those from a fully Bayesian analysis. The Bayesian model specified for imputation of  $X$  would need to include both the imputation model and the substantive model; compatibility would be ensured by the acceptance step as described in Section 2.4.

To my knowledge, there have been no publications of MI for measurement error correction when the substantive model contains a non-linear functional form of an error-prone variable.

### ***Software***

The same software which can impute missing data may be used to correct for measurement error when a validation study is present, treating  $X$  as missing and including  $W_1$  as an auxiliary variable in the imputation model. Stata, SAS, and R all have packages for imputing missing data. In R, standard packages include “mi” and “mice” and the package using the method described by Bartlett et al is called “smcfcs” [107].

The R package “smcfcs” for multiple imputation of missing data using rejection sampling was employed without modification in the validation study scenario [37]. The code was adapted for use with replicate studies as no currently available package was suitable for this setting (Appendix A).

## 2.7 SUMMARY

The three methods from the Bayesian family of methods to be investigated in this thesis have been introduced in this chapter. In Chapter 3, these will be applied for correction of measurement error when the error-prone predictor is untransformed in the linear predictor in the substantive model. The modular nature of Bayesian analysis based on the three-part conditional independence structure means that these methods have flexibility which enables them to be extended to a more complex form for one or more of the component models, as will be considered in Chapters 4-7. I have described an extension of the INLA approach to handle validation studies. I have also described an MI approach for use with replicate studies. Example code for these extensions is available in Appendix A.

### 3 MEASUREMENT ERROR CORRECTION IN GENERALIZED LINEAR MODELS WHEN THE EXPOSURE IS UNTRANSFORMED

---

#### 3.1 AIMS & OVERVIEW

While each of the methods introduced in Chapter 2 have been demonstrated for measurement error, there are no large simulation studies comparing the methods in diverse settings. I aimed to conduct a simulation study focusing on GLMs, with continuous or binary outcomes, wherein the linear predictor is made up of a single continuous exposure subject to classical measurement error. Correction methods will be applied to estimate the effect of the true exposure on the outcome in the presence of either a validation study or a replicate study performed on all or part of the study population.

The measurement error correction methods (i.e. MCMC, INLA, MI, and RC) will further be applied to a motivating example from the Framingham Heart Study relating systolic blood pressure to subsequent risk of heart disease. This example has been used in other publications evaluating measurement error correction methods [1,29,108].

#### 3.2 SIMULATION STUDY DESIGN

**Box 3.1 Simulation study design**

Simulations per scenario: 400

Sample size:  $N = \{500, 2000\}$

True exposure:  $X_i \sim N(10, 1)$

Classical error model:  $W_{ij} = X_i + U_{ij}$ ,  $U_{ij} \sim N(0, \sigma_U^2)$ ,  $j = 1, 2$

Measurement error variance:  $\sigma_U^2 = \{0.25, 1, 4\}$

<p>Linear regression substantive model:</p> $Y_i = \beta_0 + \beta_X X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1)$	<p>Logistic regression substantive model:</p> $\Pr(Y_i = 1 X_i) = \beta_0 + \beta_X X_i$
<p>Effect size: <math>\beta_X = \{1.4, 2.0, 4.0\}</math></p>	<p>log ORs: <math>\beta_X = \{0.35, 0.7, 1.4\}</math></p>

Simulation studies were designed to investigate the performance of each of the measurement error correction methods outlined in Chapter 2 in a range of scenarios. The aim was to estimate  $\beta_X$ , the regression coefficient representing the association between the true exposure  $X$  and the outcome  $Y$  (Equation 1.1). Nine scenarios were generated by varying  $\beta_X$  and  $\sigma_U^2$  (Box 3.1; Table 3.1). No error-free covariates were included. For logistic regression, the intercept  $\beta_0$  was chosen to ensure



*Table 3.1 Values used for simulation studies*

Scenario	Sample Size	Measurement error variance	Linear Regression Coefficients		Logistic Regression Coefficients	
	$N$	$\sigma_U^2$	$\beta_0$	$\beta_X$	$\beta_0$	$\beta_X$
1	500 or 2000	0.25	-0.7	1.4	-5.7	0.35
2	500 or 2000	0.25	-1.0	2.0	-9.3	0.7
3	500 or 2000	0.25	-2.0	4.0	-16.3	1.4
4	500 or 2000	1	-0.7	1.4	-5.7	0.35
5	500 or 2000	1	-1.0	2.0	-9.3	0.7
6	500 or 2000	1	-2.0	4.0	-16.3	1.4
7	500 or 2000	4	-0.7	1.4	-5.7	0.35
8	500 or 2000	4	-1.0	2.0	-9.3	0.7
9	500 or 2000	4	-2.0	4.0	-16.3	1.4

approximately 10% of individuals had the outcome event,  $Y = 1$ . For logistic regression the effect sizes, or logORs, correspond to ORs of 1.4, 2.0, and 4.1. The measurement error variances correspond to attenuation factors (Equation 1.9) of 0.8, 0.5, and 0.2. An OR of 4.1 or attenuation factor of 0.2 would be considered very large/small in most realistic applications, but these values are considered here to assess the performance of the methods in extreme scenarios. 400 simulated data sets were generated for each of the nine scenarios using the two sample sizes.

Each of the correction methods was applied to the setting of a validation study and a replicate study. In both settings,  $W_1$  is available for all individuals. For a validation study the true exposure  $X$  is observed in a random sample of 10% or 30% of individuals ( $P = 0.1$  or  $P = 0.3$ ). For a replicate study  $W_2$  is observed in 100% or 30% of individuals ( $P = 1$  or  $P = 0.3$ ;  $P = 0.3$  performed only for  $N = 2000$  due to the challenging nature of that scenario).

I also applied an analysis in which the substantive model is fitted using the true  $X$ , as well as a naïve analysis using the first error-prone exposure measure  $W_1$ .

Specific prior values and other user-defined specifications to run the methods are described in Section 3.4.

### 3.3 CRITERIA FOR EVALUATING FIT

The performance of the methods was assessed in terms of bias in  $\hat{\beta}_X$ , the empirical 95% range of the  $\hat{\beta}_X$ , coverage (the percentage of simulations in which the 95% CI/CrI for  $\beta_X$  contained the true value), and root mean square error (RMSE) (Figures 3.1-3.8; Appendix G Figures G.1-G.6). The figures also indicate Monte Carlo error (MCE) 95% CIs for  $\hat{\beta}_X$  and coverage [109] estimated as follows:

$$3.1 \quad \text{MCE of } \hat{\beta}_X = \sqrt{\frac{\text{var}(\hat{\beta}_X)}{S}}$$

$$3.2 \quad \text{MCE of Coverage, } C = \sqrt{\frac{C(1-C)}{S}}$$

where  $S$  is the number of simulations and  $C$  is the estimated coverage. Bias was assessed in terms of whether the 95% CI constructed from the MCE includes zero. Discussions of bias focus on bias greater than 5% of  $\beta_X$ , but all bias estimates are displayed in the relevant figures. A method is considered here to have poor coverage if the MCE 95% CI for coverage excludes 95%.

### 3.4 IMPLEMENTATION OF METHODS

Example code for simulated data sets and implementation of each method can be found in Appendix A.

The naïve analysis is performed by replacing  $X$  in Equation 1.1 with  $W_1$  as described in Section 1.4.1. RC is performed as described in Section 2.2 (simple RC). The delta method is used to estimate standard errors (Equation 2.6).

Bayesian analysis using MCMC is performed using the JAGS software [62] called via the R package “rjags”. For each simulation, three MCMC chains were initialized. Appropriate burn-in to reach convergence to the stationary distribution is assessed within each simulation using  $\hat{R}$  as described in Section 2.4.1.  $\hat{R}$  was assessed every 25,000 and 50,000 samples for linear and logistic regression, respectively, until  $\hat{R}$  was  $< 1.1$  for  $\hat{\beta}_X$ , with a maximum burn-in of 375,000 and 450,000, respectively, per simulation. After burn-in, 50,000 samples, thinned 1:20, were drawn for inference for linear regression and 100,000 samples, also thinned 1:20, for logistic regression.

For both MCMC and INLA, the exposure model intercept,  $\alpha_0$ , was assigned a normal prior with mean zero and variance 1,000. Variance parameters,  $\sigma_X^2$ ,  $\sigma_U^2$ , and  $\sigma_{Y|X}^2$ , were each assigned inverse-Gamma priors,  $IG(1, 1)$ . For linear regression, vague priors of  $N(0, 10000)$  were used for the substantive model regression coefficients,  $\beta_0, \beta_X$ . For logistic regression, it is reasonable to assume that most logORs lie between -4 and 4 (OR: 0.02 and 55); therefore,  $N(0, 4.01)$  was assigned as the prior for  $\beta_X$  [41]. A vague prior of  $N(0, 10000)$  was used for the logistic model intercept,  $\beta_0$ .

INLA code for the setting of a replicate study was provided by Muff et al [14]. In order to perform simulations with validation data, I modified the code as described in Section 2.5.2 and illustrated in Appendix A. While Muff et al also applied numerical integration (via the function “hyperpar”) to the estimation of the non-Gaussian hyperparameters, this is not applied to the simulations.

To perform MI, the R package “smcfcs” for multiple imputation of missing data using rejection sampling was employed without modification in the validation study scenario using the Jeffrey’s prior for variance parameters and locally uniform improper priors for the substantive model regression coefficients (Section 2.6.2) [37]. The code was adapted for use with replicate studies as described in Section 2.6.2 using inverse-Gamma priors (i.e.  $IG(1, 1)$ ) for  $\sigma_X^2$  and  $\sigma_U^2$  (Appendix A).

Trace plots of a subset of simulations were used to determine the number of iterations of the SMC-FCS algorithm per imputed data set necessary to consistently reach convergence. For logistic regression, 200 iterations per imputation was stipulated as trace plots indicated all that would converge did so before 200 iterations. For linear regression, trace plots indicated that most simulations converged before 1,000 iterations; however, with a replicate study and large measurement error ( $\sigma_U^2 = 4$ ), convergence was sometimes unclear (see Section 3.5.3 for more details) and use of too many iterations resulted in bias away from the null. Therefore, 1,000 iterations of the algorithm per imputation was stipulated for linear regression. Ten imputations were specified for each simulation. Use of a greater number of imputations did not appear to significantly alter estimates.

## 3.5 SIMULATION STUDY RESULTS

### 3.5.1 Naïve analysis

Following what is expected from theory, the naïve estimates of  $\beta_X$  were biased towards zero by 20%, 50%, and 80% for measurement error variance 0.25, 1, and 4, respectively, exactly for linear regression and approximately for logistic regression. For linear regression, coverage was 0% in all settings (Figures 3.1-3.2, 3.5-3.6). For logistic regression, coverage was close to 0% when  $\sigma_U^2 = 4$  and  $\sigma_U^2 = 1$  (Figures 3.3-3.4, 3.7-3.8), but larger for  $\sigma_U^2 = 0.25$ : e.g. when  $N = 2000$ , coverage was 82%, 47%, and 2% for  $\beta_X = \{0.35, 0.7, 1.4\}$ , respectively.

### 3.5.2 Validation study

RC resulted in unbiased estimates for linear regression (Figure 3.1, Appendix G Figures G.1-G.2) except in the setting of a small sample size and a small validation study ( $N = 500$  and  $P = 0.1$ ) which is referred to here as the least information setting (Figure 3.2). Over-coverage in the 95% CIs based on the delta method was observed when the validation study was larger ( $P = 0.3$ ). For logistic regression, RC gave unbiased estimates and appropriate coverage when  $\beta_X = 0.35$  or  $0.7$  (Figures 3.3-3.4, Appendix G Figures G.3-G.4). Estimates were biased towards the null and under-coverage was observed when  $\beta_X = 1.4$  (i.e. OR = 4.1), as is consistent with the literature [20]. The bias and under-coverage increased in severity as measurement error

**Table 3.2. Non-convergence rate of MCMC sampling after maximum burn-in for logistic regression over 400 simulations by effect size ( $\beta_X$ ) and measurement error variance ( $\sigma_U^2$ ). In the replicate study setting, the proportion ( $P$ ) of the study participants with two error-prone measures is varied between 30% and 100% and in the validation study setting the proportion of study participants with the true exposure determined is varied between 10% and 30%. Sample size ( $N$ ) is varied between 500 and 2000.**

	N	P	$\beta_X = 0.35$			$\beta_X = 0.7$			$\beta_X = 1.4$		
			$\sigma_U^2:$			$\sigma_U^2:$			$\sigma_U^2:$		
			0.25	1	4	0.25	1	4	0.25	1	4
Replicate study	500	100%	0.00	0.01	0.35	0.01	0.02	0.42	0.01	0.18	0.68
	2000	100%	0.00	0.00	0.03	0.00	0.00	0.10	0.00	0.03	0.50
	2000	30%	0.00	0.00	0.22	0.00	0.01	0.40	0.00	0.14	0.75
Validation study	2000	30%	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.02	0.06
	2000	10%	0.00	0.00	0.10	0.00	0.02	0.14	0.00	0.16	0.37
	500	30%	0.00	0.02	0.06	0.01	0.02	0.09	0.04	0.14	0.27
	500	10%	0.01	0.11	0.25	0.02	0.13	0.34	0.13	0.35	0.59

increased. While the bias was approximately the same regardless of sample size ( $N = 2000$  or  $N = 500$ ), the under-coverage was more pronounced for  $N = 2000$ .

The RMSE for RC was roughly 2-10 times higher than the other methods for linear regression and slightly elevated over the other methods for logistic regression for most scenarios. This ratio was more exaggerated as the effect size was increased.

MCMC gave unbiased estimates for linear regression (Figure 3.1-3.2, Appendix G Figures G.1-G.2). Estimates were unbiased for logistic regression (Figures 3.3, Appendix G Figures G.3-G.4) except in the least information setting (Figure 3.4). A large proportion of simulations in the least information setting had not converged according to our criteria (Gelman-Rubin convergence diagnostic  $\hat{R} < 1.1$ , Section 2.4.1) after reaching the designated maximum burn-in (Table 3.2). For  $N = 500$  with validation data on only 10% of participants ( $P = 0.1$ ), those that had not converged ranged from 1% of 400 simulations in the scenario with the lowest effect size and measurement error, to 59% in the scenario with the largest effect size and measurement error. The bias observed in this low-information setting was just at or under 10% of  $\beta_X$ . For  $N = 500$  and  $P = 0.3$ , 27% of 400 simulations had not converged for the largest effect size and measurement error; in this setting, bias was just at or under 5% of  $\beta_X$ . For  $N = 2000$  and  $P = 0.3$ , only 6% had yet to converge for the largest effect size and measurement error and no bias was observed. Coverage was consistent with the nominal level throughout, and RMSE was low compared with the other methods.

MI resulted in unbiased estimates for linear regression but had mild under-coverage (above 90%) (Figures 3.1-3.2, Appendix G Figures G.1-G.2). For logistic regression, MI gave unbiased estimates (Figures 3.1, Appendix G Figures G.1-G.2) except in the least information setting (Figure 3.2). Though, for  $N = 500$  and  $P = 0.3$ , bias of roughly 10% of  $\beta_X$  was observed for  $\beta_X = 1.4$ . In all logistic regression settings, MI had coverage at the nominal 95% level. RMSE

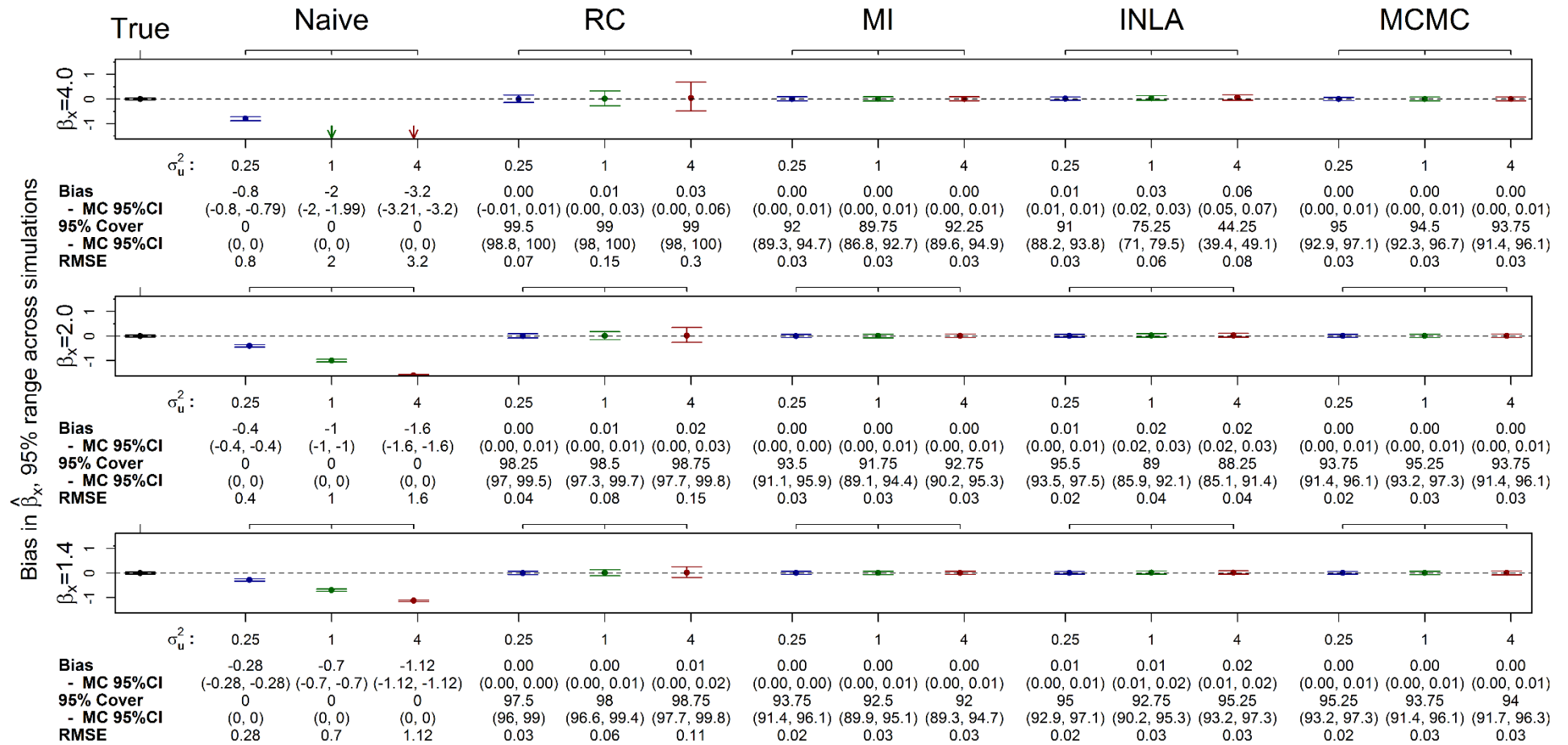


Figure 3.1. Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure  $X$  is observed for 30% ( $P = 0.3$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

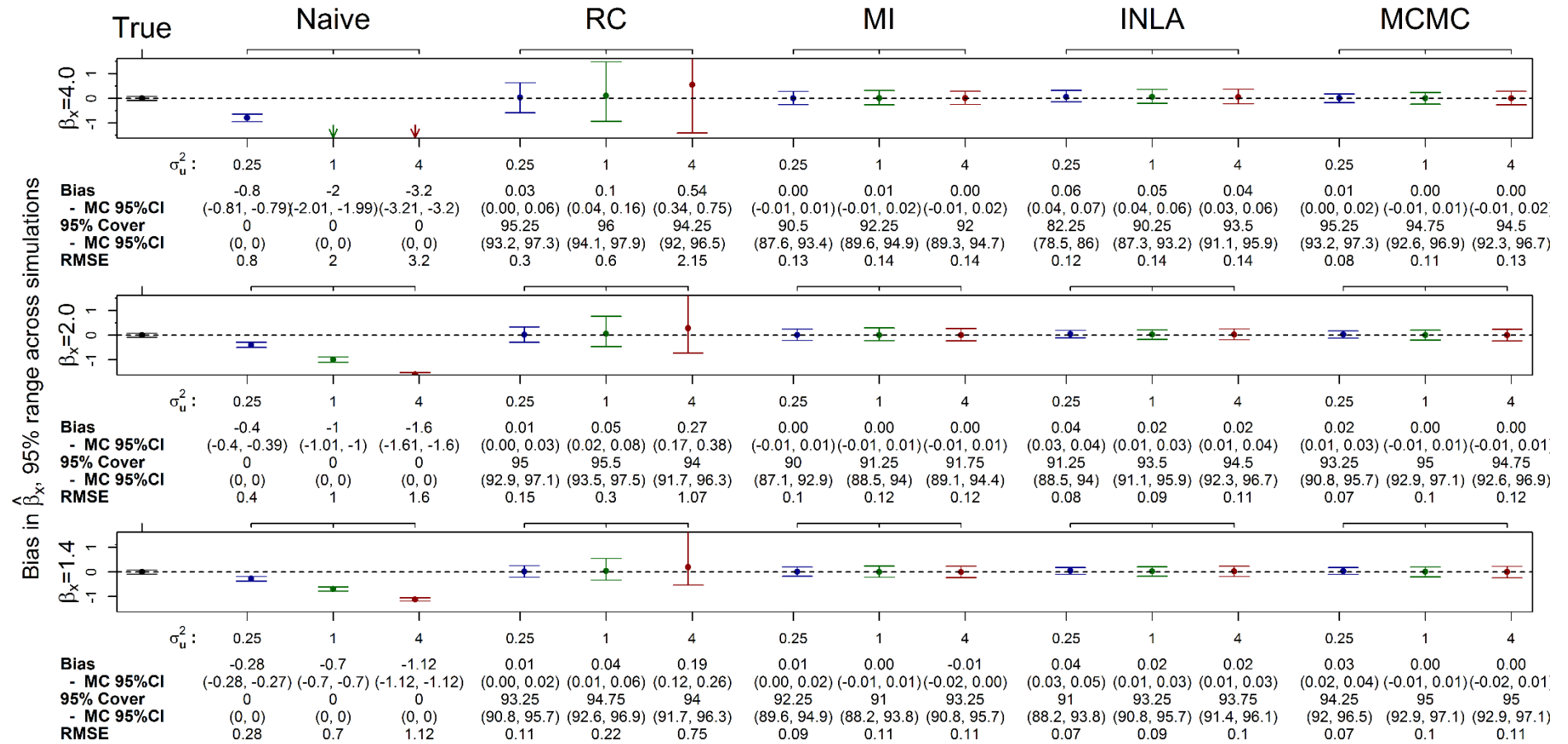


Figure 3.2 Simulation study results in the linear regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure  $X$  is observed for 10% ( $P = 0.1$ ) of the study participants, i.e. the least information setting. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

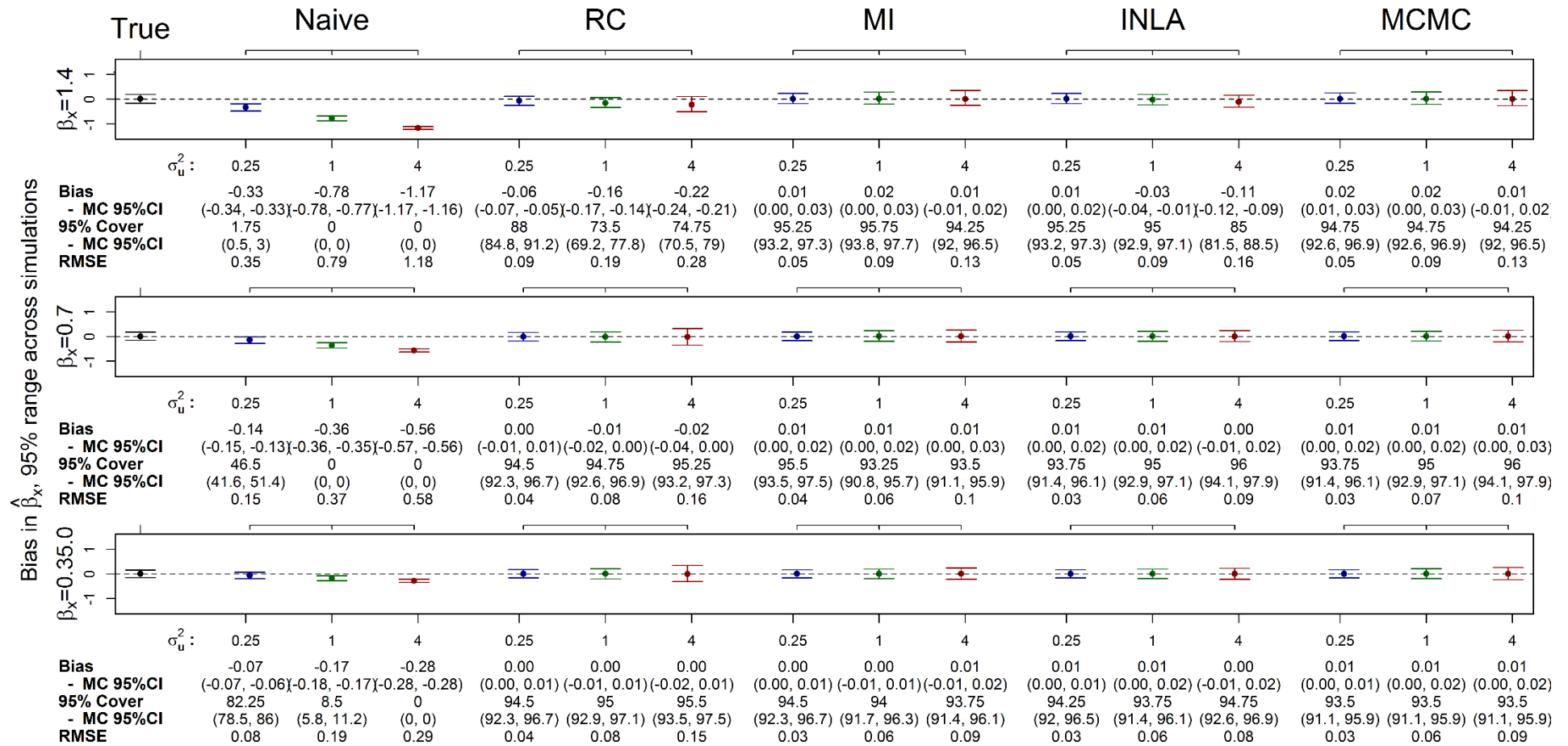


Figure 3.3. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure  $X$  is observed for 30% ( $P = 0.3$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

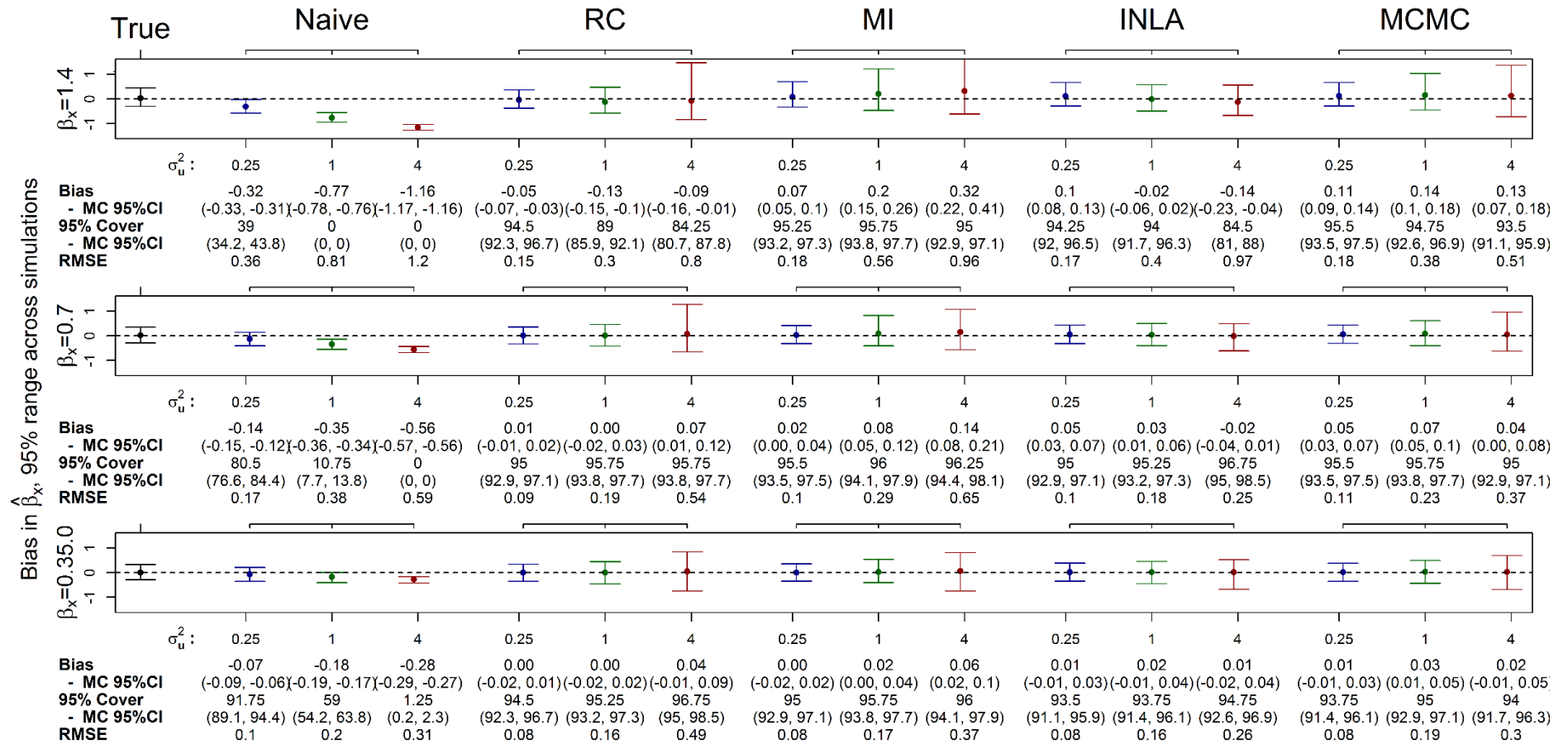


Figure 3.4 Simulation study results in the logistic regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure  $X$  is observed for 10% ( $P = 0.1$ ) of the study participants, i.e. the least information setting. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).



was similar to that of MCMC or just above for both linear and logistic regression (except for the low-information setting for the latter).

INLA gave unbiased estimates for linear regression but had under-coverage which was more severe for higher sample size ( $N=2000$ ) than lower ( $N = 500$ ) and increased in severity as  $\beta_X$  increased from 2.0 to 4.0 and as measurement error increased (Figures 3.1-3.2, Appendix G Figures G.1-G.2). Under-coverage, following the same trend, was seen for logistic regression as well (Figures 3.3-3.4, Appendix G Figures G.3-G.4). INLA gave unbiased estimates for most scenarios in logistic regression, but had significant bias when  $\beta_X = 1.4$  and  $\sigma_U^2 = 4$ , particularly when the sample size was large and the validation study size was small. In general, RMSE was similar to that of MCMC but was elevated in the scenarios with biased estimates. Bootstrapping was also performed with INLA using 1,000 bootstrap samples for the first 100 simulations with no improvement in coverage observed.

There were a few trends specific to the least-information setting (Figures 3.2 and 3.4). RC had large bias away from the null for linear regression with high measurement error ( $\sigma_U^2 = 4$ ) for all values of  $\beta_X$  (bias of 0.19, 0.27, and 0.54 for  $\beta_X = 1.4, 2.0,$  and  $4.0$ , respectively). For logistic regression there was very large variability in the RC estimates resulting in RMSE values that were sometimes larger than those from the naïve method; though, for  $\beta_X = 0.35$  or  $0.7$  the confidence intervals accurately reflected this decrease in precision. For logistic regression, MI had large bias away from the null when  $\sigma_U^2 = 1$  or  $4$  (bias 10% or 20% of  $\beta_X$ , respectively). MI also had high variability in  $\hat{\beta}_X$  for these scenarios ( $\sigma_U^2 = 1$  or  $4$ ) resulting in RMSE near to or higher than the naïve estimates (again accurately reflected in wider confidence intervals). For INLA, only the most extreme scenario for logistic regression ( $\beta_X = 1.4$  and  $\sigma_U^2 = 4$ ) resulted in bias towards the null (10% of  $\beta_X$ ) and under-coverage (85%); otherwise, it had the lowest RMSE of the methods in this setting.

Overall, all methods give unbiased estimates for linear regression, except for RC in the least information setting. Only MCMC had proper coverage in all scenarios. For logistic regression, except for the least information setting, MCMC and MI are unbiased with proper coverage and low RMSE. INLA gave unbiased estimates in all settings and had the greatest efficiency gains among the methods for all logistic regression scenarios except when  $\beta_X = 1.4$  and  $\sigma_U^2 = 4$ .

### 3.5.3 Replicate study

Results for a replicate study (Figures 3.5-3.8, Appendix G Figures G.5-G.6) were very similar to those for a validation study for RC and had similar trends for INLA and MCMC. MI had the starkest differences in performance between replicate study and validation study settings. Performance for each method, where different from that with a validation study, is discussed below.

For linear regression when RC was applied, when sample size was low ( $N = 500$ ) and measurement error was high ( $\sigma_U^2 = 4$ ), significant bias away from the null was observed with commensurate high RMSE (Figure 3.6). Over-coverage was observed for all other linear regression scenarios where replicate measures on all participants ( $P = 1$ ) were available (Figure 3.5).

For linear regression, MCMC had the lowest RMSE of the methods (Figures 3.5-3.6, Appendix G Figure G.5). However, unlike with a validation study, for linear regression, MCMC had some over-coverage when the measurement error variance or the  $\beta_X$  value was high. For logistic regression (Figures 3.7-3.8, Appendix G Figure G.6), non-convergence was higher than seen for a validation study when the measurement error variance was high and either the sample size was low (35% for  $\beta_X = 0.35$  and 68% for  $\beta_X = 1.4$ ) or the proportion with replicate measures was low ( $P = 0.3$ ) (22% for  $\beta_X = 0.35$  and 75% for  $\beta_X = 1.4$ ).

Estimates from INLA had larger bias in the replicate study setting than in the validation study setting, whenever the measurement error was high ( $\sigma_U^2 = 4$ ) (Figures 3.5-3.8, Appendix G Figures G.5-G.6). There was also significant bias for scenarios with  $\sigma_U^2 = 1$  when the sample size was reduced ( $N = 500$ ).

For lower values of measurement error, all linear regression MI simulations converged to the stationary distribution (Figures 3.5-3.6, Appendix G Figure G.5). For linear regression when the measurement error is large ( $\sigma_U^2 = 4$ ), the MI algorithm does not always appear to clearly converge to a stationary distribution; detailed investigations suggest that in some cases the sampled parameter values are rising incrementally upwards and not appearing to be settling down to a particular value. In a sensitivity analysis in this setting,  $\beta_X$  estimates were observed to continually increase as the number of iterations increased. Examination of a trace plot for every simulation to identify convergence was not possible; therefore, for scenarios with high measurement error, there is slight bias away from the null when  $\beta_X = 1.4$  and slight bias towards the null when  $\beta_X = 4$ . This pattern is observed for all replicate study settings (Figures 3.5-3.6, Appendix G Figure G.5).

For logistic regression, when the measurement error was large ( $\sigma_U^2 = 4$ ), MI was also observed to have wildly varying estimates of  $\beta_X$  (Figures 3.7-3.8, Appendix G Figure G.6). For the higher sample size when all participants had a replicate ( $N = 2000, P = 1$ ), only a small bias away from the null was observed for scenarios with  $\sigma_U^2 = 4$  (Figure 3.7). When the proportion of participants with replicate measures was dropped to 30% (Appendix G Figure G.6), scenarios with lower effect sizes ( $\beta_X = 0.35$  or  $0.7$ ) had bias away from the null of roughly 20% of the effect size and the scenario with  $\beta_X = 1.4$  gave estimates with wild variability (RMSE = 37) and bias 10-fold the effect size (toward the null, greater than bias in the naïve estimate). Finally, when

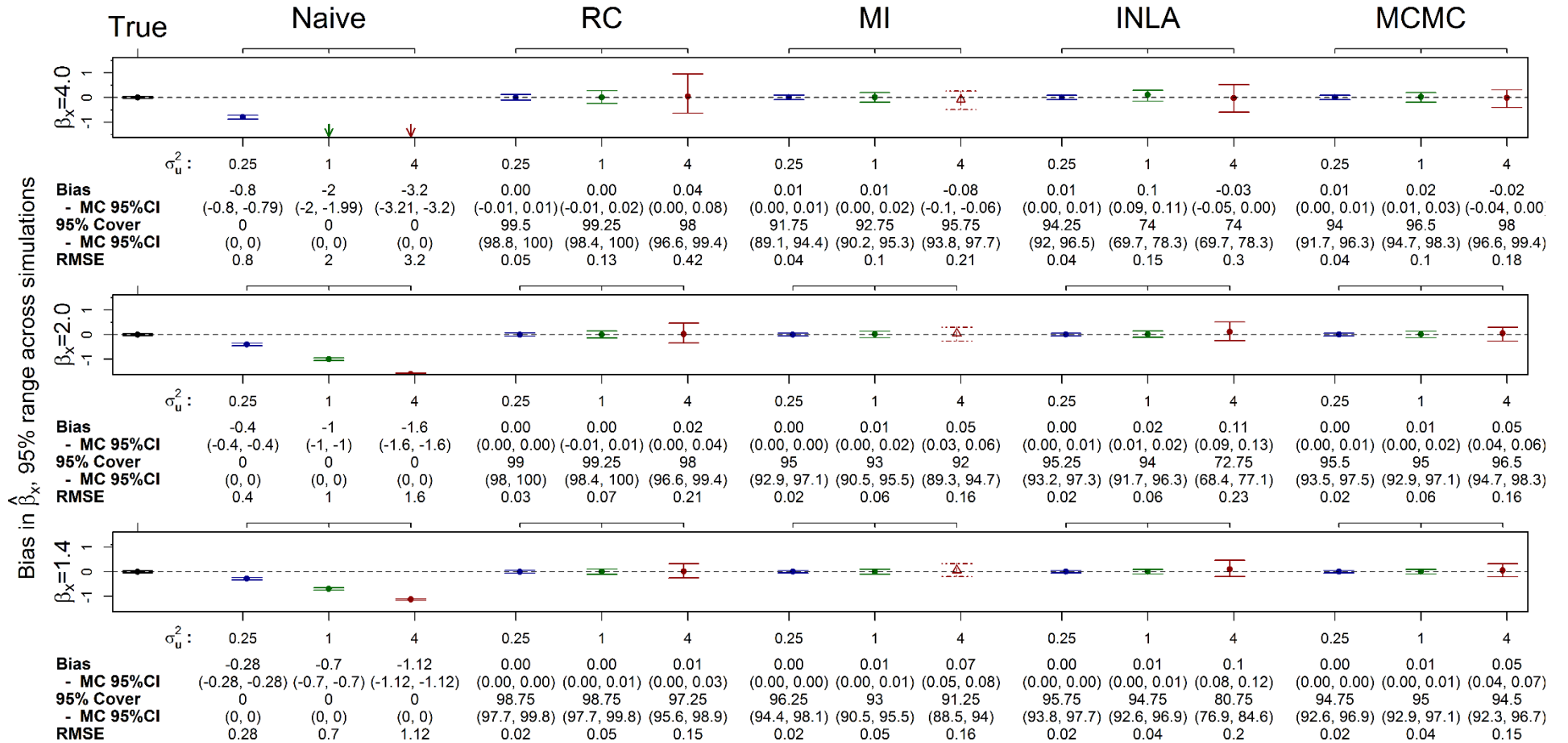


Figure 3.5. Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ). Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_x$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_x$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_x$ ,  $\hat{\beta}_x$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE). Scenarios where some simulations failed to converge are indicated with a dotted line and a triangle point estimate.

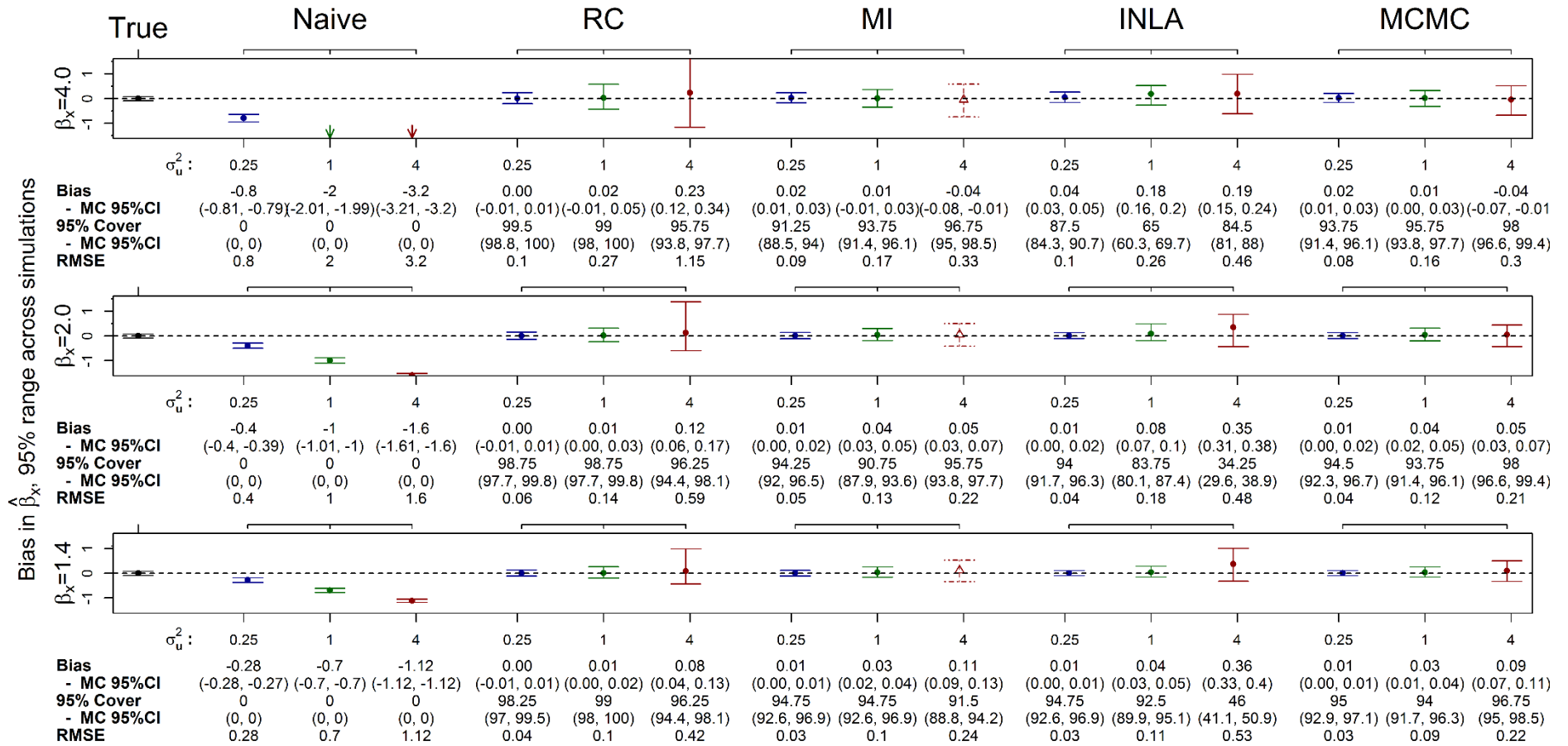


Figure 3.6. Simulation study results in the linear regression setting with sample size ( $N$ ) of 500 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ). Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_x$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_x$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_x$ ,  $\hat{\beta}_x$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE). Scenarios where some simulations failed to converge are indicated with a dotted line and a triangle point estimate.

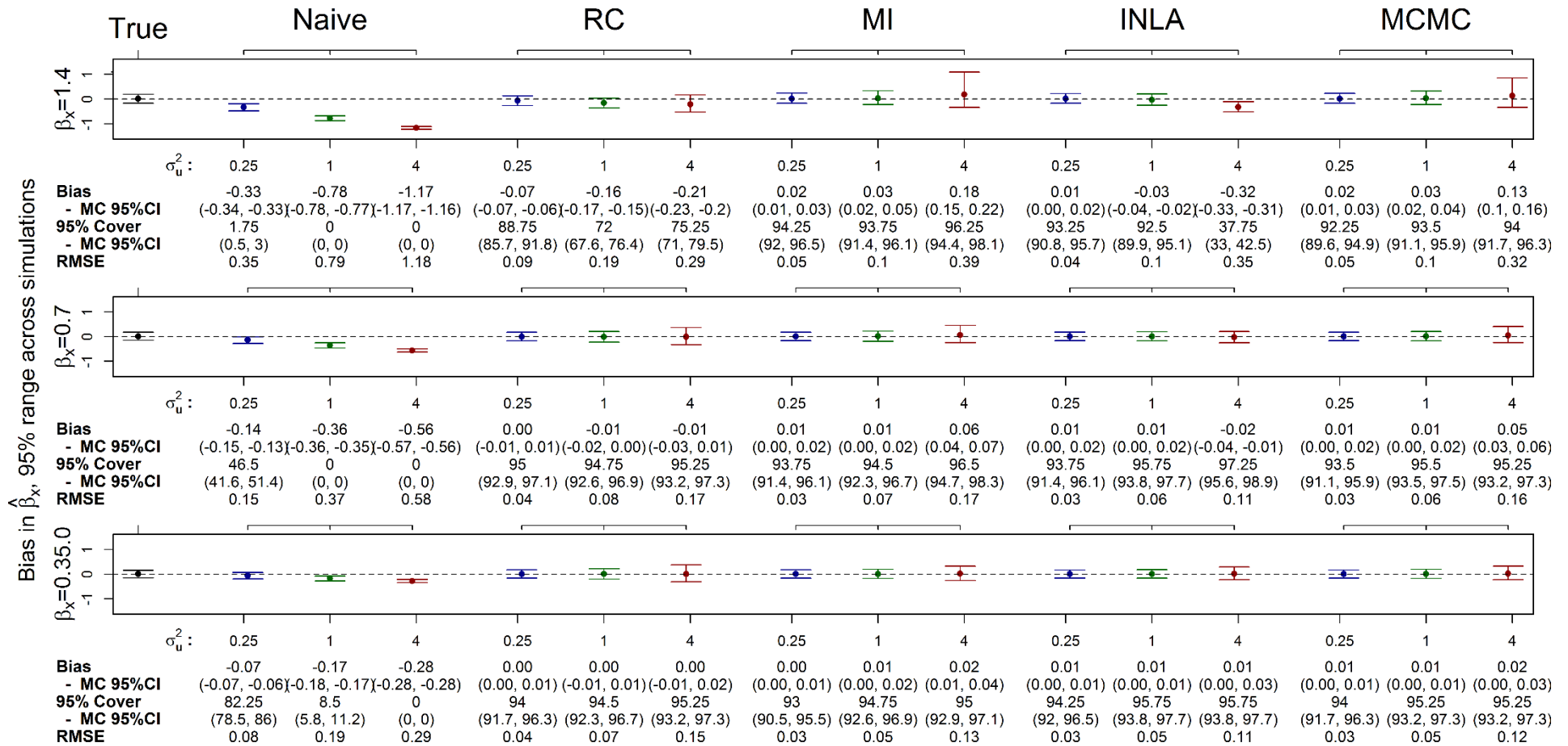


Figure 3.7. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ). Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_x$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_x$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_x$ ,  $\hat{\beta}_x$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

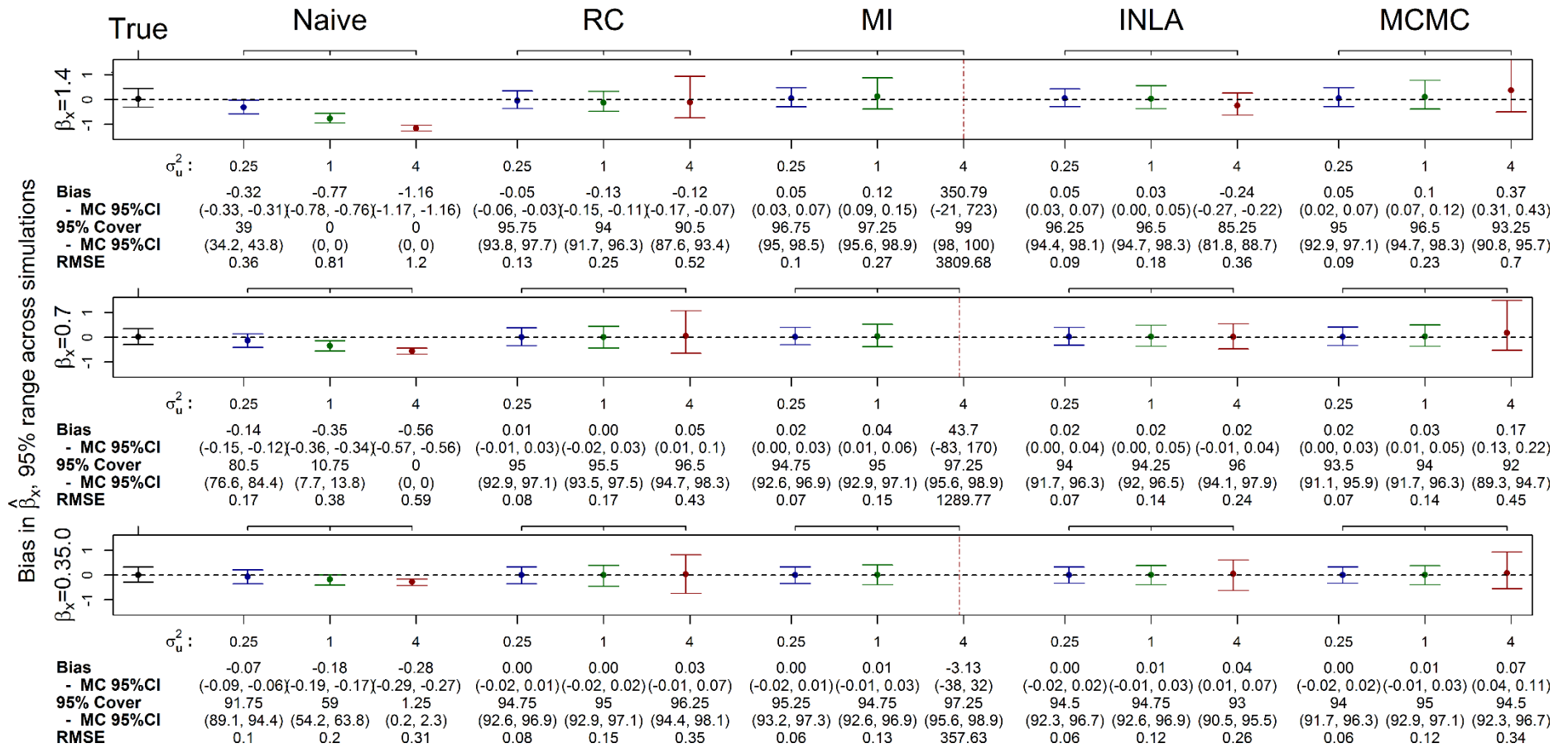


Figure 3.8. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 500 and a replicate study in which a replicate measure is available for all participants ( $P = 1$ ). Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. The association between the true predictor and the outcome,  $\beta_x$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_x$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_x$ ,  $\hat{\beta}_x$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE). Scenarios where some simulations failed to converge are indicated with a dotted line and a triangle point estimate.

sample size was smaller ( $N = 500$ ; Figure 3.8), this implementation of MI frequently failed to converge to reasonable estimates.

In order to understand whether this was a property of MI or a result of the uninformative priors used for MI, given that very little information was obtained from the likelihood at the lower sample size, a sensitivity analysis was performed with MCMC with extremely vague priors. With regression coefficient Gaussian priors of  $1 \times 10^9$ , the most challenging scenario was run (i.e.  $N = 500, \sigma_U^2 = 4, \beta_X = 1.4$ ). The resulting mean MCMC estimate of  $\beta_X$  was 60-fold the true effect size ( $\hat{\beta}_X = 84$ ; RMSE = 100). When the sample size was raised to  $N = 2000$ , this effect disappeared ( $\hat{\beta}_X = 1.56$ ; RMSE = 0.33).

When all individuals had a replicate measure, MI had negligible bias when  $\sigma_U^2 = 0.25$  or 1, the more common scenarios, for linear regression, and RMSE similar to or slightly above that of MCMC (slightly lower than for RC) (Figures 3.5-3.6). When the proportion with a replicate measure was reduced to 30% (Appendix G Figure G.5), bias less than 10% of  $\beta_X$  was seen for all values of measurement error. Similarly, for logistic regression, when  $N = 2000$  and all individuals had a replicate measure, MI had negligible bias when  $\sigma_U^2 = 0.25$  or 1 (Figure 3.7). However, when either sample size was reduced to  $N = 500$  or the proportion with a replicate measure reduced to 30% ( $P = 0.3$ ), bias less than 10% of  $\beta_X$  was observed for  $\sigma_U^2 = 1$  (Figures 3.8, Appendix G Figure G.6).

All methods produced estimates which were unbiased or had negligible bias (<5% of  $\beta_X$ ) when measurement error was small ( $\sigma_U^2 = 0.25$ ) for all settings. For linear regression, MCMC is the most reliable method across the scenarios and simulations with the lowest RMSE, no bias or small bias ( $\leq 6\%$  of  $\beta_X$ ), and over-coverage or coverage at the nominal level. For logistic regression, INLA has the lowest RMSE except when  $\beta_X = 1.4$  and  $\sigma_U^2 = 4$ .

### 3.5.4 Computing time

Because reducing computation time is an overarching goal of INLA and an important consideration to researchers, I explored the running time of each method. For each scenario, the average run-time was calculated from the first 100 simulations (Table 3.3).

Each simulation was run on a single computing core running at 2.3-2.5 Ghz given 5GB of virtual memory. All were run from R version 3.3.1. For INLA and MCMC, R packages used version 0.0-1468872408 of the INLA software and JAGS version 3.4.0, respectively.

RC ran at a fraction of a second for all scenarios. For linear regression MCMC via JAGS had an average time per scenario ranging from 40 seconds to 30 min. MI was slower than MCMC for linear regression, requiring 6 min to 74 min depending on the scenario. INLA as implemented in

*Table 3.3. Time per single simulated data set as an average of 100 simulations (minutes) in each setting for each method: multiple imputation (MI), integrated nested Laplace approximations (INLA), and Markov chain Monte Carlo (MCMC) sampling from a Bayesian model. For each simulation setting, i.e. for a study population of size  $N$  with a validation study where the true exposure,  $X$ , is observed in a proportion  $P$  of the participants or a replicate study in which a replicate measure is available for all participants ( $P = 1$ ), the effect size,  $\beta_X$ , and the measurement error variance,  $\sigma_U^2$ , are varied.*

Linear regression			$\beta_X$	MI			INLA			MCMC		
$N$	$\sigma_U^2$		0.35	0.7	1.4	0.35	0.7	1.4	0.35	0.7	1.4	
Validation study $P = 0.3$	500	0.25	5.9	7.0	11.4	0.3	0.3	0.3	0.6	0.6	0.6	
		1	7.2	9.2	18.6	0.3	0.3	0.4	0.6	0.6	0.6	
		4	8.4	11.5	24.1	0.3	0.3	1.1	0.6	0.6	0.6	
	2000	0.25	21.3	25.2	39.7	2.2	2.2	2.4	7.8	7.9	7.2	
		1	26.1	33.9	55.4	2.0	2.0	3.3	7.3	7.4	7.2	
		4	30.8	39.2	82.0	2.1	2.1	4.3	7.6	6.6	5.7	
Validation study $P = 0.1$	500	0.25	7.1	8.5	13.8	0.3	0.3	0.3	0.7	0.7	0.7	
		1	8.8	11.4	20.7	0.3	0.3	0.4	5.4	0.7	0.7	
		4	10.1	13.8	27.1	0.3	0.3	0.4	6.5	0.7	0.7	
	2000	0.25	28.3	30.3	46.1	2.0	2.0	2.2	7.2	8.1	8.3	
		1	32.8	41.5	75.1	2.0	2.0	2.1	8.2	8.2	7.7	
		4	37.3	47.1	93.4	2.1	2.1	10.8	8.1	7.4	8.2	
Replicate study	500	0.25	6.7	7.9	12.7	0.2	0.2	0.2	0.8	0.8	0.8	
		1	8.6	11.2	17.0	0.2	0.2	0.3	0.8	1.0	1.2	
		4	10.7	12.5	13.3	0.3	0.3	0.3	1.7	2.0	2.8	
	2000	0.25	23.3	25.9	40.6	0.8	0.8	0.9	7.0	8.4	8.2	
		1	27.5	34.5	62.2	0.8	0.8	0.9	8.4	10.0	22.6	
		4	31.5	38.7	73.8	0.8	1.3	1.0	26.7	33.3	32.8	
Logistic regression			$\beta_X$	MI			INLA			MCMC		
$N$	$\sigma_U^2$		0.35	0.7	1.4	0.35	0.7	1.4	0.35	0.7	1.4	
Validation study $P = 0.3$	500	0.25	0.8	0.8	0.8	0.3	0.3	0.3	3.6	3.7	3.7	
		1	0.8	0.8	0.7	0.3	0.3	0.3	3.7	3.8	3.8	
		4	0.8	0.8	0.7	0.3	0.3	0.4	3.9	4.1	3.9	
	2000	0.25	1.8	2.0	1.9	1.9	2.0	2.1	26.8	27.3	25.1	
		1	1.9	2.0	1.9	2.0	2.1	2.2	28.6	26.3	27.0	
		4	1.9	2.0	1.8	2.1	2.3	2.4	29.5	28.4	27.1	
Validation study $P = 0.1$	500	0.25	0.9	0.8	0.8	0.3	0.3	0.3	4.7	4.8	4.9	
		1	0.9	0.8	0.8	0.3	0.3	0.3	5.2	5.4	5.8	
		4	0.9	0.8	0.7	0.3	0.3	0.3	6.3	5.9	7.3	
	2000	0.25	2.2	2.3	2.1	1.9	2.1	2.2	33.1	31.8	31.3	
		1	2.3	2.3	2.0	1.9	2.1	2.2	35.5	35.2	37.2	
		4	2.2	2.2	1.9	2.1	2.3	2.4	46.4	45.0	49.7	
Replicate study	500	0.25	1.0	1.0	1.0	0.1	0.2	0.2	6.2	6.1	6.2	
		1	1.0	1.0	0.9	0.2	0.1	0.2	6.1	6.2	6.3	
		4	1.0	1.0	0.9	0.2	0.2	0.2	8.8	9.2	9.4	
	2000	0.25	2.8	2.7	2.5	0.6	0.6	0.6	35.5	32.3	32.9	
		1	2.6	2.6	2.4	0.6	0.6	0.6	35.9	32.4	34.9	
		4	2.6	2.6	2.5	0.6	0.6	0.6	38.2	43.0	61.0	



this paper required at most 11 min on average for any linear regression scenario, but for most scenarios the average time ranged between 20 seconds and 3 min.

Only MCMC required more time for logistic regression than for linear regression, ranging from 4 min to an hour across scenarios. Whereas for MI and INLA, the average time for a run was under 3 min for all logistic regression scenarios.

### 3.6 APPLICATION IN THE FRAMINGHAM HEART STUDY

The methods presented in Chapter 2 were applied to investigate the association between an individual's underlying 'usual' SBP and subsequent risk of CHD using data from the Framingham Heart Study. SBP is subject to random fluctuations within an individual from day to day, which we assume can be modelled as classical measurement error.

The Framingham Heart Study was initiated in 1948 to identify common factors leading to CHD [66]. The study initially enrolled 4,641 individuals aged 40–69 (44% male) [67]. This example has the advantage of having been used by Carroll et al and Muff et al for illustrating their implementations of MCMC or INLA for measurement error correction [1,29]; to aid comparison to these publications, the same model and subset of data is used here.

The subset included only men aged 45 or over at the second exam with a serum cholesterol between 200 and 300 at the third exam ( $N = 641$ ). At the second and third study exam (two and four years after baseline, respectively), two separate measures of each individual's SBP were taken. At the sixth exam, eight years from the first SBP measurement and 10 years after baseline, incidence of CHD was established. There were 78 incidences of CHD observed. CHD is modelled as a binary outcome in a logistic regression model. Binary smoking status, determined at the 1<sup>st</sup> exam and assumed to be measured without error, is included as a potential confounder.

As in Carroll et al [1] and Muff et al [29],  $X$  denotes usual SBP on the log scale and centered at 50,  $(\log(SBP - 50))$ , and  $W_1$  and  $W_2$  denote the error-prone measures taken from health examinations two and three (each based on the mean of two SBP measurements taken at the same exam by two separate technicians). The transformed observations appeared to be approximately normally distributed. A Bland-Altman plot did not find any systematic differences between  $W_1$  and  $W_2$  and few outliers. The substantive model of interest is a logistic regression of  $Y$  (CHD incidence) on  $X$  (usual SBP) and  $Z$  (smoking status):  $\text{logit}(\Pr(Y = 1|X, Z)) = \beta_X X + \beta_Z Z$ .

The analysis began with an exploration of the degree of measurement error. The posterior means of  $\hat{\sigma}_{X|Z}^2$  and  $\hat{\sigma}_Y^2$  from MCMC were 0.05 and 0.017, respectively, resulting in a measurement error variance 34% of the variance of  $X|Z$ , or equivalently an attenuation factor of  $\hat{\lambda}=0.75$ . By simply regressing  $W_2$  on  $W_1$  controlling for  $Z$  we obtain  $\hat{\lambda} = 0.74$ .

**Table 3.4. Estimate of the association between underlying systolic blood pressure (SBP) and coronary heart disease (CHD),  $\beta_X$ , adjusting for smoking in logistic regression after the application of the following methods: Mean, true SBP is replaced by the mean of the two measurements of SBP,  $W_1$  and  $W_2$ ; Naïve, true SBP is replaced by a single measurement,  $W_1$ ; regression calibration (RC); multiple imputation (MI); Markov chain Monte Carlo (MCMC) sampling from a Bayesian model; and integrated nested Laplace approximations (INLA). For frequentist methods, standard errors,  $\widehat{SE}(\hat{\beta}_X)$ , and the 95% confidence intervals (CI) are given; for Bayesian methods, standard deviation of the posterior samples of  $\hat{\beta}_X$ ,  $\widehat{SD}(\hat{\beta}_X)$ , and 95% credible intervals (CrI) are given.**

	Mean	Naïve	RC	MI		MCMC	INLA
$\hat{\beta}_X$	1.66	1.47	1.97	1.97	$\hat{\beta}_X$	1.80	1.81
$\widehat{SE}(\hat{\beta}_X)$	0.49	0.46	0.63	0.60	$\widehat{SD}(\hat{\beta}_X)$	0.57	0.56
<b>95% CI</b>	(0.69, 2.6)	(0.56, 2.4)	(0.74, 3.2)	(0.81, 3.1)	<b>95% CrI</b>	(0.62, 2.9)	(0.71, 2.9)

Each of the methods was then applied to the Framingham subset (Table 3.4). For MI, a trace graph of  $\hat{\beta}_X$  was used to determine that 50 cycles of the algorithm were required for convergence. Furthermore, 100 imputed data sets were used to ensure that stable parameter estimates were reached. The naïve estimate of the logOR,  $\hat{\beta}_X$ , was 1.47 (SE: 0.46). The point estimate can be improved by regressing on the mean of  $W_1$  and  $W_2$  ( $\hat{\beta}_X = 1.66$ ) but the variance is still underestimated and likely the point estimate is still biased towards the null. RC and MI have the highest values of  $\hat{\beta}_X$  at 1.97 (SE: 0.63 and 0.60, respectively) and INLA and MCMC have  $\hat{\beta}_X$  values of 1.81 and 1.80, respectively (SD: 0.57 and 0.56).

Assuming the structure outlined in Section 2.3 with the substantive model taking the form of a logistic regression, the maximum likelihood estimate of  $\hat{\beta}_X$  from Carroll et al was 1.76 (SE: 0.54) and the Bayesian estimate was 1.91 (SD: 0.56) (where exact prior parameters are not given) [1]. Muff et al used informative priors and found estimates just under that found by Carroll et al using MCMC and INLA (exact values not given). Use of the same informative priors as used in Muff et al leads to  $\hat{\beta}_X$  values of 1.90 and 1.87 for INLA and MCMC, respectively (SD: 0.56 and 0.58) [29].

Overall, the pattern of results obtained in this application is similar to those seen in the simulation study. This data set is most similar to the simulated scenario depicted in Figure 3.8 as blue along the topmost line, for which all methods produced reliable and relatively unbiased corrected  $\hat{\beta}_X$ . All four methods perform similarly in measurement error correction when measurement error variance is low. Point estimates are substantially further from the null than the naïve estimate, and the variance of the corrected estimates are all larger than the naïve variance. However, use of either INLA or MCMC results in efficiency gains via improved precision of the estimates. MI, as expected, has better precision than RC but not quite as good as that of INLA and MCMC.

### 3.7 SUMMARY

In this chapter, I have compared methods for correcting for the effects of measurement error in a continuous exposure. The methods considered were based on Bayesian principles (e.g. MCMC,

MI, and INLA) and they were compared with the popular RC method. Although the focus is on error in a single main exposure, well known to result in an attenuation of the estimated association with the outcome, the methods described apply equally if the error-prone variable is a confounder, though the impact on the main exposure-outcome association in that case is not necessarily an attenuation.

The simulation study demonstrated that all correction methods outperformed the naïve method in realistic scenarios. RC is straightforward to implement and performs well in many circumstances. However, the method requires an approximation in non-linear outcome models, such as logistic regression, which was seen to result in biased estimates for larger ORs, consistent with scenarios tested in other publications [20]. RC was also shown to be much more variable than the other methods in this chapter.

MCMC works very well for correction of classical measurement error in the predictor, particularly for linear regression across all settings and scenarios. It was shown that when little information is contained in the likelihood, the use of very non-informative priors results in MCMC chains that are very slow to converge. This was particularly egregious for logistic regression which is constrained by both the lack of information and the slower acceptance rate of the sampler. However, this approach gave approximately correct coverage even when the chains had not converged and some bias was present.

Given that MI can be performed in widely available software without much additional work on the part of the user, MI is a strong choice for measurement error correction when a validation study is present. This is especially true where measurement error correction might be performed alongside imputation of missing data. MI was shown in this chapter to produce reliable corrected estimates with standard uninformative priors as long as a validation study was available. However, use of uninformative priors in the replicate setting led to subsequent poor performance when the likelihood was empirically unidentifiable. The method could, in theory, be modified to use proper and weakly informative priors for all variables, including the substantive model regression coefficients; in this case, there is little to be gained over running an MCMC sampler for the model.

INLA was shown to produce unbiased results with good coverage in most scenarios using only a few minutes of computational time.

It is of note that the MCMC and INLA models were slightly misspecified as the exposure values and error-prone measures were not centered yet the prior for the exposure model had an expected value of zero. Were appropriate centering included, the MCMC chain would converge more rapidly and potentially both methods would have slightly less bias in the estimates of  $\beta_X$ .

Use of MCMC and hybrid Bayesian methods will be explored in the next chapter (Chapter 4) for the correction of measurement error when the functional form of the exposure includes a squared term.

## 4 MEASUREMENT ERROR CORRECTION FOR A QUADRATIC TRANSFORMATION OF THE ERROR-PRONE EXPOSURE

---

### 4.1 AIMS & OVERVIEW

Adding a squared transformation of a predictor in a regression model is the simplest extension to entering the predictor into the model using a linear term. It is arguably the most common non-linear transformation found in the literature.

In this chapter, the effect of measurement error on the exposure-outcome association is explored for the case when the error-prone variable takes a quadratic transformation in the substantive model. Potential methods for measurement error correction are evaluated via simulation studies considering several different quadratic shapes. RC extends easily to the inclusion of a squared term [1,49]. Bayesian MCMC is easily extended to any specified transformation [1]. The limitations of MI were explored in Chapter 3 after ensuring compatibility by use of rejection sampling [46]. For INLA, I introduce in this chapter the particular challenges for extension to a non-linear transformation given the method assumption of Gaussian relationships to the latent exposure [41,43].

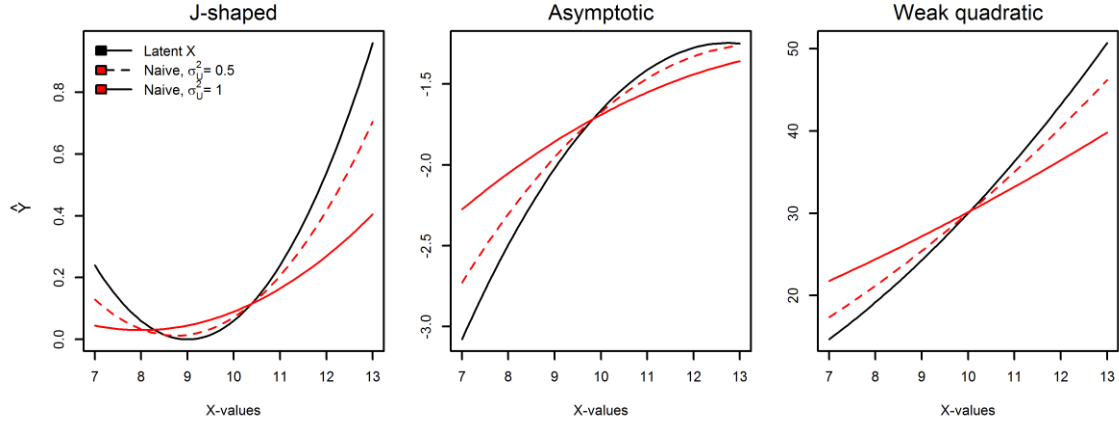
Methods are demonstrated for both continuous and binary outcomes with either a validation study or a replicate study.

### 4.2 QUADRATIC SHAPES

The form of the substantive model in this chapter is a generalized linear model with  $X$  and  $X^2$  in the linear predictor:

$$4.1 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_{X_1}X_i + \beta_{X_2}X_i^2 + \boldsymbol{\beta}_Z^T\mathbf{Z}_i, \quad i = 1, 2, \dots, n.$$

This model, a quadratic model, encompasses a range of different shapes of associations between  $X$  and the outcome, corresponding to different values for the parameters  $\beta_{X_1}$  and  $\beta_{X_2}$ . A  $\beta_{X_2}$  value of 0 corresponds to a purely linear association. In this chapter, three non-linear shapes which may be fit using a quadratic model are explored with an identity link for  $g(\cdot)$ , i.e. linear regression; two shapes are explored with a logit link, i.e. logistic regression. For linear regression and thus a continuous outcome, the substantive model includes a normally distributed residual,  $\varepsilon_{Y|XZ}$ , with mean zero and variance  $\sigma_{Y|XZ}^2$ . For the remainder of this chapter, all models have a continuous outcome modelled using linear regression unless otherwise stated.



**Figure 4.1** Plots of exposure-outcome associations fit to the quadratic model using 1) the latent  $X$  where  $X \sim N(10, 1)$ ; 2) a naïve fit of the error-prone measure of  $X$  ( $W$ ) with a measurement error variance,  $\sigma_U^2$ , of 0.25; and 3) a naïve fit of  $W$  with  $\sigma_U^2 = 1$ . Mean predicted  $Y$ -values ( $\hat{Y}$ ), from exposure values 7 to 13 in steps of 0.1, are the point-wise mean of curve fits from 200 simulated data sets of sample size 2000.

Several shapes of the exposure-outcome association were chosen to illustrate the impact of classical measurement error, and later to evaluate the correction methods when assuming a quadratic model fit and a continuous outcome. Keogh et al explored the effect of classical measurement error on a number of possible exposure-outcome curve shapes in the context of the fractional polynomial method and p-splines [23]. I use two of the four quadratic models explored in that paper: a “J-shaped” model and an “asymptotic” model. Additionally, in this work a “weak quadratic” model, similar to what might be found for the association between blood pressure and mortality, is considered. The shapes of association are illustrated in Figure 4.1 and the models used to generate these shapes are given in Table 4.1. Note that the asymptotic model is not a strictly quadratic model; however, the quadratic model fits this shape relatively well. In Chapter 6, alternative fits will be explored for this association.  $R^2$  is the proportion of the variance in the outcome predictable from the exposure. Residual variance ( $\sigma_{Y|X}^2$ ) values were chosen to obtain for each model an  $R^2$  of approximately 0.5.

### 4.3 NAÏVE ANALYSIS AND THE IMPACT OF MEASUREMENT ERROR

The naïve analysis uses the first error-prone measure  $W_1$  instead of the true exposure  $X$ , i.e. the latent  $X$ , in the linear predictor of a quadratic model:

$$4.2 \quad E[Y_i|W_i, \mathbf{Z}_i] = \beta_0^* + \beta_{W_1} W_{i1} + \beta_{W_2} W_{i1}^2 + \beta_{\mathbf{Z}}^T \mathbf{Z}_i,$$

assuming the classical error model from Equation 1.7,  $W = X + U$ , where  $U$  is a normally distributed measurement error with a mean of zero and variance  $\sigma_U^2$ . Because the outcome,  $Y$ , remains fixed while the observed mis-measured exposure values,  $W_1$ , become more scattered along the x-axis than the latent  $X$ , the effect of the measurement error is to flatten the observed curve and increase the apparent linearity of the relationship. The impact of classical measurement

*Table 4.1 Models for exposure-outcome association shapes with a continuous outcome*

Shape	Linear predictor	Residual variance ( $\sigma_{Y X}^2$ )
J-shaped	$0.06 * (X - 9)^2$	0.14
Asymptotic	$10 * (4 - X)^{-1}$	0.3
Weak quadratic	$0.3 * X^2$	6

error is illustrated for the shapes of association in Figure 4.1 using simulated data. The naïve curve results in overestimates of the outcome for some values of the exposure and underestimates for others [6,23]. The estimates of  $\beta_{W_1}$  and  $\beta_{W_2}$  are biased estimates of the true parameters of interest  $\beta_{X_1}$  and  $\beta_{X_2}$ .

In the case of a quadratic substantive model (Equation 4.1), there is no concise measure of the impact of measurement error such as the attenuation factor for a linear model. The dilution effect of measurement error on the  $X^2$  coefficient can be shown to be the square of the attenuation factor,  $\lambda$ , in Equation 1.9 [6]:

$$4.3 \quad \beta_{W_2} = \beta_{X_2} \lambda^2 = \beta_{X_2} \left( \frac{\sigma_{X|Z}^2}{\sigma_{X|Z}^2 + \sigma_U^2} \right)^2.$$

The effect on the linear coefficient in the quadratic model is dependent on both the  $X$  and  $X^2$  coefficients as well as the attenuation factor and the mean of  $X$  [6]:

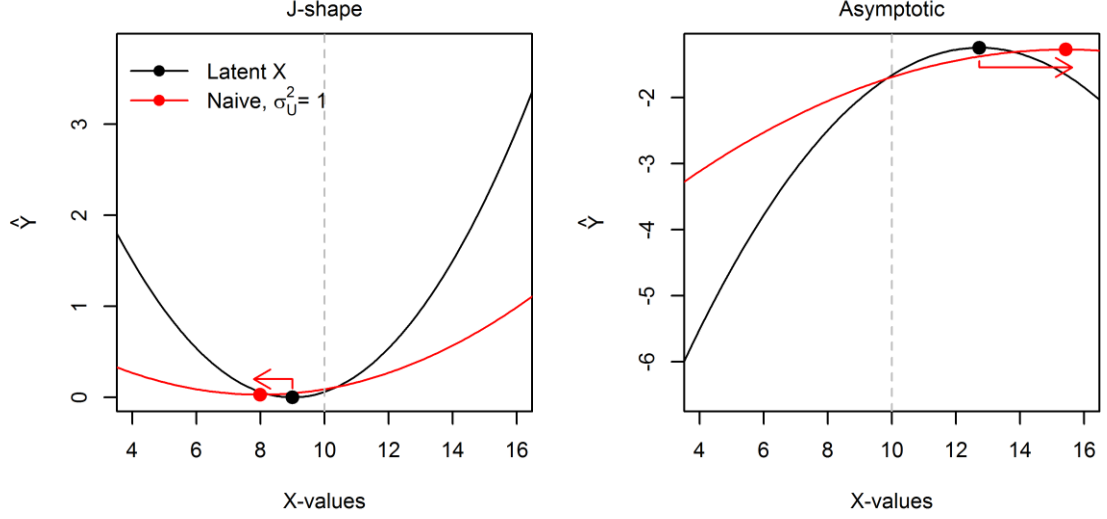
$$4.4 \quad \beta_{W_1} = \beta_{X_1} \lambda + 2\beta_{X_2} \lambda \mu_X.$$

Measurement error in  $X$  can affect the shape of the association in such a way that the turning point of the quadratic curve is biased. The value of  $X$  at the true turning point under Equation 4.1,  $\eta_X$ , can be found at  $-\beta_{X_1}/2\beta_{X_2}$ . As illustrated in Figure 4.2, the naïve estimate of the turning point under Equation 4.2 using  $W_1$ ,  $\eta_W$ , will be found biased away from  $\eta_X$  in the direction of the mean of  $X$ ,  $\mu_X$  [6]. The relationship between  $\eta_X$  and  $\eta_W$  is:

$$4.5 \quad \eta_W = -\frac{\beta_{W_1}}{2\beta_{W_2}} = \eta_X + \frac{\sigma_U^2}{\sigma_{X|Z}^2} (\eta_X - \mu_X).$$

Furthermore, the  $Y$  value of the naïve estimate of the turning point will be biased upwards if the turning point is a minima (Figure 4.2, left) or biased downwards if the turning point is a maxima (Figure 4.2, right) [6]. The turning point may be of particular importance in terms of setting public health recommendations as it will be the point of highest or lowest risk from the exposure.

The power to detect the non-linearity and select the best model is also reduced in the naïve analysis; this will be discussed further in Chapter 5.



**Figure 4.2** *Demonstration of the effect on turning point bias after measurement error is added to the exposure,  $X$ . The J-shaped and asymptotic associations are fit to the quadratic model using 1) the latent  $X$  where  $X \sim N(10, 1)$ ; 2) a naive fit of the error-prone measure of  $X$  with a measurement error variance,  $\sigma_U^2$ , of 1. Mean predicted  $Y$ -values ( $\hat{Y}$ ), from exposure values 7 to 13 in steps of 0.1, are the point-wise mean of curve fits from 200 simulated data sets of sample size 2000. The vertical, dashed gray line indicates the mean of  $X$ ,  $\mu_X$ .*

## 4.4 CORRECTION METHODS FOR A QUADRATIC MODEL

In this section, I will describe the modifications required to adapt the correction methods presented in Chapter 2, where a linear form of the error-prone variable appeared in the substantive model, to the quadratic substantive model presented in Equation 4.1. As in previous chapters, it is assumed that either a validation study has been performed on some fraction of the study population or that a replicate study has been performed providing at least two measures of the error-prone measure on all or part of the study population. In all cases, a classical error model is assumed unless specifically stated otherwise.

### 4.4.1 Regression calibration (RC)

RC operates by replacing  $X$  in the substantive model with the expectation of  $X_i$  given the error-prone measure(s),  $\mathbf{W}_i$ , and any adjustment variables,  $\mathbf{Z}_i$ . When the substantive model includes transformations of  $X$ , e.g.  $X^p$ , RC is extended by replacing  $X^p$  with  $E[X^p | \mathbf{W}, \mathbf{Z}]$ . For a quadratic model with a continuous outcome, RC therefore works by regressing  $Y_i$  on  $E[X_i | \mathbf{W}_i, \mathbf{Z}_i]$  and  $E[X_i^2 | \mathbf{W}_i, \mathbf{Z}_i]$ :

$$4.6 \quad E[Y_i | \mathbf{W}_i, \mathbf{Z}_i] = \beta_0 + \beta_{X_1} E[X_i | \mathbf{W}_i, \mathbf{Z}_i] + \beta_{X_2} E[X_i^2 | \mathbf{W}_i, \mathbf{Z}_i] + \boldsymbol{\beta}_Z^T \mathbf{Z}_i.$$

The expectation  $E[X_i | \mathbf{W}_i, \mathbf{Z}_i]$  may be estimated as described in Chapter 2 by regressing  $W_1$  on  $X$  if a validation study is present or by regressing  $W_2$  on  $W_1$  if a replicate study is present (Equations 2.2 and 2.3). If a validation study is present, the variance of  $X$  given the error-prone measure and any accurately measured covariates,  $\sigma_{X|WZ}^2$ , may be directly estimated. Alternatively, in a replicate study,  $\sigma_{X|WZ}^2$  may be estimated as the covariance of the error-prone measures conditional on  $\mathbf{Z}_i$ .



The definition of variance ( $\text{var}(X) = E[X^2] - E[X]^2$ ) can be used to rearrange  $E[X_i^2|\mathbf{W}_i, \mathbf{Z}_i]$  [1,49]. Therefore, we replace  $E[X_i^2|\mathbf{W}_i, \mathbf{Z}_i]$  in Equation 4.6 with  $\sigma_{X|WZ}^2 + E[X_i|\mathbf{W}_i, \mathbf{Z}_i]^2$ :

$$4.7 \quad E[Y_i|\mathbf{W}_i, \mathbf{Z}_i] = \beta_0 + \beta_{X_1}E[X_i|\mathbf{W}_i, \mathbf{Z}_i] + \beta_{X_2}E[X_i|\mathbf{W}_i, \mathbf{Z}_i]^2 + \beta_{X_2}\sigma_{X|WZ}^2 + \boldsymbol{\beta}_Z^T\mathbf{Z}_i.$$

Equation 4.7 can be rearranged to put all constant terms together:

$$4.8 \quad E[Y_i|\mathbf{W}_i, \mathbf{Z}_i] = (\beta_0 + \beta_{X_2}\sigma_{X|WZ}^2) + \beta_{X_1}E[X_i|\mathbf{W}_i, \mathbf{Z}_i] + \beta_{X_2}E[X_i|\mathbf{W}_i, \mathbf{Z}_i]^2 + \boldsymbol{\beta}_Z^T\mathbf{Z}_i.$$

It follows from Equation 4.8 that if  $E[X_i|\mathbf{W}_i, \mathbf{Z}_i]$  is used in place of  $X$  and  $E[X_i|\mathbf{W}_i, \mathbf{Z}_i]^2$  in place of  $X^2$  in the quadratic substantive model, the desired  $\hat{\beta}_{X_1}$  and  $\hat{\beta}_{X_2}$  are equivalent to the observed linear and quadratic parameters and the desired  $\hat{\beta}_0$  is equivalent to  $\hat{\beta}_0^* - \hat{\beta}_{X_2}\hat{\sigma}_{X|WZ}^2$ .

This may be extended to logistic regression with the same caveats regarding approximation as for the untransformed model (Section 2.2). The Cox proportional hazards model is a special case of this method as the term  $\beta_{X_2}\text{var}(X_i|\mathbf{W}_i, \mathbf{Z}_i)$  is subsumed by the baseline hazard (15).

Bootstrapping [1] or an extension of the delta method can be used to obtain SEs [6].

#### 4.4.2 Bayesian analysis using MCMC

A three-part conditional independence structure was introduced in Section 2.3 and applied in the context of an untransformed predictor in Chapter 3. This approach may be extended easily for use when the substantive model also includes an  $X^2$  term, i.e. the quadratic model in Equation 4.1.

In this chapter, the substantive model specified,  $f(Y_i|X_i, \mathbf{Z}_i; \boldsymbol{\beta})$ , is updated to be the quadratic model given in Equation 4.1. The measurement error model,  $f(\mathbf{W}_i|X_i; \boldsymbol{\pi})$ , remains the classical error model (Equation 1.7) and the exposure model,  $f(X_i|\mathbf{Z}_i; \boldsymbol{\alpha})$ , remains the distribution of  $X$  dependent on any accurately measured covariates,  $\mathbf{Z}$ , here assumed to be normal.

Scaling of the variables may be necessary to specify plausible prior distributions [28,85] and centering the exposure and its squared term will reduce correlation between the terms and improve MCMC convergence [28]. Scaling and centering must be performed after the transformation of the squared term. The mean and standard deviation of the latent  $X$  and  $X^2$  can be estimated from  $\mathbf{W}$  when no validation study is available.

#### 4.4.3 INLA

It was discussed in Section 2.5 that the joint distribution of the latent Gaussian parameters, including the latent  $X$ , is assumed to have the attributes of a GMRF. Whether the simplified Laplace approximation used for the latent Gaussian parameters is reliable depends on the accuracy of this assumption. In Chapter 2, the latent Gaussian parameters,  $\mathbf{v}$ , included the regression parameters from the substantive model, the exposure model, and the latent  $X$ , i.e.

$\beta_0, \boldsymbol{\beta}_Z, \alpha_0, \boldsymbol{\alpha}_Z,$  and  $X_i$ . When the substantive model is the quadratic model,  $\boldsymbol{v}$  includes  $\beta_0, \boldsymbol{\beta}_Z, \alpha_0, \boldsymbol{\alpha}_Z, X_i,$  and  $X_i^2$ . The method cannot treat both  $X$  and  $X^2$  as approximately normally distributed latent variables as the square of a normally distributed variable cannot also be normally distributed. Therefore, a significant extension to the INLA method would be required to accommodate transformations of a latent variable.

Furthermore, the software for applying INLA cannot accommodate a transformation of a latent Gaussian variable within the substantive model; therefore, the impact of this violation of principle cannot be easily assessed.

Given these limitations, in this work I will no longer pursue INLA as previously described as a method of measurement error correction when the error-prone measure has been transformed within the substantive model. However, in the next section, a hybrid method using MCMC or INLA and attributes of RC is described.

#### 4.4.4 Bayesian regression calibration

In this chapter and the previous, correction methods are applied in the context of relatively simple models (i.e. the classical error model and the use of a specified functional form of the error-prone exposure). In this relatively straightforward setting, I would like to propose a novel correction method which is a hybrid between Bayesian methods and RC. This method is expected to easily adapt to settings with a complex error model and an unknown functional form of the error-prone predictor, i.e. model selection. MCMC solutions are powerful and flexible, but slow to converge, particularly when model selection is incorporated (Chapters 5 and 6). While the time involved to run standard RC is negligible, estimation of the maximum likelihood estimates of  $E[X_i | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  for more complex error models can be cumbersome or even prohibitive where the likelihood cannot be expressed in closed form [110]. When assuming the classical error model, this is not a limiting problem with linear and quadratic substantive models but becomes so for more complex substantive models such as the full set of models required for the fractional polynomial method, the topic of Chapter 6.

Standard RC relies on the estimation of  $E[X_i | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  and  $E[X_i^2 | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  via maximum likelihood estimation. An alternative means of obtaining  $E[X_i | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  and  $E[X_i^2 | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  is via the posterior mean of  $f(X_i | \boldsymbol{W}_i, \boldsymbol{Z}_i; \boldsymbol{\alpha}, \boldsymbol{\pi})$ , where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\pi}$  represent the parameters of the exposure model and the error model, respectively (Section 2.3). Posterior samples of the latent  $X$ , denoted  $\tilde{X}_i$ , are drawn using MCMC after the chains have reached convergence. By squaring all samples of  $\tilde{X}_i$ , one can estimate  $E[\tilde{X}_i^2 | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  directly.  $E[\tilde{X}_i^2 | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  will be a good estimate of  $E[X_i^2 | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  as long as the error and exposure models are not misspecified and enough samples have been drawn to approximate the distribution. Each estimated expectation,  $E[\tilde{X}_i | \boldsymbol{W}_i, \boldsymbol{Z}_i]$  and  $E[\tilde{X}_i^2 | \boldsymbol{W}_i, \boldsymbol{Z}_i]$ , can

then be inserted directly into Equation 4.6 to fit the quadratic model. Inference can then be made according to frequentist principles.

The MCMC chains would be expected to converge more quickly using this simpler model than the fully Bayesian model which incorporates the substantive model. This is particularly true for non-linear outcome models such as logistic regression.

Either MCMC or INLA may be used to generate the posterior samples of  $\tilde{X}_i$ . While sampling is not inherent to the INLA method, samples may be obtained from the estimated posterior distribution. This operation is still much faster than MCMC because there is no need to wait for convergence or concerns about autocorrelation, i.e. the ESS is equivalent to the number of samples drawn. In this work, each method used in this way will be referred to as MCMC-RC or INLA-RC, respectively. R code demonstrating the implementation of each is provided in Appendix B. Any other Bayesian method of analysis, such as Hamiltonian Monte Carlo algorithms, may be used similarly.

Bayesian RC would be expected to underestimate variance in the regression parameters of the substantive model as it does not propagate the uncertainty due to measurement error from the MCMC model to the substantive model (Section 2.2.1). In theory, bootstrapping may be used for better estimates of the SEs; however, this could only be done for simple models and small data sets for MCMC-RC. For INLA-RC, bootstrapping is more feasible but was not used in simulation studies in this thesis.

#### **4.4.5 Multiple imputation**

##### ***Multiple imputation of squared terms in the missing data context***

The desirability of compatibility of the substantive model and the imputation model when performing MI was discussed in Section 2.6. However, several authors have considered MI methods for imputing covariates when they appear as transformed terms in the substantive model which use imputation models that are not compatible with the substantive model [111–113].

The simplest method is to impute the missing variable  $X$  assuming a linear relationship to  $Y$ , then transform it to  $X^2$  for the substantive model. This preserves the  $X$  and  $X^2$  relationship but violates the theoretical properties of the joint model for the substantive model and imputation model underlying the multiple imputation. This method, sometimes referred to as the “passive approach”, has been shown to result in biased regression estimates [111,112]. Alternatively, in what’s been called the “Just Another Variable” (JAV) approach, one can impute both  $X$  and  $X^2$  separately using chained equations as if they were different variables [111]. Unlike the “passive approach”, JAV does not preserve the relationship between  $X$  and  $X^2$ . In some settings, JAV may

improve estimates over the “passive approach”, but in many common settings it still results in bias [111–113].

In a method called polynomial combination, Vink and van Buuren proposed to impute not  $X$ , but  $X + X^2$  [112]. While this method results in less bias than JAV and preserves the  $X$  and  $X^2$  relationship, it is not easily extendable to other transformations of the latent exposure.

SMC-FCS, which uses rejection sampling to ensure compatibility (Section 2.6), was demonstrated in its original publication for use with a quadratic substantive model (Equation 4.1) [46]. For this transformation of the latent exposure as well as others, SMC-FCS was demonstrated to be effective at minimizing bias due to missing data.

### *Multiple imputation of squared terms in the measurement error context*

None of the above MI approaches have, to my knowledge, been applied to exposure measurement error in the published literature.

In Chapter 3, I explored the use of SMC-FCS for measurement error correction with either a validation study or a replicate study present. When a validation study has been performed, SMC-FCS is an effective tool for minimizing bias without any alteration to the method as used for missing data. However, for replicate studies, it was necessary to alter the method to incorporate the measurement error model and to stipulate proper priors for both the  $\sigma_X^2$  and  $\sigma_U^2$  variances to ensure reliable posterior inference [27]. From the simulation studies performed (Section 3.5.3), it was further shown that without stipulation of somewhat informed priors for the substantive model regression coefficients, estimated posterior distributions of the regression parameters are uninformative when the likelihood contains little data, i.e. due to high measurement error variance and/or small sample size. Therefore, use of a fully Bayesian model from which to draw samples of the latent  $X$  may be required for reliability across scenarios.

The same MCMC model as used for the fully Bayesian analysis may be used to obtain samples of  $\tilde{X}_i$  and  $\tilde{X}_i^2$  from  $f(X_i|\mathbf{W}_i, Y_i, \mathbf{Z}_i; \boldsymbol{\theta})$  which could then be imputed into data sets to be used in the standard MI fashion. That is, the quadratic model would be fit to each imputed data set and pooled regression estimates obtained by applying Rubin’s Rules. In this way, estimates with frequentist properties may be obtained in lieu of Bayesian posterior means.

Given the limitations of MI as demonstrated in Chapter 3 and as explored by others in the field of missing data, in this thesis I will no longer pursue MI as a method of measurement error correction where the error-prone measure has been transformed.

## 4.5 SIMULATION STUDY WITH A CONTINUOUS OUTCOME

### 4.5.1 Simulation study design

In order to evaluate the performance of the measurement error correction methods presented in this chapter, I performed a simulation study with the data generated as described in Box 4.1. See an example of a single simulated data set for each of the three shapes of association in Figure 4.3. The error prone measure,  $W$ , was generated according to the classical measurement error model (Equation 1.7), i.e.  $W = X + U$ .

Three study settings were assessed: 1) a validation study where 30% of the study population has the true  $X$  observed and all participants have a single error-prone measure,  $W_1$ , and the measurement error variance is a quarter the variance of  $X$ , i.e.  $\sigma_U^2 = 0.25$ ; 2) a similar validation study

where the measurement error variance is equal to the variance of  $X$ , i.e.  $\sigma_U^2 = 1$ ; and 3) a replicate study where all participants have two error-prone measures,  $\mathbf{W}_i = \{W_1, W_2\}$ , and  $\sigma_U^2 = 1$ . Each setting was generated in combination with each shape of association for a total of nine scenarios.

While accurately observed covariates,  $\mathbf{Z}$ , are included in the explanations of each method, all simulations focus on a single error-prone exposure with no covariates for ease of inference.

### 4.5.2 Criteria for evaluating quadratic fit

Performance of the measurement error correction methods was evaluated assuming a quadratic model fit (Equation 4.1). Below is a summary of the performance measures used to assess the methods.

The bias in the turning point of the mean corrected curve was assessed for the J-shaped association and the asymptotic association; the weak association did not have a turning point within the observed range of  $X$ . Bias was estimated as the mean difference between the estimated turning point, i.e.  $-\hat{\beta}_{X_1}/2\hat{\beta}_{X_2}$ , and the true turning point, determined as outlined in Section 4.3 and illustrated in Figure 4.2.

#### Box 4.1 Simulation study design

Simulations per scenario: 200

Sample size:  $N = 2000$

True exposure:  $X_i \sim N(10, 1)$

Classical error model:

$$W_{ij} = X_i + U_{ij}, \quad U_{ij} \sim N(0, \sigma_U^2), \quad j = 1, 2$$

Measurement error variance:  $\sigma_U^2 = \{0.25, 1\}$

Outcome models (Table 4.1):

J-shaped association:

$$Y_i \sim N(0.06(X_i - 9)^2, 0.14)$$

Asymptotic association:

$$Y_i \sim N(10(4 - X_i)^{-1}, 0.3)$$

Weak quadratic association:

$$Y_i \sim N(0.3X^2, 6)$$

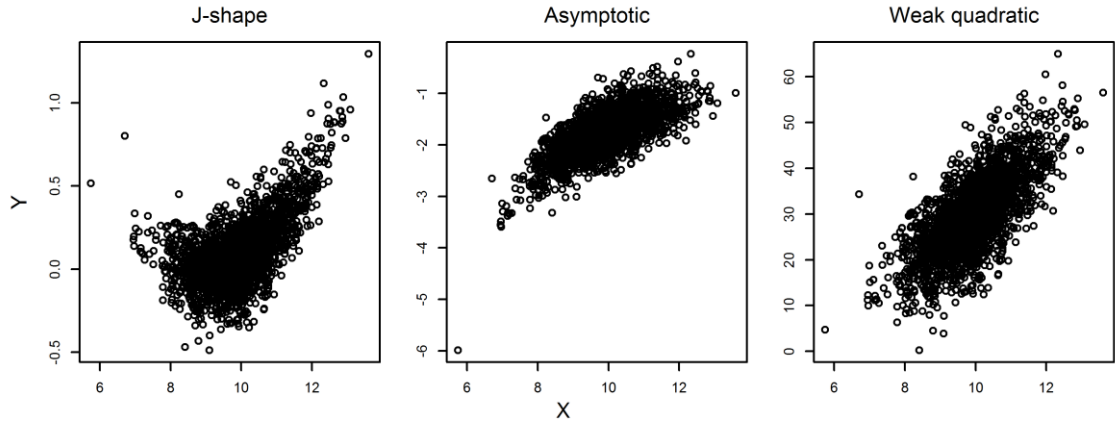


Figure 4.3 Example of simulated data for each quadratic association shape.

Bias in the estimates of the substantive model regression coefficients, i.e.  $\beta_0$ ,  $\beta_{X_1}$ , and  $\beta_{X_2}$ , was assessed. Bias was determined as the mean difference between the estimated regression coefficients after application of each method and the true regression coefficients in the quadratic model for the J-shaped association and the weak association. Because the asymptotic association was not generated from a quadratic polynomial model, the “true” regression coefficients were estimated from the mean of 10,000 simulations of the quadratic model with  $X$ . 95% MCE CIs were obtained around the estimated bias (Equation 3.1).

In order to assess the error between the estimated shape of a curve and the true underlying shape, the mean integrated square error (MISE) of the predicted outcome values,  $\hat{f}(x)$ , can be estimated over a series of support points. The MISE is the expectation of the mean square error across the curve [114]. Using a series of  $P$  evenly spaced values where  $X = x$ , the following was estimated for each simulation:

$$4.9 \quad MISE = \frac{1}{P} \sum_{x=7}^{x=13} (\hat{f}(x)_x - f(x)_x)^2.$$

In the simulation study, the MISE was estimated from  $X = 7$  to 13 (mean of  $X \pm 3$  standard deviations) in increments of 0.1. For each correction method, for each setting and association shape, the average MISE across simulations was used to assess the error between the corrected curve and the true curve. The interpretation of the MISE is akin to that for the RMSE capturing error in the estimator due to either bias or variability.

Coverage of the nominal 95% CI/CrIs for  $\beta_{X_1}$  and  $\beta_{X_2}$  was assessed. MCE 95% CIs were obtained around the coverage estimates (Equation 3.2).

In addition to the above criteria, for each scenario and correction method, a plot was generated depicting: 1) the estimated curve for every simulation from  $X = 7$  to 13; 2) the point-wise mean of all curve fits estimated by finding the mean  $\hat{Y}$  at each exposure point from  $X = 7$  to  $X = 13$  in

steps of 0.1; and 3) the point-wise mean estimated 95% confidence or credible bounds estimated by the correction method.

### 4.5.3 Method implementation

Example code for simulated data sets and implementation of each method can be found in Appendix B.

For each simulated data set, the quadratic model was fit with the latent  $X$  (Equation 4.1). Additionally, the naïve analysis was performed by fitting the quadratic model using the “observed”  $W_1$ .

#### *Regression calibration*

For standard RC, 1,000 bootstraps were used to estimate variance for the regression parameter estimates and basic bootstrap 95% confidence intervals were estimated [77].

#### *Bayesian MCMC*

For the quadratic substantive model, it is desirable that both  $X$  and  $X^2$  be centered at zero then scaled by their standard deviations [28,85]; however, this standardization cannot occur before the latent value is transformed [8] requiring the following process to be applied. In the case of a validation study, the mean and standard deviation of  $X$ ,  $\mu_X$  and  $\hat{\sigma}_X$ , and of  $X^2$ ,  $\mu_{X^2}$  and  $\hat{\sigma}_{X^2}$ , can be found directly from the observed  $X$  values. For a replicate study, these values can be estimated from the mean and covariance of  $\mathbf{W}$ . All observed  $X$  values and any error-prone measures  $\mathbf{W}$  were scaled then centered prior to running the MCMC. At each iteration of MCMC, after the latent  $X$  values were sampled ( $\tilde{X}$ ), in order to find  $X^2$ , the standardization of  $X$  was reversed, i.e.  $\tilde{X}\hat{\sigma}_X - \hat{\mu}_X$ , and the samples squared to find the latent  $\tilde{X}^2$  values. The  $\tilde{X}^2$  values were then centered near zero using  $\hat{\mu}_{X^2}$  and scaled by  $\hat{\sigma}_{X^2}$ . The scaled and centered samples of  $\tilde{X}$  and  $\tilde{X}^2$  were used in the outcome regression model. The outcome  $Y$  was also scaled by its standard deviation prior to running the MCMC. Regression coefficient estimates were back-transformed to the original scale at each iteration for ease of inference.

Within the MCMC model, each variance parameter,  $\sigma_X^2$ ,  $\sigma_U^2$ ,  $\sigma_{Y|X}^2$ , is assigned an inverse gamma prior,  $IG(1, 1)$ , which has 95% of its density between 0.27 and 39.5 and allows for only positive values. The prior for the expectation of  $X$ ,  $\mu_X$ , is  $N(0, 1000)$ . Finally, the prior for the substantive model regression coefficients is a multivariate normal distribution with a mean vector of zeros and a covariance matrix with a diagonal vector of 10,000. In the JAGS language, precision (i.e. 1/variance) is used in lieu of variance.

Three MCMC chains were run for each simulation. In the validation study setting, the burn-in for each chain was 40,000 iterations. Gelman and Rubin’s convergence diagnostic,  $\hat{R}$  (Section 2.4),

was calculated based on the last 10,000 burn-in samples to assess convergence. Another 20,000 samples were collected from each chain and used for inference. The ESS was calculated from these final samples to assess whether sufficient samples for inference had been collected (Equation 2.10). In the replicate study setting, a similar procedure was followed with a burn-in period of 80,000 iterations.

For each scenario (i.e. 200 simulations  $\times$  9 scenarios), I assessed convergence using  $\hat{R}$ . Convergence was believed to have occurred if  $\hat{R} < 1.1$ .

### ***MCMC-RC***

For the MCMC-RC method, where the MCMC model is comprised of only the measurement error model and the exposure model, the same priors were specified as in the fully Bayesian analysis for the relevant parameters. All observed  $X$  values and any error-prone measures  $\mathbf{W}$  were centered by the estimated mean of  $X$ ,  $\hat{\mu}_X$ , then scaled by the  $\hat{\sigma}_X$  prior to running the MCMC.

Three MCMC chains were run for each simulation with 40,000 iterations of burn-in.  $\hat{R}$  was calculated on the last 10,000 burn-in samples to assess convergence. Another 6,000 samples were collected from each chain and used for inference with the ESS estimated to assess that these were sufficient.

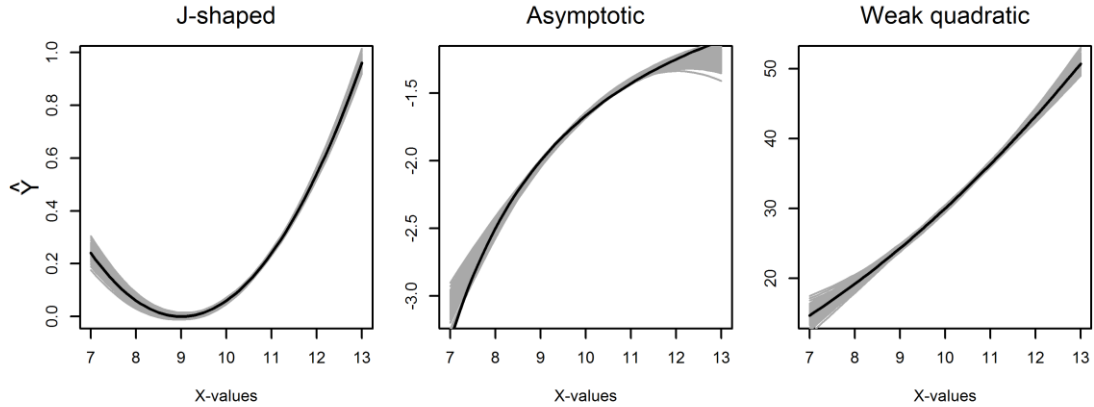
### ***INLA-RC***

INLA-RC used the same priors as MCMC-RC and similarly centered then scaled the observed measures and values of  $X$ . Starting values specified for the variances were all set at 1. After running INLA with the combined measurement error and exposure models, 3,000 posterior samples from each  $f(X_i | \mathbf{W}_i; \boldsymbol{\alpha}, \boldsymbol{\pi})$  were generated to perform INLA-RC.

#### **4.5.4 Fitting the latent $X$ and naïve models**

The results from fitting the quadratic model using the latent  $X$  can be found in Tables 4.2-4.4. Bias in either the turning point or the regression coefficients was negligible for all association shapes. The average MISE for the J-shaped, asymptotic, and weak quadratic associations was 0.00011, 0.040, and 0.19, respectively. The difference in MISE values between the shapes of association when no measurement error was present reflects their differing residual variances (Box 4.1). The relative close fit of the J-shaped association across simulations can be seen in the left-most plot of Figure 4.4. Conversely, the wider variation in the weak quadratic fit, indicated by the large  $Y$  scale, can be seen in the right-most plot of Figure 4.4. Finally, in the center plot of Figure 4.4, one can observe the deviation from the true curve resulting from fitting the quadratic model to data generated from the asymptotic shape which is not, in fact, quadratic. There is also undercoverage observed in the 95% CIs of the regression coefficients for the asymptotic shape





*Figure 4.4 Latent X fit in simulation: the quadratic model fit to the latent X for each of 200 simulations (gray lines). The black line represents the true curve used to generate the simulated data.*

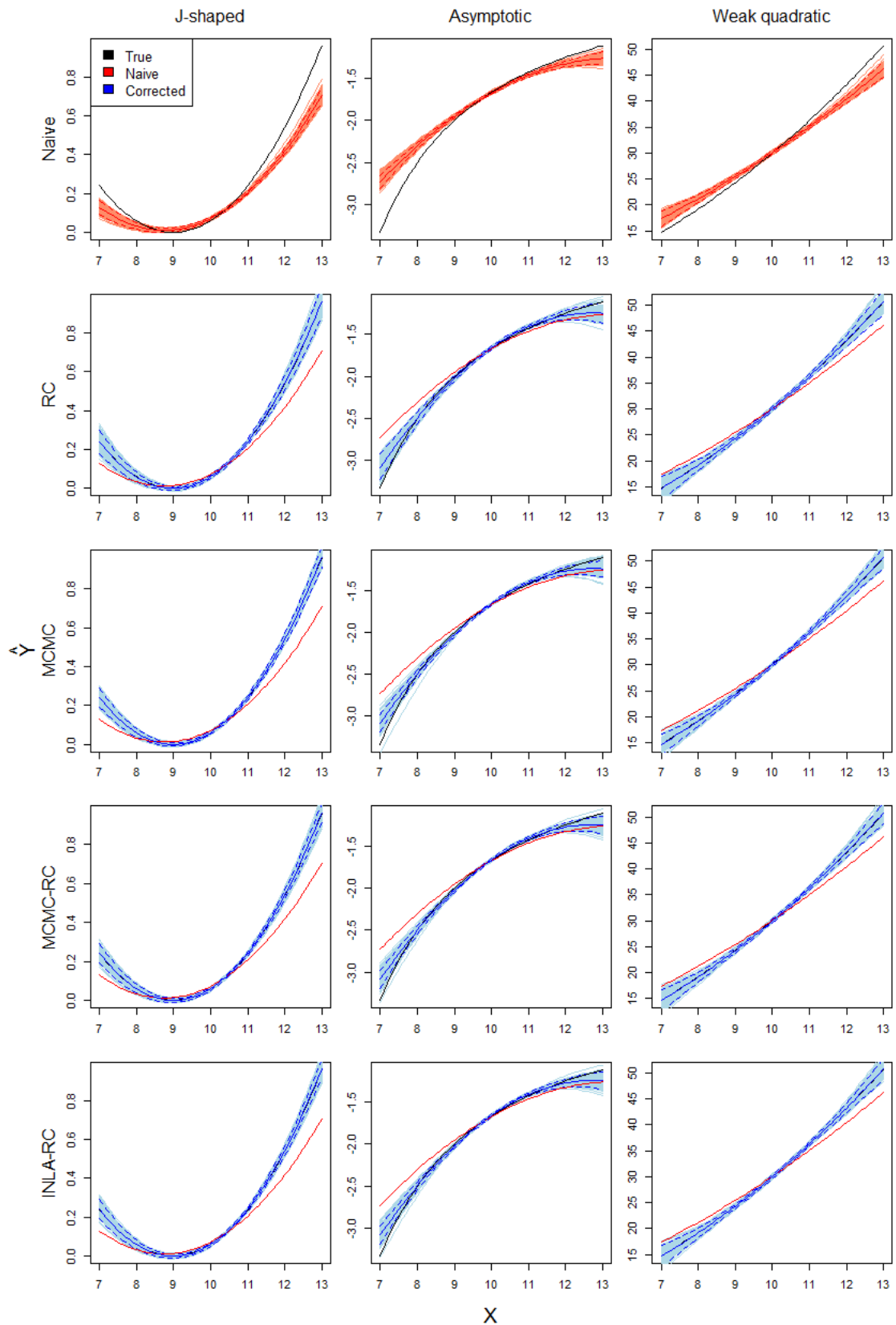
(Tables 4.2-4.4). Coverage is at the nominal level for the J-shaped and weak quadratic associations in analyses using the latent  $X$ .

The addition of measurement error resulted in bias for all three shapes. When  $\sigma_U^2 = 0.25$  (Table 4.2; Figure 4.5), the turning point for the J-shaped association falls by 0.3 (away from the mean of  $X$ ) and the turning point for the asymptotic association is observed to increase by 0.7 (away from the mean).  $\hat{\beta}_{X_2}$  is biased toward the null by 36% for each shape, as predicted by Equation 4.3.  $\hat{\beta}_{X_1}$  is biased toward the null by for the J-shaped and asymptotic associations by 38% and 33%, respectively. The slightly obscured shape of each curve can be observed in the top row of Figure 4.5. The naïve departure from the true association shape is also reflected in the average naïve curve MISE which was increased 78-fold, 7-fold, and 25-fold over the latent  $X$  average MISE for the J-shape, asymptotic, and weak quadratic association, respectively. The coverage of the 95% CIs for the J-shaped and asymptotic associations was at or very near 0%. For the weak quadratic association, significant undercoverage was observed only for  $\beta_{X_2}$  (73%).

When  $\sigma_U^2 = 1$  (Tables 4.3 and 4.4), bias in the turning point is -1.0 and 2.7 for the J-shaped association and asymptotic association, respectively.  $\hat{\beta}_{X_2}$  is biased toward the null by 75% for each shape and  $\hat{\beta}_{X_1}$  is biased toward the null for the J-shaped and asymptotic associations by 78% and 70%, respectively. In the top row of Figure 4.6, it can be seen that the higher measurement error resulted in a much more dramatic obfuscation of the shapes; each appears much more linear than the true shape. The average naïve MISE was increased 353-fold, 25-fold, and 150-fold over the latent  $X$  for the J-shape, asymptotic, and weak quadratic association, respectively. The coverage of the 95% CIs around  $\beta_{X_1}$  and  $\beta_{X_2}$  remain at 0% for the J-shaped and asymptotic associations and for the weak quadratic shape are 77% and 3%, respectively.

**Table 4.2 Simulations with a validation study and measurement error variance  $\frac{1}{4}$  the variance of  $X$ :** Mean bias in the quadratic model regression estimates, coverage of the 95% confidence intervals or credible intervals and the associated Monte Carlo error 95% confidence interval (MCE 95% CI), average mean integrated square error (MISE) over the exposure range 7-13, and bias in the turning point of the mean curve (where applicable) estimated from 200 simulations with a continuous outcome, a validation study performed on 30% of 2000 study participants after application of the designated method to data generated from three shapes of association (J-shape, asymptotic, and weak quadratic). \*indicates a MCE 95% CI for bias that excludes zero. † indicates that true values were estimated from 10,000 simulations with the latent  $X$  values.

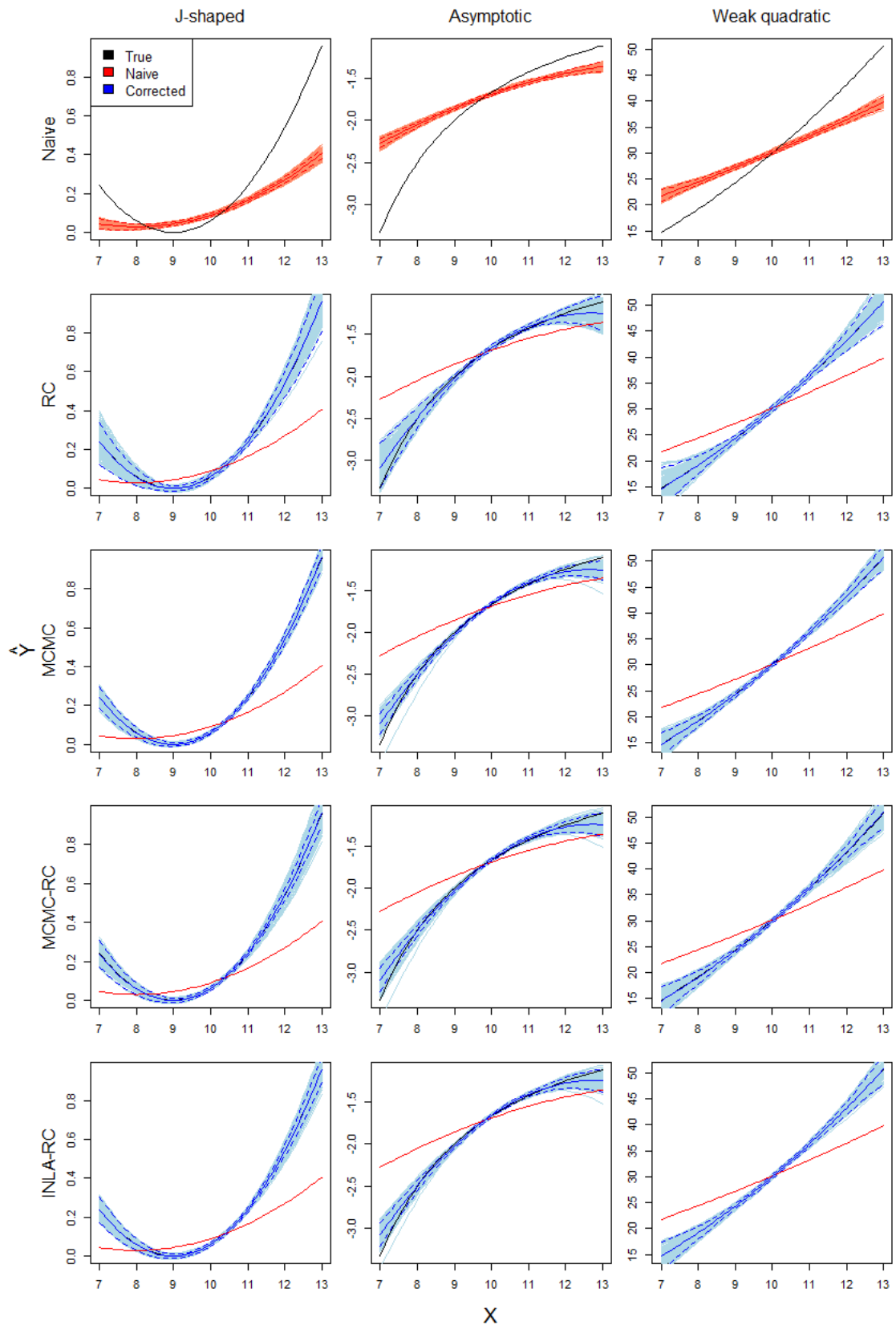
Method	Bias			Coverage (MCE 95% CI)		Average MISE	Turning point bias
	$\beta_0$	$\beta_{x_1}$	$\beta_{x_2}$	$\beta_{x_1}$	$\beta_{x_2}$		
<b>J-shaped</b>							
<b>True values</b>	4.86	-1.08	0.06				9.0
<b>Latent X</b>	-0.003	0.000	0.0000	96% (93 – 98)	96% (93 – 98)	0.00011	0
<b>Naïve</b>	-1.915*	0.409*	-0.0217*	0% (0 – 0)	0% (0 – 0)	0.00857	-0.3
<b>RC</b>	-0.002	0.000	0.0000	95% (91 – 98)	94% (91 – 97)	0.00031	0
<b>MCMC</b>	0.008	-0.002	0.0001	97% (94 – 99)	97% (94 – 99)	0.00016	0
<b>MCMC-RC</b>	0.026	-0.006	0.0003	92% (88 – 96)	92% (88 – 95)	0.00022	0
<b>INLA-RC</b>	0.014	-0.003	0.0002	92% (88 – 96)	92% (88 – 95)	0.00021	0
<b>Asymptotic</b>							
<b>True values†</b>	-10.326	1.428	-0.0562				12.7
<b>Latent X</b>	0.021	-0.005	0.0002	92% (88 – 95)	91% (87 – 95)	0.0050	0
<b>Naïve</b>	2.631*	-0.468*	0.0204*	1% (0 – 1)	1% (0 – 1)	0.0330	0.7
<b>RC</b>	0.012	-0.003	0.0002	95% (92 – 98)	95% (92 – 98)	0.0054	0
<b>MCMC</b>	0.028	-0.007	0.0004	91% (87 – 95)	92% (88 – 95)	0.0051	0
<b>MCMC-RC</b>	-0.007	0.000	0.0001	92% (88 – 95)	92% (88 – 95)	0.0051	0
<b>INLA-RC</b>	0.010	-0.003	0.0002	92% (88 – 96)	92% (88 – 96)	0.0052	0
<b>Weak quadratic</b>							
<b>True values</b>	0	0	0.3				
<b>Latent X</b>	-0.127	0.020	-0.0008	96% (93 – 98)	96% (93 – 98)	0.19	
<b>Naïve</b>	1.143*	0.976*	-0.1085*	94% (90 – 97)	73% (67 – 79)	4.74	
<b>RC</b>	-0.111	0.010	0.0002	96% (93 – 99)	96% (93 – 99)	0.33	
<b>MCMC</b>	-0.011	-0.021	0.0021	95% (91 – 98)	95% (91 – 98)	0.25	
<b>MCMC-RC</b>	0.142	-0.058	0.0042	96% (93 – 98)	96% (93 – 99)	0.28	
<b>INLA-RC</b>	0.154	-0.051	0.0035	96% (93 – 98)	95% (92 – 98)	0.28	



*Figure 4.5 Simulations with a validation study and measurement error variance  $\frac{1}{4}$  the variance of  $X$ : Fitted curves from 200 simulations with a continuous outcome for each correction method (by row) and each shape (by column) compared to the latent  $X$  (black line) and the naïve (red line) fitted curves when a validation study is performed on 30% of 2000 study participants. Darker solid red/blue line is the mean fitted curve, the point-wise mean of curve fits, and the dashed red/blue lines are the mean point-wise 95% confidence or credible bounds as determined by each method.*

**Table 4.3 Simulations with a validation study and measurement error variance equal to the variance of X:** Mean bias in the quadratic model regression estimates, coverage of the 95% confidence intervals or credible intervals and the associated Monte Carlo error 95% confidence interval (MCE 95% CI), average mean integrated square error (MISE) over the exposure range 7-13, and bias in the turning point of the mean curve (where applicable) estimated from 200 simulations with a continuous outcome, a validation study performed on 30% of 2000 study participants after application of the designated method to data generated from three shapes of association (J-shape, asymptotic, and weak quadratic). \*indicates a MCE 95% CI for bias that excludes zero. † indicates that true values were estimated from 10,000 simulations with the latent X values.

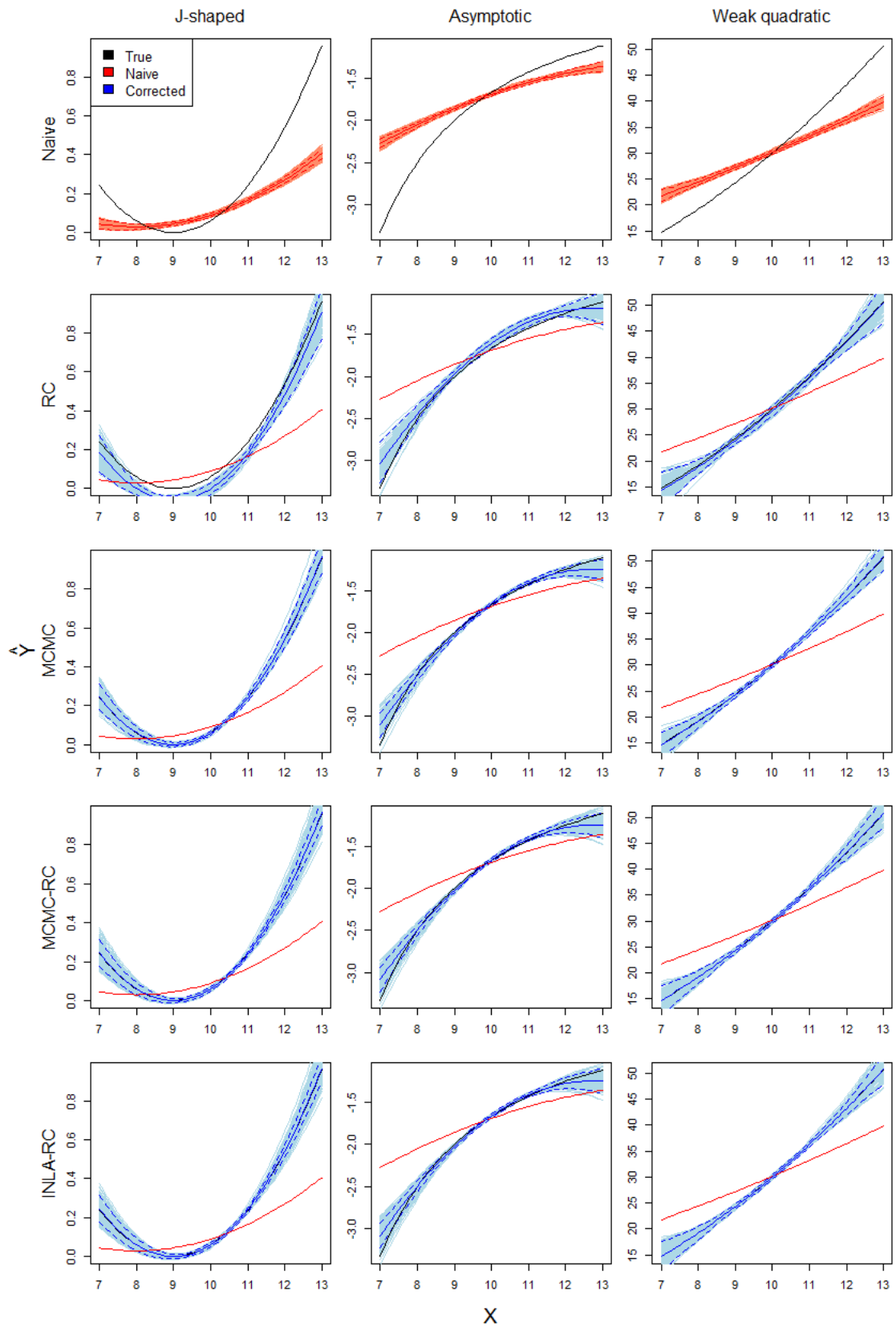
Method	Bias			Coverage (MCE 95% CI)		Average MISE	Turning point bias
	$\beta_0$	$\beta_{X_1}$	$\beta_{X_2}$	$\beta_{X_1}$	$\beta_{X_2}$		
<b>J-shaped</b>							
<b>True values</b>	4.86	-1.08	0.06				9.0
<b>Latent X</b>	-0.003	0.000	0.0000	96% (93 – 98)	96% (93 – 98)	0.00011	0
<b>Naïve</b>	-3.872*	0.840*	-0.0450*	0% (0 – 0)	0% (0 – 0)	0.03885	-1.0
<b>RC</b>	-0.021	0.004	-0.0002	95% (92 – 98)	95% (92 – 98)	0.00088	0
<b>MCMC</b>	0.002	-0.001	0.0001	97% (95 – 99)	97% (95 – 99)	0.00022	0
<b>MCMC-RC</b>	-0.004	0.000	0.0000	89% (85 – 93)	89% (85 – 93)	0.00054	0
<b>INLA-RC</b>	-0.008	0.001	-0.0001	95% (91 – 98)	93% (89 – 96)	0.00035	0
<b>Asymptotic</b>							
<b>True values†</b>	-10.326	1.428	-0.0562				12.7
<b>Latent X</b>	0.021	-0.005	0.0002	92% (88 – 95)	91% (87 – 95)	0.0050	0
<b>Naïve</b>	5.698*	-0.993*	0.0421*	0% (0 – 0)	0% (0 – 0)	0.1271	2.7
<b>RC</b>	-0.002	0.000	0.0000	98% (96 – 100)	98% (96 – 100)	0.0073	0
<b>MCMC</b>	-0.095	0.019	-0.0009	93% (89 – 96)	93% (89 – 96)	0.0055	0
<b>MCMC-RC</b>	-0.005	-0.002	0.0002	93% (89 – 97)	95% (91 – 98)	0.0060	0.1
<b>INLA-RC</b>	0.036	-0.008	0.0004	95% (92 – 98)	96% (93 – 98)	0.0059	0
<b>Weak quadratic</b>							
<b>True values</b>	0	0	0.3				
<b>Latent X</b>	-0.127	0.020	-0.0008	96% (93 – 98)	96% (93 – 98)	0.19	
<b>Naïve</b>	7.475*	1.526*	-0.2260*	77% (71 – 82)	3% (0 – 5)	28.58	
<b>RC</b>	-0.485	0.092	-0.0043	97% (95 – 99)	97% (95 – 99)	0.84	
<b>MCMC</b>	0.257	-0.064	0.0037	97% (95 – 99)	97% (94 – 99)	0.33	
<b>MCMC-RC</b>	-0.081	-0.037	0.0044	99% (98 – 100)	99% (98 – 100)	0.61	
<b>INLA-RC</b>	0.311	-0.072	0.0039	99% (98 – 100)	99% (98 – 100)	0.45	



**Figure 4.6** Simulations with a validation study and measurement error variance equal to the variance of  $X$ : Fitted curves from 200 simulations with a continuous outcome for each correction method (by row) and each shape (by column) compared to the latent  $X$  (black line) and the naïve (red line) fitted curves when a validation study is performed on 30% of 2000 study participants. Darker solid red/blue line is the mean fitted curve, the point-wise mean of curve fits, and the dashed red/blue lines are the mean point-wise 95% confidence or credible bounds as determined by each method.

**Table 4.4 Simulations with a replicate study and measurement error variance equal to the variance of X:** Mean bias in the quadratic model regression estimates, coverage of the 95% confidence intervals or credible intervals and the associated Monte Carlo error 95% confidence interval (MCE 95% CI), average mean integrated square error (MISE) over the exposure range 7-13, and bias in the turning point of the mean curve (where applicable) estimated from 200 simulations with a continuous outcome a replicate performed with two error-prone measures on all study participants after application of the designated method to data generated from three shapes of association (J-shape, asymptotic, and weak quadratic). \*indicates a MCE 95% CI for bias that excludes zero. † indicates that true values were estimated from 10,000 simulations with the latent X values.

Method	Bias			Coverage (MCE 95% CI)		Average MISE	Turning point bias
	$\beta_0$	$\beta_{X_1}$	$\beta_{X_2}$	$\beta_{X_1}$	$\beta_{X_2}$		
<b>J-shaped</b>							
<b>True values</b>	4.86	-1.08	0.06				9.0
<b>Latent X</b>	-0.003	0.000	0.0000	96% (93 – 98)	96% (93 – 98)	0.00011	0
<b>Naïve</b>	-3.872*	0.840*	-0.0450*	0% (0 – 0)	0% (0 – 0)	0.03885	-1.0
<b>RC</b>	-0.005	-0.012	0.0006	96% (93 – 98)	95% (92 – 98)	0.00422	0
<b>MCMC</b>	0.051	-0.010	0.0005	95% (91 – 98)	95% (93 – 98)	0.00040	0
<b>MCMC-RC</b>	0.058	-0.012	0.0006	81% (76 – 86)	79% (73 – 85)	0.00062	0
<b>INLA-RC</b>	0.061	-0.012	0.0006	80% (74 – 86)	79% (73 – 84)	0.00062	0
<b>Asymptotic</b>							
<b>True values†</b>	-10.326	1.428	-0.0562				12.7
<b>Latent X</b>	0.021	-0.005	0.0002	92% (88 – 95)	91% (87 – 95)	0.0050	0
<b>Naïve</b>	5.698*	-0.993*	0.0421*	0% (0 – 0)	0% (0 – 0)	0.1271	2.7
<b>RC</b>	-0.031	0.015	-0.0006	98% (95 – 100)	98% (95 – 100)	0.0092	0
<b>MCMC</b>	-0.140*	0.025*	-0.0011	92% (88 – 96)	93% (89 – 96)	0.0053	0
<b>MCMC-RC</b>	0.013	-0.004	0.0003	89% (85 – 93)	91% (86 – 95)	0.0060	0
<b>INLA-RC</b>	0.006	-0.003	0.0002	90% (85 – 94)	90% (86 – 94)	0.0060	0
<b>Weak quadratic</b>							
<b>True values</b>	0	0	0.3				
<b>Latent X</b>	-0.127	0.020	-0.0008	96% (93 – 98)	96% (93 – 98)	0.19	
<b>Naïve</b>	7.475*	1.526*	-0.2260*	77% (71 – 82)	3% (0 – 5)	28.58	
<b>RC</b>	-0.584	0.015	0.0012	97% (94 – 99)	96% (93 – 99)	0.75	
<b>MCMC</b>	0.151	-0.065	0.0048	95% (91 – 98)	94% (91 – 97)	0.39	
<b>MCMC-RC</b>	0.883	-0.197	0.0106	96% (93 – 99)	96% (93 – 99)	0.51	
<b>INLA-RC</b>	0.846	-0.192	0.0104	96% (93 – 99)	96% (93 – 98)	0.52	



*Figure 4.7 Simulations with a replicate study and measurement error variance equal to the variance of  $X$ : Fitted curves from 200 simulations with a continuous outcome for each correction method (by row) and each shape (by column) compared to the latent  $X$  (black line) and the naïve (red line) fitted curves when replicate study is performed with two error-prone measures on 2000 study participants. Darker solid red/blue line is the mean fitted curve, the point-wise mean of curve fits, and the dashed red/blue lines are the mean point-wise 95% confidence or credible bounds as determined by each method.*

#### 4.5.5 Results: Model fit after the application of measurement error correction methods

All correction methods resulted in considerably better curve fit than the naïve analysis (Figures 4.5-4.7). The bias in the turning point, the bias and coverage of the regression coefficients, and the average MISE for each method and association shape can be found in Tables 4.2-4.4.

##### 4.5.5.1 Validation study setting

Overall, for either measurement error variance when a validation study was present, all measurement error correction methods were able to minimize bias in the turning point and regression coefficients (Tables 4.2 and 4.3). One can see in Figures 4.5 and 4.6 where the mean predicted outcomes are plotted that all correction methods recover the qualitative true shape of the curve, though within the limitation of the quadratic model for the asymptotic shape.

RC had the largest average MISE for all three shapes in this setting while MCMC had the most minimal average MISE. The hybrid methods had average MISE values similar to each other and between those of MCMC and RC; for the higher measurement error, MCMC-RC had average MISE values (4.9-fold, 1.2-fold, and 3.2-fold the latent  $X$ , respectively) slightly higher than those observed for INLA-RC (3.2-fold, 1.2-fold, and 2.4-fold the latent  $X$ , respectively).

The 95% CI/CrIs for  $\beta_{X_1}$  and  $\beta_{X_2}$  maintained coverage at or above the nominal level for all association shapes except for MCMC-RC for the J-shaped association where undercoverage was observed (89% for both  $\beta_{X_1}$  and  $\beta_{X_2}$ ). For the weak quadratic association when  $\sigma_{\eta}^2 = 1$ , both MCMC-RC and INLA-RC showed overcoverage (99%). It is important to note that for all settings the SEs for MCMC-RC and INLA-RC are smaller than those seen for RC. The estimated SEs for MCMC were the smallest of the methods while for RC are always the largest (approximately 2-fold that of MCMC). MCMC-RC and INLA-RC fall in between and very close to each other. Therefore, this overcoverage occurs even when the 95% CIs for hybrid methods are much narrower than the 95% CIs for RC.

##### 4.5.5.2 Replicate study setting

No significant bias in the turning point or regression coefficients was observed for any of the correction methods in this setting. However, while RC may have unbiased regression coefficients, one can observe a departure from the true shape of the association for the J-shaped and asymptotic associations in Figure 4.7 (second row) where the estimated curves and mean predicted outcome values are plotted. For the J-shaped association, the RC mean predicted curve is shifted below the true curve for the range of exposure values plotted. For the asymptotic association, between the range of  $X$  from approximately 9 to 12, the RC mean predicted curve overestimates the outcome. The average MISE is also biased for RC for the J-shaped association. All other correction methods appear to recover close to the true shapes of the curves.



For the asymptotic shape, the RC average MISE was higher than seen for a validation study with the same measurement error variance, but only slightly so (1.8-fold the latent  $X$  average MISE). The average MISE values for MCMC-RC and INLA-RC were much closer in value to each other in this setting reflecting the fact that the estimated SEs for each method were nearly identical.

There was significant undercoverage observed for MCMC-RC and INLA-RC for the J-shaped association (81% and 80% for  $\beta_{X_1}$ , respectively and 79% and 79% for  $\beta_{X_2}$ , respectively) (Table 4.4). The undercoverage was more severe with a replicate study than was observed for a validation study with the same measurement error variance. This is due to greater uncertainty in the estimation of  $E[X_i|W_i, Z_i]$  and  $E[X_i|W_i, Z_i]$  and this uncertainty not being included in the SEs of the regression estimates when only replicate measures are available in lieu of a validation study. While there was apparent undercoverage for the asymptotic shape, given similar undercoverage observed for the latent  $X$ , this was not considered to be a result of the measurement error correction method.

#### *MCMC convergence diagnostics*

All MCMC chains converged to the stationary distribution as determined by  $\hat{R} < 1.1$ . The mean ESS for regression coefficients in the full MCMC model for the J-shaped, asymptotic, and weak quadratic associations was between 18,000 and 30,000 for a validation study with  $\sigma_U^2 = 0.25$ , 11,000 and 21,000 for a validation study with  $\sigma_U^2 = 1$ , and 5,000 and 16,000 for a replicate study with  $\sigma_U^2 = 1$ . In all settings, the J-shaped association had the lowest ESS in the full MCMC model; the only setting in which the ESS was less than 10,000 was the J-shaped association with a replicate study and  $\sigma_U^2 = 1$ . The mean ESS for the latent  $X_i$  values in the MCMC-RC model was approximately 18,000 for all association shapes.

## **4.6 SIMULATION STUDY WITH A BINARY OUTCOME**

In this section, the methods laid out for measurement error correction in the presence of a quadratic transformation of the error-prone variable are applied in a simulation study when the outcome is binary and the substantive model is therefore logistic regression.

### **4.6.1 Simulation study design and evaluation criteria**

The latent exposure  $X$  and the error prone measure  $W$  were generated exactly as for the previous simulation studies and no accurately observed covariates were included (Box 4.1). Only two quadratic shapes are considered in this simulation study, a J-shaped association and a weak quadratic association (Table 4.5). The binary outcome,  $Y_i$ , was generated as a Bernuolli random variable with probability equal to the logit transformation of the linear predictor. The value of each  $\beta_0$  was chosen to result in an event rate for the outcome  $Y$  of 20%.

*Table 4.5 Models for exposure-outcome association shapes with a binary outcome.*

Shape	Linear predictor	Sample size (N)
J-shape	$-1.8 + 0.2 * (X - 9)^2$	2000
Weak quadratic	$-4.5 + 0.03 * X^2$	10,000

The only setting considered was a validation study in 30% of the study population with measurement error variance  $\frac{1}{4}$  the variance of  $X$ , i.e.  $\sigma_U^2 = 0.25$ . For each association shape, 100 simulations were performed ( $2 \times 100$ ).

The criteria for evaluating the fit of the quadratic model after application of measurement error correction can be found in Section 4.5.2. The average MISE for each method was estimated using the outcome probabilities, i.e. the logit transformation of the linear predictor, in lieu of a continuous outcome (Equation 4.9).

#### 4.6.2 Implementation of methods

All methods were implemented exactly as they were in Section 4.5.3 except for the following changes.

For the implementation of the fully Bayesian method in MCMC, there was no scaling of the binary outcome. Three chains were run for each simulation with a burn-in period of 300,000 samples with the last 20,000 burn-in samples evaluated for convergence using  $\hat{R}$ . Another 20,000 samples were collected from each chain for inference. ESS was estimated from these samples.

#### 4.6.3 Fitting the latent $X$ and naïve models

Only the J-shaped association had a turning point observed in the range of exposure simulated at  $X = 9.0$  (Table 4.6). The average MISE for the J-shaped association was 0.027 and for the weak quadratic association was 0.0073.

The addition of measurement error,  $\sigma_U^2 = 0.25$ , in the naïve analysis resulted in the turning point of the J-shaped association being biased -0.3 away from the mean of  $X$  (Table 4.6). The bias toward the null observed in  $\hat{\beta}_{X_2}$  for the J-shaped association was 37% of  $\beta_{X_2}$ . This was near to the 36% reduction predicted by Equation 4.3 which was only approximately true for logistic regression. The bias in  $\hat{\beta}_{X_1}$  was 39% of  $\beta_{X_1}$ . For the weak quadratic association,  $\hat{\beta}_{X_2}$  is biased toward the null by 49%, much higher than predicted by Equation 4.3. Similar to what was observed for linear regression, the undercoverage for the J-shaped association (42% and 33% for  $\beta_{X_1}$  and  $\beta_{X_2}$ , respectively) was much more severe than for the weak quadratic association (92% and 82%, respectively).

*Table 4.6 Simulations with a validation study and a binary outcome and measurement error variance ¼ the variance of X: Mean bias in the quadratic model regression estimates, coverage of the 95% confidence intervals or credible intervals and the associated Monte Carlo error 95% confidence interval (MCE 95% CI), average mean integrated square error (MISE) over the exposure range 7-13, and bias in the turning point of the mean curve (where applicable) estimated from 200 simulations with a continuous outcome, a validation study performed on 30% of 2000 study participants after application of the designated method to data generated from two shapes of association. \*indicates a MCE 95% CI for bias that excludes zero.*

Method	Bias			Coverage (MC error 95% CI)		Average MISE	Turning point bias
	$\beta_0$	$\beta_{X_1}$	$\beta_{X_2}$	$\beta_{X_1}$	$\beta_{X_2}$		
J-shaped, N=2000							
<b>True values</b>	14.4	-3.6	0.2				9.0
<b>Latent X</b>	-0.034	0.010	-0.0007	97% (94 – 100)	97% (94 – 100)	0.027	0
<b>Naïve</b>	-6.620*	1.414*	-0.0746*	42% (32 – 52)	33% (24 – 42)	0.115	-0.3
<b>RC</b>	-0.360	0.073	-0.0035	94% (89 – 99)	94% (89 – 99)	0.043	0
<b>MCMC</b>	0.110	-0.025	0.0013	93% (88 – 98)	93% (88 – 98)	0.039	0
<b>MCMC-RC</b>	-0.057	0.012	-0.0005	93% (88 – 98)	94% (89 – 99)	0.037	0
<b>INLA-RC</b>	-0.100	0.021	-0.0010	93% (88 – 98)	94% (89 – 99)	0.037	0
Weak quadratic, N=10000							
<b>True values</b>	-4.5	0	0.03				
<b>Latent X</b>	-0.267	0.048	-0.0021	97% (94 – 100)	96% (92 – 100)	0.0073	
<b>Naïve</b>	-0.243	0.173*	-0.0146*	92% (87 – 97)	82% (74 – 90)	0.0506	
<b>RC</b>	-0.585*	0.120*	-0.0060*	94% (89 – 99)	95% (91 – 99)	0.0106	
<b>MCMC</b>	-0.424	0.078	-0.0036	97% (94 – 100)	98% (95 – 100)	0.0096	
<b>MCMC-RC</b>	-0.441*	0.090*	-0.0044*	96% (92 – 100)	97% (94 – 100)	0.0086	
<b>INLA-RC</b>	-0.439*	0.089*	-0.0044*	96% (92 – 100)	97% (94 – 100)	0.0086	

#### 4.6.4 Results: Model fit after application of measurement error correction methods

No bias in the turning point of the J-shaped association was observed after correction by any method (Table 4.6). All correction methods reduced bias and improved coverage over the naïve analysis; however, all RC methods (RC, MCMC-RC, and INLA-RC) had significant bias remaining in the regression coefficients for the weak quadratic association. The Bayesian RC methods, MCMC-RC and INLA-RC, are subject to the same approximation for logistic regression as standard RC.

MCMC had no discernable bias in the mean regression coefficients for either association shape (given that there were only 100 simulations per scenario), but the average MISE was slightly higher than for MCMC-RC and INLA-RC.

No indication of undercoverage, within the limits of the smaller simulation study, was observed. The uncertainty in the estimation of the substantive model parameters was going to be much higher for logistic regression than for linear regression; this has the effect of minimizing the relative impact of not including the uncertainty from the measurement error and exposure models for MCMC-RC and INLA-RC.

### *MCMC convergence diagnostics*

All MCMC chains converged to the stationary distribution as determined by  $\hat{R} < 1.1$ . The mean ESS for the regression coefficients in the full MCMC model for the J-shaped and weak quadratic associations was 3,955 and 3,206, respectively. The mean ESS for the latent  $X_i$  values in the MCMC-RC model was approximately 18,000 in both settings.

## 4.7 SUMMARY

### **Box 4.2 Simulation study summary**

- For linear regression, only RC was observed to have significant bias in the corrected curve in the replicate study setting (Figure 4.7). The mean regression coefficients did not appear to have significant bias. For logistic regression, RC and the hybrid methods had significant bias in the mean regression coefficients for the weak quadratic association.
- RC had the greatest variance in the regression coefficients and the highest average MISE across simulations. The 95% CIs for RC were wider than the 95% CI/CrIs for other methods although the nominal coverage was maintained.
- The fully Bayesian method using MCMC consistently resulted in the lowest average MISE across all linear regression settings while maintaining the nominal coverage for the regression estimates.
- The hybrid methods performed similarly to each other and consistently between MCMC and RC. However, INLA-RC performed slightly better than MCMC-RC in a number of scenarios either having a lower average MISE, better coverage, or less variability.

In this chapter, I introduced hybrid Bayesian RC methods using either MCMC or INLA for the Bayesian posterior estimates of  $E[X_i | \mathbf{W}_i, \mathbf{Z}_i]$  and  $E[X_i^2 | \mathbf{W}_i, \mathbf{Z}_i]$ . These were evaluated in simulation studies alongside previously published extensions of RC and MCMC.

All methods of measurement error correction applied in this chapter improved the fit of the quadratic model over the naïve analysis. That is, the true turning point was recovered and, in most cases, bias in the regression coefficients was minimized. The average MISE was consistently most minimal for MCMC and largest for RC, reflecting the greater variability in the curve fits estimated by this method, with the hybrid methods in between. A more specific summary of the simulation studies can be found in Box 4.2.

MCMC-RC had a much higher ESS (approx. 18,000) while INLA-RC drew only 3,000 samples (equivalent to an ESS of 3,000). This number of samples was sufficient in this setting as evidenced by the performance of INLA-RC but a larger sample size may be necessary to minimize Monte Carlo error in other settings. It is important to note that the INLA method for the classical measurement error model and a normally distributed exposure is exact.

The hybrid methods had below nominal coverage in general (with the exception of the weak quadratic association for a validation study with  $\sigma_{ij}^2 = 1$ ). This was anticipated given that the uncertainty in the estimation of the measurement error and exposure model parameters was not incorporated into the final SEs; only the uncertainty in the estimation of the substantive model parameters were included. Use of bootstrapping, most feasible with INLA-RC, would improve these estimates.

While MI and INLA have promising application to measurement error correction in some settings, for the setting described in this and future chapters, it was felt that each had reached its limit.

In the next chapter, I will evaluate whether RC, MCMC, MCMC-RC, and INLA-RC can recover power lost to measurement error in the naïve analysis while maintaining appropriate type I error rates.

# 5 MEASUREMENT ERROR CORRECTION FOR SELECTION OF A QUADRATIC TERM

---

## 5.1 AIMS & OVERVIEW

In the previous chapter, the substantive model was specified to be the quadratic model introduced in Equation 4.1. In this chapter, I aim to extend the measurement error correction methods from the previous chapter (RC, MCMC, MCMC-RC, and INLA-RC) to the situation in which the best functional form of the error-prone exposure is unknown. It is often of interest to test whether a squared term should be included in the model; this can be done easily by testing the null hypothesis that  $\beta_{X_2} = 0$ . However, in order to lay the groundwork for model selection when the models are not nested, I aim to assess the use of measures of predictive accuracy as well.

In this chapter, I also aim to introduce an additional fully Bayesian approach which makes use of methods typically applied to Bayesian variable selection. In this chapter, this approach is applied to find the posterior probability that a squared term should be included in the model.

The effect of classical measurement error on the naïve model selection process is to reduce the power to detect non-linear features due to the increased variability in the data [1–3]. Specifically, the power to select the quadratic model over the linear model when the true association curve is, in fact, quadratic is reduced. Furthermore, the validity of hypothesis testing, i.e. whether the type I error rate remains at its nominal level asymptotically, is not necessarily retained in the naïve analysis with multiple error-prone predictors [1,2]. To my knowledge, no publications have specifically addressed the selection of a squared term in the presence of measurement error.

Methods are demonstrated in the same settings and for the same shapes of association as in the previous chapter.

## 5.2 HYPOTHESIS TESTING

In this chapter, the null hypothesis that  $\beta_{X_2} = 0$ , where  $\beta_{X_2}$  is the regression coefficient of the squared term from Equation 4.1, is evaluated at a nominal type I rate of 5%. For an  $X$  observed without error, this hypothesis may be tested after fitting the quadratic model to the data by comparing  $\hat{\beta}_{X_2}/SE(\hat{\beta}_{X_2})$  (the “Wald statistic”) to the appropriate distribution and rejecting the null hypothesis if  $p < 0.05$ . This is the equivalent of assessing whether the 95% CI excludes zero.

Alternatively, if both the linear (Equation 1.1) and quadratic model (Equation 4.1) are fit to the data, the likelihood ratio test may be performed. Under the null hypothesis that  $\beta_{X_2} = 0$ , to perform the likelihood ratio test twice the difference in the log-likelihoods between the models is

compared to a  $\chi^2$ -distribution with one degree of freedom. The likelihood ratio test and the Wald test described above are asymptotically equivalent.

Assuming  $X_i^{(1)}$  and  $X_i^{(2)}$  are two correlated error-prone variables subject to classical measurement error in a linear regression model (i.e.  $E[Y_i|X_i, \mathbf{Z}_i] = \beta_0 + \beta_{X,1}X_i^{(1)} + \beta_{X,2}X_i^{(2)} + \boldsymbol{\beta}_Z^T\mathbf{Z}_i$ ), Carroll et al showed that the naïve test of the hypothesis that  $\beta_{X,2} = 0$  is not necessarily valid [1]. That is because each naïve regression coefficient is a function of both  $\beta_{X,1}$  and  $\beta_{X,2}$ . I showed in Section 4.3 that for the quadratic model (Equation 4.1,  $g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_{X_1}X_i + \beta_{X_2}X_i^2 + \boldsymbol{\beta}_Z^T\mathbf{Z}_i$ ), the naïve coefficient of the squared term is equal to  $\beta_{X_2}\lambda^2$  (Equation 4.3). Therefore, the naïve test of the hypothesis that  $\beta_{X_2} = 0$  is valid because if  $\beta_{X_2} = 0$  then  $\beta_{X_2}\lambda^2 = 0$ . However, a similar test of the linear coefficient,  $\beta_{X_1} = 0$ , would not necessarily be valid (Equation 4.4).

Due to the known effect of measurement error to reduce power [1], investigators may decide to apply hypothesis testing with a relaxed type I error rate of 10% or 20% (i.e.  $\alpha$ -level of 0.10 or 0.20). Whether using a higher  $\alpha$ -level than the traditional 0.05 in the presence of measurement error for model selection is appropriate will depend on the application but may aid in exploratory research or sensitivity analyses.

For MCMC-RC and INLA-RC, the SE for  $\beta_{X_2}$  is estimated and compared to the  $t$ -distribution to find the 95% CI around  $\beta_{X_2}$ . For RC, basic bootstrap 95% confidence intervals were estimated [77]. Fitting the quadratic substantive model using MCMC, the HPD 95% CrI is analogous to the 95% CIs used in frequentist methods (Section 2.4.2). One may apply similar hypothesis testing to test whether the interval excludes zero; however, it is not necessarily expected to maintain a type I error rate at 5%. Whereas frequentist 95% CIs summarize the long-run repeatability of the likelihood distribution given the model, Bayesian 95% CrIs summarize the estimated posterior distribution of the parameter which is comprised of both the likelihood and the prior information. The likelihood ratio test is inappropriate for Bayesian analysis for this same reason – the posterior is a combination of both the likelihood and the prior distributions.

### 5.3 MEASURES OF PREDICTIVE ACCURACY (INFORMATION CRITERIA)

When it is of interest to compare models that are not nested, the above hypothesis testing is not applicable. Instead, measures of predictive accuracy are applied. These are termed “information criteria” [87] and are a function of the log-likelihood with some penalty for model complexity. These metrics are an attempt to put models on the same scale for direct comparison rather than estimating probability or performing hypothesis testing [87]. While unnecessary for model selection when comparing a quadratic model to a linear model, the effect of measurement error

on the use of the measures was examined as it may inform whether they should be used in more general model selection settings.

For an  $X$  observed without error, the Akaike information criterion (AIC) can be used to determine the best model after fitting both the quadratic model (Equation 4.1) and the linear model (Equation 1.1) [115]. The AIC is equal to twice the negative log-likelihood with an added penalty for complexity that is twice the number of parameters. The model with the lowest AIC is accepted as the best model. The AIC can be used as a measure of predictive accuracy for any of the frequentist methods presented in this chapter, including the hybrid methods.

For nested models, it can be shown that the AIC is asymptotically equivalent to hypothesis testing as described above (i.e. the likelihood ratio test or the Wald test) at an  $\alpha$ -level of 16% [116,117].

For standard Bayesian analysis, the deviance information criterion (DIC) is an alternative metric to the AIC used in Bayesian analysis [118,119]. The DIC, introduced in 2002, is constructed by taking negative twice the log-likelihood (relying on the posterior mean for the log-likelihood instead of the maximum likelihood estimate, i.e. the deviance) and adding a measure of the effective number of parameters [118]. The effective number of parameters can be estimated from two MCMC chains by taking the log ratio of the likelihoods of each chain, i.e.  $p_D = \log\{f(D|\theta^1)/f(D|\theta^2)\}$ , where  $D$  indicates the data and  $\theta^1$  and  $\theta^2$  independent samples of the posterior of  $\theta$ , the model parameters. This solution may be derived from the Kullback-Leibler information divergence between predictive distributions for different values of  $\theta$  [118].

To use the DIC, one fits a Bayesian model using MCMC. For our purposes, a Bayesian model with the quadratic model and the appropriate measurement error and exposure models as described in Section 4.4.2 is fit; then another Bayesian model with the linear model and appropriate measurement error and exposure models as implemented in Chapter 3 (Section 2.3) are fit. The DIC is estimated from samples drawn from the posterior distribution for each model and the model with the lowest DIC is accepted as the best model.

#### **5.4 BAYESIAN TRANSFORMATION SELECTION (BTS)**

As an alternative to the use of either DIC or using the 95% CrIs to determine the best model, one may employ methods traditionally used for Bayesian variable selection; that is, selecting a subset of variables from a larger set. Bayesian variable selection encompasses a wide range of specific methods [120]. A few publications have specifically explored Bayesian variable and transformation selection together [121–123], although only one publication specifically explores selection of fractional polynomial transformations [55]. In this thesis, I focus solely on selection of the best transformation for a single error-prone variable; therefore, this approach is referred to as “Bayesian transformation selection” (BTS) when adapted to this context. As with DIC, this



approach is unnecessary for the selection of the quadratic model over the linear model. BTS was applied in this setting in order to assess the performance in a simpler setting before extending the method to the selection of fractional polynomial models for an error-prone exposure in the next chapter.

I will first describe the Kuo and Mallick method of Bayesian variable selection [60] (Section 5.4.1). This method is arguably the simplest to implement and requires no tuning variables (i.e. parameters with no other purpose than to ensure good mixing of the sampler). In Section 5.4.2, I adapt the Kuo and Mallick method to determine the posterior probability of a quadratic model (Equation 4.1) over a linear model (Equation 1.1) in the presence of exposure measurement error. To my knowledge, this method has not previously been used to select the best transformation(s) of a single variable.

Incorporation of the selection of the squared term into the Bayesian model has the additional advantage of being able to include uncertainty due to model selection in the 95% CrIs of the regression coefficients. Uncertainty due to model selection is not likewise incorporated into the 95% CI/CrIs of any of the other methods in this chapter.

#### 5.4.1 Kuo and Mallick method applied to Bayesian variable selection

Kuo and Mallick outlined a Bayesian method of variable selection designed to select a reduced number of variables for a generalized linear model [60]. Given a number of predictor variables,  $\mathbf{V}$ , from  $j = 1, 2, \dots, J$ , the regression coefficient,  $\beta_j$ , for each variable is replaced with a composite parameter which can take a “spike and slab” prior. The composite parameter  $\varphi_j$  is composed of the original regression coefficient which takes a Gaussian prior and an indicator value,  $I_j$ , which takes a Bernoulli prior with probability  $\pi$ , i.e.  $\varphi_j = I_j\beta_j$ . If  $I_j$  takes on the value one, then the value of  $\varphi_j$  is drawn from the Gaussian distribution associated with  $\beta_j$  or the “slab”. If  $I_j$  takes on the value zero, then  $\varphi_j$  is also zero or takes its value from the density “spike” at zero. If  $\varphi_j$  is zero, then the variable  $V_j$  contributes nothing to the linear predictor.

A generalized linear model including the spike and slab prior for all regression coefficients is:

$$5.1 \quad g(E[Y_i|X_i]) = \beta_0 + \beta_1 I_1 V_{1i} + \dots + \beta_J I_J V_{Ji} = \beta_0 + \mathbf{I}\boldsymbol{\beta}\mathbf{V}_i = \beta_0 + \boldsymbol{\varphi}\mathbf{V}_i.$$

An appropriate hyperprior distribution may be specified for  $\pi$  reflecting either no prior knowledge (i.e. a uniform distribution from 0 to 1) or some prior knowledge (i.e. a normal distribution centered at an expected value). Alternatively,  $\pi$  may be assigned a fixed value reflecting the desired proportion of variables to be retained (or selected) in the final model.

In the samples drawn from this model using MCMC, sampled values being indicated by  $\tilde{\cdot}$  (e.g.  $\tilde{\mathbf{I}}$ ), different values of the vector  $\mathbf{I}$  can be said to define different substantive models. The model

drawn most frequently or having the highest posterior probability [60] may be said to be the nominal best model. As several models may have a high posterior probability or the performance of a specific model within the set of possible models may be of interest, the magnitude of the evidence in favor of one model versus another may be assessed by use of the Bayes factor (BF) which is described in Section 5.4.2.

When  $\tilde{I}_j = 0$ , particularly over many consecutive samples,  $\beta_j$  continues to be sampled but contributes nothing to the linear predictor. As a result, the value of each  $\tilde{\beta}_j$  may rarely be in a region of the model space where there is posterior support for  $I_{X_2}$  to change from 0 to 1. Good mixing in this setting is defined by the frequency with which  $\tilde{I}_j$  changes value [120]. Appropriate mixing can be quite challenging in this setting. Therefore, the priors for the regression coefficients  $\beta$  cannot be very vague.

#### 5.4.2 Kuo and Mallick method applied to selection of the quadratic model (Bayesian transformation selection)

In order to apply the above method to the selection of the quadratic model (Equation 4.1) versus the linear model (Equation 1.1), the “spike and slab” prior is applied only to the regression coefficient of the squared term:

$$5.2 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) = \beta_0 + \beta_{X_1}X_i + I_{X_2}\beta_{X_2}X_i^2 + \beta_{\mathbf{Z}}^T\mathbf{Z}_i.$$

Each sample of  $I_{X_2}$  takes the value either 1 or 0 as dictated by the draw from the Bernoulli distribution with probability  $\pi$ ; if  $\tilde{I}_{X_2} = 1$  then the  $X_i^2$  term is included in the model and if  $\tilde{I}_{X_2} = 0$  then it is not. If the hyperparameter  $\pi$  is assigned the fixed value 0.5, neither inclusion nor exclusion of the  $X_i^2$  term is favored; this value is suggested by George and McCulloch in order to be uninformative [124]. Further discussion on the effect of the choice of value for  $\pi$  can be found in Section 5.5.8 as demonstrated in a sensitivity analysis.

A BF may be used to assess whether there is significant evidence in favor of the quadratic model over the linear model (Box 5.1 presents a common interpretation of the value of the BF [125,126]). The BF may be considered as a constant which quantifies the magnitude of the evidence in the data for one model (or hypothesis) over another [127]. Specifically, the posterior odds are equivalent to the prior odds multiplied by the BF:

<b>Box 5.1 Bayes Factor (BF)</b>	
BF =	$\frac{\text{posterior odds}}{\text{prior odds}}$
BF = 1 – 3	weak evidence
BF = 3 – 10	moderate evidence
BF = 10 – 100	strong evidence

$$5.3 \quad \frac{f(I_{X_2} = 1|\mathbf{W}, X, Y, \mathbf{Z}; \theta)}{f(I_{X_2} = 0|\mathbf{W}, X, Y, \mathbf{Z}; \theta)} = \frac{f(I_{X_2}=1)}{f(I_{X_2}=0)} \times \frac{f(\mathbf{W}, Y, \mathbf{Z}|X, I_{X_2} = 1)}{f(\mathbf{W}, Y, \mathbf{Z}|X, I_{X_2} = 0)}$$

The posterior probability of the quadratic model is equivalent to the proportion of samples wherein  $\tilde{I}_{X_2} = 1$  is sampled. If  $\pi$  is assigned a fixed value, then the prior probability of the quadratic model is  $\pi$  and the prior probability of the linear model is  $1 - \pi$ . The prior odds are then  $\pi/(1 - \pi)$ .

For very strong quadratic associations, when the model is believed to have converged to the stationary distribution, all samples will be from the quadratic model. In this case,  $I_{X_2}$  will have a posterior mean value of 1,  $E[\tilde{I}_{X_2} | \mathbf{W}, X, Y, \mathbf{Z}, \boldsymbol{\theta}] = 1$ , as well as a 95% CrI from 1 to 1. Regression coefficients and their variances can then be estimated from the posterior in the same manner as for standard MCMC. However, for a weaker quadratic association, when  $E[\tilde{I}_{X_2} | \mathbf{W}, X, Y, \mathbf{Z}, \boldsymbol{\theta}]$  is between 0 and 1, non-inclusive, some of the samples will be from the quadratic model and some will be from the linear model. In these cases,  $E[\tilde{I}_{X_2} | \mathbf{W}, X, Y, \mathbf{Z}, \boldsymbol{\theta}]$  represents the posterior probability that inclusion of the squared term is appropriate.

Unfortunately, in cases where there are a significant number of samples drawn from both models, convergence diagnostics such as  $\hat{R}$  cannot be used to assess convergence. This is characteristic of MCMC sampling with mixture models wherein “label-switching” occurs between samples to indicate different models being sampled [28,128].

As in the standard application of the approach, the prior for  $\beta_{X_2}$  must not be too vague (specific prior values are discussed in Section 5.5.2), otherwise mixing will be poor and the sampler will rarely if ever change  $\tilde{I}_{X_2} = 0$  to  $\tilde{I}_{X_2} = 1$  [120]. Appropriate scaling of the untransformed and transformed terms, i.e.  $X$  and  $X^2$ , as outlined in Section 4.5.3, is important for efficient mixing and convergence of this model [28,85].

In order to select the best model describing the exposure-outcome relationship while accommodating measurement error, this modified substantive model (Equation 5.2) and specified priors are included in the joint model with the specified exposure and measurement error models in the modular fashion discussed previously (Section 2.3).

The 95% CrI and variance estimates for the regression coefficients include uncertainty due to model selection as well as measurement error and population sampling.

## **5.5 SIMULATION STUDY EXTENSION TO MODEL SELECTION FOR A CONTINUOUS OUTCOME**

### **5.5.1 Simulation study design**

The simulations study in Chapter 4 (Section 4.5; Box 4.1) was extended to include a linear model in order to enable assessment of type I error rates in selection of the quadratic model over the

linear model alongside assessment of power to select the quadratic model for the three quadratic associations. In this section, the four models used to generate  $Y_i$  include the three quadratic association shapes illustrated in Figure 4.1 and a linear association defined in Box 5.1. The parameters of the linear association model were chosen because they are the maximum likelihood estimates when fitting a simple linear regression to data generated from the weak quadratic model; the residual variance ( $\sigma_{Y|X}^2$ ) is identical in the two models. Simulation study details are otherwise identical to Section 4.5 but can be found in Box 5.2 for reference.

### Box 5.2 Simulation study design

Simulations per scenario: 200

Sample size:  $N = 2000$

True exposure:  $X_i \sim N(10, 1)$

Classical error model:

$$W_{ij} = X_i + U_{ij}, \quad U_{ij} \sim N(0, \sigma_U^2), \quad j = 1, 2$$

Outcome models:

$$\text{J-shaped association: } Y_i \sim N(0.06(X_i - 9)^2, 0.14)$$

$$\text{Asymptotic association: } Y_i \sim N(10(4 - X_i)^{-1}, 0.3)$$

$$\text{Weak quadratic association: } Y_i \sim N(0.3X^2, 6)$$

$$\text{Linear association: } Y_i \sim N(-30 + 6X, 6)$$

Settings:

- 1) Validation study in 30% of participants,  $\sigma_U^2 = 0.25$
- 2) Validation study in 30% of participants,  $\sigma_U^2 = 1$
- 3) Replicate study with 2 measures in all participants,  $\sigma_U^2 = 1$

### 5.5.2 Criteria for evaluating power and type I error

The power to select the quadratic model was assessed using the proportion of simulated data sets for which the quadratic model was selected when the association shape is one of the three non-linear shapes (J-shaped, asymptotic, and weak quadratic). The type I error rate, that is, the rate at which the quadratic model was selected when the association was linear, was assessed using the proportion of simulated data sets for which the quadratic model was selected when the association shape was linear. The power and type I error rate were assessed for the latent  $X$  analysis (fitting the latent  $X$  to the linear and quadratic models), the naïve analysis (fitting the first error-prone measure to the linear and quadratic models), and for each measurement error correction method (RC, MCMC, MCMC-RC, INLA-RC) using hypothesis testing (Section 5.2) and AIC/DIC (Section 5.3).

The concepts of power and type I error are frequentist in nature as they are measures of long-run frequencies in hypothesis testing. Bayesian posteriors are not expected to maintain nominal type I error rates in the same fashion due to the introduction of prior information. However, examination of power and type I error of Bayesian methods can be informative as to the ability

of the method to discern evidence for one model over another given the available data and specified priors.

Additionally, for the BTS method incorporating measurement error correction, the fit of the quadratic model in samples where it was selected was evaluated according to the criteria in Section 4.5.2 (i.e. bias in the turning point, bias in the regression coefficients, and MISE). The average posterior probability that  $I_{X_2} = 1$  (and therefore the squared term is included), and the empirical 95% range of this quantity, was estimated alongside the median BF and the interquartile range of the BF. Whether enough evidence was observed to select the quadratic model was determined using several cut-off values of the BF (i.e. 1, 3, and 10) comparing the samples for the quadratic model to the samples for the linear model; these quantities were used to assess power (when the underlying model was a quadratic association) and type I error (when the underlying model was a linear association).

### 5.5.3 Method implementation

For RC, the quadratic model was estimated as detailed in Section 4.4.1 and the linear model was estimated as detailed in Section 2.2. For the quadratic model fit, 1,000 bootstraps were used to estimate variance for the regression parameter estimates and basic bootstrap 95% confidence intervals were estimated [77].

The implementation of MCMC as outlined in this chapter requires that two separate Bayesian models be fit, one with the quadratic substantive model and the other with the linear substantive model. Prior values for fitting the linear substantive model are identical to their quadratic counterparts exempting the prior for  $\beta_{X_2}$ . Standardization of variables, burn-in, and samples collected are identical to those outlined in Section 4.5.3.

No change to the Bayesian part of the MCMC-RC or INLA-RC methods was required in order to fit additional models; therefore, the prior values and implementation details were identical to those outlined in Section 4.5.3.

The prior values specified for BTS are also identical to those specified for standard MCMC with the following exceptions. The prior for  $\beta_{X_2}$  is a normal distribution with mean zero and variance of 100 and the prior for  $I_{X_2}$  is a Bernoulli distribution with  $\pi = 0.5$ . A sensitivity analysis is performed where  $\pi = 0.05$ . Six chains are run for each simulation, three with  $I_{X_2}$  starting at zero and three with  $I_{X_2}$  starting at one, with a burn-in of 80,000.  $\hat{R}$  is estimated from the last 10,000 burn-in samples. An additional 10,000 samples are then drawn from each chain for inference. The effective sample size is estimated from these samples.

BTS was also performed with the latent  $X$  without the measurement error or exposure models. In this case, only priors applying to the substantive model were specified as above.

#### 5.5.4 Model selection using the latent $X$

To better understand the properties of the four associations simulated in this section, the latent  $X$  for each association was fitted to the linear (Equation 1.1) and quadratic models (Equation 4.1). The likelihood ratio test was applied, and the mean p-value associated with the test statistic indicates the ease of appropriate model selection for that association.

The mean p-value associated with the likelihood ratio test comparing the two models for the J-shaped association was  $7 \times 10^{-109}$  and for the asymptotic association was  $2 \times 10^{-16}$ . These very significant p-values indicate that the quadratic model is a much better fit for these associations than the linear model. Either use of hypothesis testing or AIC selected the quadratic model in all simulations.

For the weak quadratic association, after fitting the latent  $X$  to the linear and quadratic models, the likelihood ratio test had a mean p-value of 0.022. Observing whether the 95% CI for  $\beta_{X_2}$  included zero resulted in the selection of the quadratic model in 87% of simulations when the  $\alpha$ -level was 5% (Table 5.1). Use of the AIC resulted in selection of the quadratic model in 97% of simulations for weak quadratic association.

When the latent  $X$  was fit to both the quadratic and linear models when the underlying association was linear, the likelihood ratio test had an associated mean p-value of 0.05 for selection of the quadratic model. The quadratic model was selected by observing whether the 95% CI for  $\beta_{X_2}$  included zero at the nominal  $\alpha$ -level rate of 5% (Table 5.1). Use of the AIC resulted in selection of the quadratic model in 16% of simulations for the linear association.

#### 5.5.5 Model selection using the naïve analysis

In the naïve analysis, the first error-prone measure was fit to the linear model (Equation 1.2) and the quadratic model (Equation 4.2).

For either  $\sigma_U^2 = 0.25$  or 1, the naïve analysis selected the quadratic model for the J-shaped or asymptotic associations in all simulations using either hypothesis testing or AIC. When  $\sigma_U^2 = 1$ , the mean p-value associated with the likelihood ratio test was still very significant for both the J-shaped and asymptotic associations ( $3 \times 10^{-17}$  and  $8 \times 10^{-4}$ ).

For the weak quadratic association, hypothesis testing with an  $\alpha$ -level of 5% had a reduction in power of 28% and 65% for  $\sigma_U^2 = 0.25$  and 1, respectively, highlighting the loss of power

**Table 5.1 Simulations to determine the power and type I error rate:** 200 simulations performed to find the power and type I error rate for selection of the quadratic model when the shape is either a weak quadratic or a linear association as determined by 1) the exclusion of the zero from the confidence interval or credible interval at three nominal type I error rates ( $\alpha$ ); 2) the Akaike information criteria (AIC) or the deviance information criteria (DIC); for the MCMC method only). Applied in three settings: a validation study in 30% of 2000 study participants with measurement error variance  $\frac{1}{4}$  the variance of  $X$  ( $\sigma_U^2=0.25$ ); a validation study with  $\sigma_U^2 = 1$ ; or a replicate study with two measures in all participants with  $\sigma_U^2 = 1$ .

		Hypothesis testing $H_0: \beta_{X_2} = 0$			AIC/DIC	Hypothesis testing $H_0: \beta_{X_2} = 0$			AIC/DIC
		$\alpha=5\%$	$\alpha=10\%$	$\alpha=20\%$		$\alpha=5\%$	$\alpha=10\%$	$\alpha=20\%$	
		Power to select quadratic fit for a weak quadratic association				Type I error rate in selecting the quadratic fit given a linear association			
Validation study $\sigma_U^2 = 0.25$	<b>Latent X</b>	87%	96%	98%	97%	5%	10%	20%	16%
	<b>Naïve</b>	59%	72%	87%	83%	4%	8%	17%	13%
	<b>RC</b>	59%	72%	87%	83%	5%	8%	17%	13%
	<b>MCMC</b>	78%	87%	93%	70%	6%	11%	19%	17%
	<b>MCMC-RC</b>	72%	82%	92%	89%	5%	11%	17%	14%
	<b>INLA-RC</b>	72%	82%	93%	89%	5%	11%	17%	15%
Validation study $\sigma_U^2 = 1$	<b>Latent X</b>	87%	96%	98%	97%	5%	10%	20%	16%
	<b>Naïve</b>	22%	32%	50%	42%	3%	9%	16%	13%
	<b>RC</b>	19%	32%	49%	42%	3%	8%	16%	13%
	<b>MCMC</b>	72%	84%	90%	73%	4%	8%	18%	19%
	<b>MCMC-RC</b>	48%	61%	72%	70%	1%	5%	16%	10%
	<b>INLA-RC</b>	50%	61%	73%	70%	1%	5%	13%	10%
Replicate study $\sigma_U^2 = 1$	<b>Latent X</b>	87%	96%	98%	97%	5%	10%	20%	16%
	<b>Naïve</b>	22%	32%	50%	42%	3%	9%	16%	13%
	<b>RC</b>	19%	31%	50%	42%	4%	7%	17%	13%
	<b>MCMC</b>	63%	75%	85%	76%	5%	11%	18%	31%
	<b>MCMC-RC</b>	47%	58%	73%	70%	4%	10%	20%	14%
	<b>INLA-RC</b>	47%	59%	73%	68%	4%	9%	20%	15%

associated with an increase in measurement error. Using the AIC, the reduction in power was 14% for  $\sigma_U^2 = 0.25$  and 55% for  $\sigma_U^2 = 1$ .

The naïve analysis had a slight drop in type I error observed from the latent  $X$  analysis for both methods of selection.

### 5.5.6 Results: Model selection using hypothesis testing

All methods resulted in selection of the quadratic model in all simulations and settings for the J-shaped association and in all simulations but one for the asymptotic association. When the underlying association was linear, all methods had type I error rates at or below the nominal level. The Bayesian and partly Bayesian methods were not expected to maintain nominal rates of type I error, but with appropriate prior specification, there is no reason to expect an increase in type I error for MCMC-RC and INLA-RC. Therefore, the focus in this section was on the power to select the quadratic model for the weak quadratic association.

Of the methods, MCMC was the most successful at recovering power lost due to measurement error in all settings (Table 5.1). This was due to the incorporation of prior information. Even with relatively uninformative priors distributions, the use of Bayesian analysis allowed us to incorporate more of what we knew about the parameters (e.g. that variance estimates after scaling were near one). Therefore, estimates of  $\beta_{X_2}$  were both less biased and had improved efficiency allowing one to more frequently reject the null hypothesis that  $\beta_{X_2} = 0$ . Power gains were much larger with a validation study than with a replicate study with equivalent measurement error (5-9%).

RC had the same or slightly lower power to detect the weak quadratic association as the naïve analysis (Table 5.1). In the previous chapter, it was observed that RC had the highest average MISE among the methods. The MISE reflects both bias and variability in the predicted outcome values. The empirical SE for  $\beta_{X_2}$  for RC was approximately double that of MCMC (0.24 vs 0.12 for a validation study with  $\sigma_U^2 = 1$ ). Bootstrapping incorporates the uncertainty in estimating the parameters for RC resulting in appropriately wide CIs to achieve good coverage. While RC is successful in reducing bias, it does not increase efficiency of the estimates. If bootstrapping is not used, RC results in power identical to the naïve analysis (confirmed in sensitivity analysis not shown). Given the bootstrapped 95% CIs for  $\beta_{X_2}$ , RC will include zero slightly more often than the naïve analysis as was observed.

MCMC-RC and INLA-RC had very similar power gains to each other, though less than observed for MCMC (Table 5.1). Power gains were slightly larger with a validation study than with a replicate study (0-3% larger). The hybrid methods still benefit from the prior information specified for the measurement error and exposure models but derive no such benefit in fitting the substantive model. As the uncertainty due to measurement error is not incorporated into the 95% CIs for  $\beta_{X_2}$ , incorporation of this uncertainty may reduce gains in power. This effect is likely to be relatively small (i.e. there would still be a significant gain in power) because 1) the observed undercoverage for MCMC-RC and INLA-RC is not severe (Table 4.2-4.4) and 2) the empirical SEs for MCMC-RC and INLA-RC for  $\beta_{X_2}$  (both 0.16 for a validation study with  $\sigma_U^2$ ) are still significantly smaller than observed for RC.

I performed a sensitivity analysis wherein the INLA-RC method was implemented with bootstrapping to estimate a 95% CI around  $\beta_{X_2}$  for only the validation study setting (same 200 simulated data sets per scenario as above) with  $\sigma_U^2 = 1$  for the weak quadratic and linear associations. For each simulation, 100 bootstraps were used to estimate basic bootstrap 95% CIs [77]. As expected, at an  $\alpha$ -level of 5%, only a small drop in power to select the weak quadratic association was observed (49%) while the type I error in selecting the quadratic model for the



**Table 5.2 Quadratic model fit from Bayesian transformation selection (BTS) samples including mean bias in the quadratic model regression estimates, coverage of the 95% confidence intervals or credible intervals and the associated Monte Carlo error 95% confidence interval (MCE 95% CI), average mean integrated square error (MISE) over the exposure range 7-13, and bias in the turning point of the mean curve (where applicable) estimated from 200 simulations for each setting and association (J-shaped, asymptotic, and weak quadratic). Applied to three settings with either a validation study performed on 30% of 2000 study participants or a replicate study where all 2000 participants have two error-prone measures and measurement error variance ( $\sigma_U^2$ ) as indicated. The prior probability ( $\pi$ ) for the Bernoulli distribution for  $I_{X_2}$  is 0.5. \*indicates an MCE 95% CI for bias that excludes zero. † indicates that true values were estimated from 10,000 simulations with the latent  $X$  values. TP indicates turning point.**

Method	$\sigma_U^2$	Bias			Coverage (MCE 95% CI)		Average MISE	TP bias
		$\beta_0$	$\beta_{X_1}$	$\beta_{X_2}$	$\beta_{X_1}$	$\beta_{X_2}$		
<b>J-shaped</b>								
<b>True values</b>		4.86	-1.08	0.06				9.0
Validation	0.25	0.004	-0.001	0.0001	97% (94 – 99)	97% (94 – 99)	0.00016	0
Validation	1	-0.004	0.000	0.0000	97% (95 – 99)	98% (95 – 99)	0.00022	0
Replicate	1	0.040	-0.008	0.0004	95% (92 – 98)	95% (92 – 98)	0.00040	0
<b>Asymptotic</b>								
<b>True values†</b>		-10.326	1.428	-0.0562				12.7
Validation	0.25	0.032	-0.008	0.0004	91% (86 – 95)	92% (88 – 95)	0.0051	0
Validation	1	-0.090	0.018	-0.0009	93% (89 – 97)	93% (89 – 97)	0.0055	0
Replicate	1	-0.133*	0.024*	-0.0011	93% (89 – 96)	94% (90 – 97)	0.0053	0
<b>Weak quadratic</b>								
<b>True values</b>		0	0	0.3				
Validation	0.25	-0.029	-0.017	0.0019	95% (92 – 98)	95% (91 – 98)	0.25	
Validation	1	0.242	-0.062	0.0037	97% (94 – 99)	97% (94 – 99)	0.33	
Replicate	1	-0.090	-0.021	0.0028	96% (93 – 98)	96% (93 – 99)	0.38	

linear association was 2%. The coverage of the 95% CIs estimated in this fashion was 98% (MCE 95% CI: 96 – 100) for both the weak quadratic association and linear association.

### 5.5.7 Results: Model selection using AIC/DIC

It can be inferred from Table 5.1 that for the RC and the hybrid methods the observed power and type I error rates associated with use of the AIC were similar to what would be expected from those methods using hypothesis testing at a nominal  $\alpha$ -level of 16%. Use of AIC is therefore valid and useful after measurement error correction by these methods.

Using DIC with Bayesian analysis using MCMC, when  $\sigma_U^2 = 0.25$  and a validation study is present, the quadratic model was selected in far fewer simulations than the naïve analysis when the underlying association was the weak quadratic association (70% vs. 83%) (Table 5.1).

The use of DIC resulted in selection of the quadratic model for the linear association in 17% of simulations when  $\sigma_U^2 = 0.25$  and 19% when  $\sigma_U^2 = 1$  with a validation study present (Table 5.1). However, when a replicate study was present, the use of DIC resulted in selection of the quadratic model in 31% of simulations where the underlying association was linear. Bayesian joint model estimation allows for “feedback” wherein the choice of the outcome model parameters effect the estimation of the measurement error and exposure model parameters [129]. In the case of a

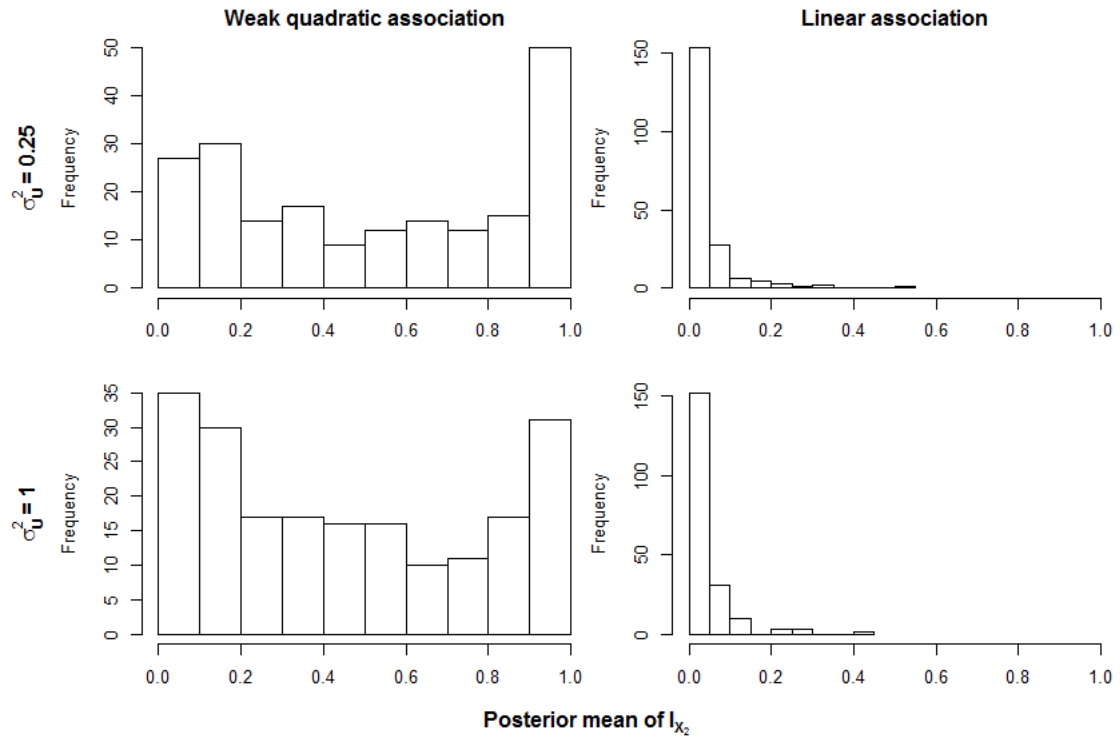


Figure 5.1 Histograms of the posterior mean of  $I_{X_2}$  estimated from 200 simulations with a continuous outcome, a validation study performed on 30% of 2000 study participants, and measurement error variance ( $\sigma_U^2$ ) either  $\frac{1}{4}$  the variance of  $X$  or equal to the variance of  $X$  after application of Bayesian transformation selection.

replicate study where the exposure was entirely latent, the estimation of  $X_i$  as a model parameter was changed depending on which substantive model was selected. This made attempts at model discrimination in this setting challenging.

### 5.5.8 Results: Model selection using BTS

#### *Fitting the quadratic model*

In all settings and for all three non-linear associations (Table 5.2), the BTS samples selecting the quadratic model had bias in the substantive model regression coefficients equal to or less than the bias observed for the standard Bayesian method assuming a quadratic model (“MCMC”) in Section 4.5.5 (Tables 4.2-4.4). Any reduction in bias and increase in precision over MCMC is likely due to the stronger prior on  $\beta_{X_2}$ . Coverage of the 95% CrIs for  $\beta_{X_1}$  and  $\beta_{X_2}$  was at or near the nominal level. The average MISE for BTS was equal to that of MCMC.

#### *Selection of the quadratic model*

When BTS was applied to the J-shaped or asymptotic associations, all samples select the quadratic model, i.e.  $\tilde{I}_{X_2} = 1$ , in all simulations indicating a very high probability that the quadratic term should be included in the model. This is in agreement with the selection criteria from other methods.

**Table 5.3 Bayes Factor (BF) evidence for the quadratic model using the Bayesian transformation selection (BTS) approach** given the mean posterior probability of the inclusion of a squared term in addition to a linear term ( $I_{X_2}$ ) in ten settings with a continuous outcome. Either a validation study was performed on 30% of 2000 study participants or a replicate study was performed on all 2000 participants. Measurement error variance ( $\sigma_U^2$ ) was either ¼ the variance of  $X$  or equal to the variance of  $X$ . The prior probability ( $\pi$ ) for the Bernoulli distribution for  $I_{X_2}$  was either 0.5 or 0.05. IQR is the interquartile range. In the setting with no measurement error (\*), the Bayesian analysis includes only the substantive model described for BTS; all other settings incorporate the exposure and measurement error models. 200 simulations performed for each setting.

Setting:	$\sigma_U^2$	Prior for $\pi$	Posterior mean of $I_{X_2}$ (IQR)	Median BF (IQR)	% of simulation for which BF $\geq x$		
					$x = 1$	$x = 3$	$x = 10$
<b>Weak quadratic association</b>							
No measurement error*	0	0.5	0.596 (0.29 – 0.92)	2.13 (0.40 – 11.9)	61	45	30
Validation	0.25	0.5	0.528 (0.16 – 0.90)	1.15 (0.19 – 8.6)	52	36	23
Validation	1	0.5	0.458 (0.14 – 0.78)	0.69 (0.17 – 3.5)	43	26	15
Replicate	1	0.5	0.392 (0.09 – 0.68)	0.47 (0.11 – 1.8)	34	19	10
Validation	1	0.05	0.150 (0.01 – 0.13)	0.68 (0.13 – 2.9)	42	25	15
<b>Linear association</b>							
No measurement error*	0	0.5	0.046 (0.02 – 0.05)	0.03 (0.02 – 0.05)	1	0	0
Validation	0.25	0.5	0.051 (0.03 – 0.05)	0.03 (0.03 – 0.05)	1	0	0
Validation	1	0.5	0.051 (0.03 – 0.05)	0.04 (0.03 – 0.05)	0	0	0
Replicate	1	0.5	0.059 (0.03 – 0.06)	0.04 (0.03 – 0.06)	0	0	0
Validation	1	0.05	0.003 (0.00 – 0.00)	0.04 (0.02 – 0.06)	1	0	0

When BTS was applied to the weak quadratic association when a validation study was present, the average posterior mean  $E[\tilde{I}_{X_2} | \mathbf{W}, X, Y; \boldsymbol{\theta}]$  was 0.53 and 0.46 for  $\sigma_U^2 = 0.25$  and 1, respectively, (Table 5.3). However, the histogram in Figure 5.1 highlights the wide spread of posterior mean values for  $I_{X_2}$ . The interquartile range (IQR) presented in Table 5.3 also highlights this spread.

The mean BF was sensitive to extreme values, therefore the median BF across simulations was more informative (Table 5.3). At a BF of 1, the linear and quadratic models were equally likely. For the weak quadratic association when no measurement error was present, the median Bayes Factor comparing the quadratic model to the linear model was 2.1, indicating weak evidence on average for the quadratic model. If only requiring that the quadratic model have greater evidence than the linear model, 61% of simulations without measurement error select the quadratic model. This was substantially less power to select the quadratic model than observed for the latent  $X$  frequentist analysis (87%; Table 5.1).

When the measurement error variance was ¼ the variance of  $X$  ( $\sigma_U^2 = 0.25$ ) and a validation study was present (Table 5.3), there was at least weak evidence (BF  $\geq 1$ ) for the quadratic model in 52% of simulations (compared to the naïve frequentist analysis with power of 59%; Table 5.1).

When measurement error variance is equal to the variance of  $X$ , there was at least weak evidence in favor of the quadratic model in 42% of simulations. By comparison, the naïve analysis selected the quadratic model in only 22% of simulations. In this same setting, when the prior for  $\pi$ , i.e. the prior probability of inclusion of the squared term, was reduced from 0.5 to 0.05 (indicated a strong prior belief that the model is linear), the average posterior mean dropped dramatically from 0.46 to 0.15. However, one can see in Table 5.3 that the BF was largely unaffected by the choice of prior (Equation 5.3).

The average posterior mean of  $I_{X_2}$  and the median BF were reduced when a replicate study was used in lieu of a validation study (Table 5.3). Power to select the quadratic model was improved in this scenario over that of the naïve analysis (34% versus 22% when considering weak evidence).

When the underlying association was linear, the posterior mean of  $I_{X_2}$  was near 0.05 with or without measurement error when the prior for  $\pi$  was 0.5 (Table 5.3). When the prior for  $\pi$  was 0.05, the posterior mean of  $I_{X_2}$  was 0.003; however, the BF was similar in either case (though slightly greater variation was observed). None of the simulations had even weak evidence for the quadratic model for this association shape.

Mixing, i.e. the change of  $\tilde{I}_{X_2} = 0$  to  $\tilde{I}_{X_2} = 1$ , was subjectively observed to be poor in this setting. This may lead to underestimation of the probability of the quadratic model.

## **5.6 SIMULATION STUDY EXTENSION TO MODEL SELECTION FOR A BINARY OUTCOME**

### **5.6.1 Simulation study design and evaluation criteria**

The simulation study for logistic regression in Chapter 4 (Section 4.6) was extended to include a model with a linear functional form of the error-prone exposure. The two quadratic associations from Table 4.5 and this “linear” association were simulated again in this section to assess model selection between the linear (Equation 1.1) and quadratic (Equation 4.1) substantive models when the outcome was a binary variable. The binary outcome,  $Y_i$ , was generated as a Bernuolli random variable with probability equal to the logit transformation of the linear predictor. The value of each  $\beta_0$  was chosen to result in an event rate for the outcome  $Y$  of 20%. As in the previous chapter, only one setting was considered.

The criteria for evaluating model selection can be found in Section 5.5.2.

### 5.6.2 Method implementation

All methods were implemented exactly as they were in Section 5.5.3 except for the following changes.

For the implementation of the fully Bayesian MCMC, there was no scaling of the binary outcome. Three chains were run for each simulation with a burn-in period of 300,000 samples with the last 20,000 burn-in samples evaluated for convergence using  $\hat{R}$ . Another 20,000 samples were collected from each chain for inference. ESS was estimated from these samples.

#### Box 5.3 Simulation study design

Simulations per scenario: 100

True exposure:  $X_i \sim N(10, 1)$

Classical error model:

$$W_{ij} = X_i + U_{ij}, \quad U_{ij} \sim N(0, \sigma_U^2), \quad j = 1, 2$$

Outcome models:

J-shaped association,  $N = 2000$ :

$$P(Y_i = 1|X_i) \sim \text{Bern}(-1.8 + 0.2(X - 9)^2)$$

Weak quadratic association,  $N = 10000$ :

$$P(Y_i = 1|X_i) \sim \text{Bern}(-4.5 + 0.03 * X^2)$$

Linear association,  $N = 10000$ :

$$P(Y_i = 1|X_i) \sim \text{Bern}(-7.66 + 0.62 * X)$$

Setting: Validation study in 30% of participants,  $\sigma_U^2 = 0.25$

The BTS method was implemented in MCMC as above, but with six chains for each simulation and 10,000 samples collected for inference from each chain.

### 5.6.3 Model selection using the latent $X$

The mean p-value associated with the likelihood ratio test comparing the linear and quadratic models for the J-shaped association was  $2 \times 10^{-4}$ . While this was a much higher p-value than that observed for the J-shaped association with a continuous outcome, it was still a strong indicator that the quadratic model was the best fit. For all simulations with the J-shaped association, when the latent  $X$  was fit to each model, the quadratic model was selected using either the 95% CI for  $\beta_{X_2}$  or AIC.

The mean p-value associated with the likelihood ratio test comparing models for the weak quadratic association was 0.28. Hypothesis testing, i.e. observing whether the 95% CI for  $\beta_{X_2}$  excludes zero, resulted in the selection of the quadratic model in only 29% of simulations and use of AIC after fitting both the linear and quadratic models resulted in the selection of the quadratic model in only 49% of simulations (Table 5.4).

When simulated data generated from the linear association was fit to both the quadratic and linear models, the p-value associated with the likelihood ratio test was 0.48.

**Table 5.4 Selection of the quadratic model with a weak quadratic or linear association with a binary outcome: 200 simulations performed to find the power and type I error rate in selection of the quadratic when the shape was either a weak quadratic or a linear association as determined by the likelihood ratio test (at  $\alpha$ -level 5%, 10%, or 20%), the exclusion of the zero from the 95% confidence interval or credible interval (CI), or the Akaike information criteria (AIC) or alternatively the deviance information criteria (DIC; for the standard MCMC method only). Applied for a validation study performed on 30% of 2000 study participants with measurement error variance  $1/4$  the variance of  $X$ .**

	Hypothesis testing $H_0: \beta_{X_2} = 0$			AIC/DIC	Hypothesis testing $H_0: \beta_{X_2} = 0$			AIC/DIC
	$\alpha=5\%$	$\alpha=10\%$	$\alpha=20\%$		$\alpha=5\%$	$\alpha=10\%$	$\alpha=20\%$	
	Power to select quadratic fit for a weak quadratic association				Type I error rate in selecting the quadratic fit given a linear association			
<b>Latent X</b>	29%	41%	52%	49%	5%	9%	21%	17%
<b>Naïve</b>	12%	20%	42%	39%	7%	12%	20%	17%
<b>RC</b>	15%	23%	43%	39%	6%	12%	20%	17%
<b>MCMC</b>	20%	35%	46%	47%	6%	9%	23%	36%
<b>MCMC-RC</b>	16%	34%	44%	41%	6%	10%	22%	12%
<b>INLA-RC</b>	16%	34%	45%	41%	5%	10%	21%	11%
	Power to select quadratic fit for a J-shaped association							
<b>Latent X</b>	100%	100%	100%	100%				
<b>Naïve</b>	98%	98%	98%	98%				
<b>RC</b>	97%	98%	98%	98%				
<b>MCMC</b>	98%	98%	99%	96%				
<b>MCMC-RC</b>	98%	98%	99%	99%				
<b>INLA-RC</b>	98%	98%	99%	99%				

#### 5.6.4 Model selection using the naïve analysis

After the addition of  $\sigma_U^2 = 0.25$  to the exposure, in the naïve analysis the 95% CI for  $\beta_{X_2}$  excluded zero for the J-shaped association in 98% of simulations (Table 5.4). The naïve AIC similarly selected the quadratic model over the linear model in 98% of simulations for the J-shaped association.

For the weak quadratic association, in the naïve analysis the 95% CI for  $\beta_{X_2}$  excluded zero in only 12% of simulations (Table 5.4). Use of AIC resulted in a selection of the quadratic model in 39% of simulations.

Empirically, slightly elevated type I error rates were observed for the linear association, but this was likely due to random variability given the relatively small number of simulations.

#### 5.6.5 Results: Model selection using hypothesis testing

For the J-shaped association, nearly all methods resulted in the selection of the quadratic model in 97-98% of simulations using either the 95% CI/CrI.

For the weak quadratic association, assuming an  $\alpha$ -level of 5%, RC selected the quadratic model in 3% more simulations than the naïve analysis on the basis of hypothesis testing (Table 5.4). MCMC-RC and INLA-RC selected the quadratic model in 4% more simulations than the naïve

*Table 5.5 Quadratic model fit with a binary outcome from Bayesian transformation selection (BTS) samples including mean bias in the quadratic model regression estimates, coverage of the 95% confidence intervals or credible intervals and the associated Monte Carlo error 95% confidence interval (MCE 95% CI), average mean integrated square error (MISE) over the exposure range 7-13, and bias in the turning point of the mean curve (where applicable) estimated from 200 simulations for each setting and association (J-shaped, asymptotic, and weak quadratic). Assumes a validation study performed on 30% of 2000 study participants with a measurement error variance 1/4 the variance of X. The prior probability ( $\pi$ ) for the Bernoulli distribution for  $I_{x_2}$  is 0.5. \*indicates a MCE 95% CI for bias that excludes zero. TP indicates turning point. † indicates 1 simulation wouldn't run for that method.*

Association shape:	Bias			Coverage (MCE 95% CI)		Average MISE	TP bias
	$\beta_0$	$\beta_{x_1}$	$\beta_{x_2}$	$\beta_{x_1}$	$\beta_{x_2}$		
<b>J-shaped, N=2000</b>							
True values	14.4	-3.6	0.2				9.0
BTS†	-0.411	0.078	-0.0038	92% (87 – 97)	92% (87 – 97)	0.044	0
<b>Weak quadratic, N=10000</b>							
True values	-4.5	0	0.03				
BTS	-1.452*	0.278*	-0.0133*	27% (18 – 36)	28% (19 – 37)	0.015	

analysis. MCMC selected the quadratic model on the basis of the 95% CrI exclusion of zero in 8% more simulations than the naïve analysis.

Type I error rates for selection of the quadratic model when the underlying association was linear were similar to the naïve analysis for all methods.

### 5.6.6 Results: Model selection using AIC/DIC

The use of the AIC to select the quadratic model over the linear model when the association was J-shaped resulted in the selection of the quadratic model in 98% of simulations for RC (same as the naïve analysis) and 99% for MCMC-RC and INLA-RC (Table 5.4). Use of DIC to select the best MCMC model resulted in selection of the quadratic model in only 96% of simulations when the association was J-shaped (3% less than the naïve analysis).

For the weak quadratic association, use of the DIC with MCMC resulted in selection of the quadratic model in 47% of simulations (8% more than the naïve analysis). Use of the AIC after correction with RC selected the quadratic model in only 39% (same as the naïve analysis). MCMC-RC and INLA-RC selected the quadratic model using AIC in 41% of simulations.

RC, MCMC-RC, and INLA-RC had a type I error rate when using AIC between that observed for the naïve analysis and the latent X. MCMC, using the DIC, selected the quadratic model for the linear association at more than twice the rate of the naïve analysis, i.e. 36% vs 17%. This further exemplifies the phenomenon of Bayesian feedback (Section 5.5.7) when the likelihood contributes too little information to the posterior [129].

### 5.6.7 Results: Model selection using BTS

Using only the samples from the quadratic model, for the J-shaped association, the curve fits resulting from the implementation of BTS were very similar to those from the Bayesian analysis with MCMC wherein a quadratic substantive model was specified (Table 5.5). For the weak

**Table 5.6 Bayes Factor (BF) evidence for the quadratic model with a binary outcome using the Bayesian transformation selection (BTS) approach** given the mean posterior probability of the inclusion of a squared term in addition to a linear term ( $I_{X_2}$ ) in ten settings with a continuous outcome. Either a validation study was performed on 30% of 2000 study participants or a replicate study was performed on all 2000 participants. Measurement error variance ( $\sigma_U^2$ ) was either  $\frac{1}{4}$  the variance of  $X$  or equal to the variance of  $X$ . The prior probability ( $\pi$ ) for the Bernoulli distribution for  $I_{X_2}$  was either 0.5 or 0.05. IQR was the interquartile range. In the setting with no measurement error (\*), the Bayesian analysis included only the substantive model described for this method; all other settings incorporated the exposure and measurement error models. 200 simulations performed for each setting. † indicates 1 simulation wouldn't run for that method. ‡ indicates a value greater than 100 or less than 0.01.

Association shape:	Prior for $\pi$	Posterior mean of $I_{X_2}$ (IQR)	Median BF (IQR)	% of simulation for which $BF \geq x$		
				$x = 1$	$x = 3$	$x = 10$
J-shaped†	0.5	0.866 (0.83 – 1.00)	61.8 (5.0 – 100‡)	94	81	51
Weak quadratic	0.5	0.127 (0.01 – 0.17)	0.07 (0.01‡ – 0.21)	5	0	0
Linear	0.5	0.061 (0.01 – 0.08)	0.01 (0.01‡ – 0.09)	0	0	0

quadratic association, because there were far fewer samples where  $I_{X_2} = 1$ , the ESS for  $\hat{\beta}_{X_2}$  was very small in most simulations (median 17.7; IQR 9.6 – 37.0). Consequently, the quadratic curve fit from these limited samples was very poor.

From Table 5.6, it can be seen that the J-shaped association had the highest posterior probability that the quadratic model was the most appropriate model, i.e. the mean  $E[\tilde{I}_{X_2} | \mathbf{W}, X, Y, \boldsymbol{\theta}] = 0.65$ . By comparison, the weak quadratic association had a mean posterior probability for the quadratic model of only 0.29. The linear association had a much higher mean posterior mean of  $I_{X_2}$  for logistic regression than that observed for linear regression; this is likely due to the fact that the likelihood had less information in this setting and the prior information held more weight.

Using the BF, there was at least weak evidence for the quadratic model in 35% of simulations with the J-shaped association and 8% with the weak quadratic association. For both association shapes, this was a dramatic reduction from the naïve analysis.

## 5.7 SUMMARY

In this chapter, I introduced hybrid Bayesian/RC methods using either MCMC or INLA for the Bayesian posterior estimates. These methods reliably improved the power to select the quadratic model over the naïve analysis or standard RC while maintaining type I error rates at or below the nominal level. The hybrid methods were also much faster than fully Bayesian methods and, as will be seen in the next chapter, could easily accommodate more complex model selection procedures.

I also introduced the use of a method of BTS in order to determine the probability that a squared term should be included in the model. This model requires a Bayesian framework to interpret, but can inform the user as to the specific probability of each model.



All methods were observed to have type I error rates at or below the nominal level when making use of hypothesis testing for model selection (within the bounds of empirical variability). For the Bayesian methods (MCMC, MCMC-RC, and INLA-RC), whether the type I error is at or below the nominal level will depend on appropriate specification of the prior distributions. Misspecification of the prior or the model may lead to increased type I error and loss of power. AIC is likewise affected for MCMC-RC and INLA-RC but likely maintains the relative type I error rates as would be observed for hypothesis testing at a nominal  $\alpha$ -level of 16%.

RC recovered no power lost to the effects of classical measurement error for the model demonstrated in this section. In the logistic regression simulation study, RC appeared to recover a small proportion of the power (Table 5.4), but that was due to bias remaining in the regression coefficients (Table 4.6). If additional fully measured covariates related to the error-prone exposure were available or other sources of information were integrated into the measurement error model, power gains would more likely be observed [130,131]. However, the reduction in bias using correction by RC comes at a cost of greater variability (referred to as the “bias versus variance” tradeoff in Carroll et al [1]) and thus no power is gained.

As implemented in this chapter, the hybrid methods and MCMC had improved power to select the quadratic model for a weak quadratic association when hypothesis testing or AIC was used.

DIC was unreliable: sometimes resulting in power gains and others in loss of power significantly below that of the naïve analysis. Use of DIC also resulted in greatly inflated type I error when used in combination with a replicate study as a result of Bayesian feedback wherein the choice of parameters or model specification for the substantive model impacts the choice of parameters for the measurement error or exposure models [129]. This feedback resulted in overselection of the more complex model (i.e. high type I error).

When looking exclusively at samples for which  $I_{X_2} = 1$ , BTS was observed to result in the same reduction of bias and good coverage properties observed in Chapter 4 for a Bayesian model with the quadratic model specified. The use of BTS resulted in improved power to select the quadratic model for the higher measurement error variance (i.e. when the model was correctly specified). Bayesian feedback was not observed in this model for linear regression with a replicate study or for logistic regression.

The performance of the measurement error correction methods when the substantive model was logistic regression was similar to when the substantive model was linear regression. The most important caveat being that the likelihood will have less information for a binary outcome with the same effect size (or strength of association) and sample size.

Lessons learned about model selection in the context of these measurement error correction methods will be applied in the next chapter to fractional polynomial model selection.

## 6 MEASUREMENT ERROR CORRECTION FOR POLYNOMIAL MODEL SELECTION USING THE FRACTIONAL POLYNOMIAL METHOD

---

### 6.1 AIMS & OVERVIEW

Often model-building requires consideration of an array of potential transformations of the error-prone exposure; the fractional polynomial method (or “fractional polynomials”) is a systematic method of selection of the best polynomial transformation of an exposure (Section 1.3.3). In the previous chapter, it was shown that classical measurement error in the exposure leads to obscuring of the shape of the association leading to a reduction in the power to detect any non-linearity [1–3]. Accurate estimation of the exposure-outcome association at all relevant exposure levels or even detecting the presence of any non-linearity in the relationship can therefore be quite challenging, as demonstrated in simulation studies in Chapter 5.

Previously, Strawbridge described how to use RC in the context of fractional polynomials wherein the maximum likelihood estimates for the necessary expectations were derived under specific distributional assumptions [49]. Keogh, Strawbridge, and White performed simulation studies across a range of association shapes relating an error-prone exposure to survival time, using the Cox proportional hazards model [23]. They demonstrated the loss of power to detect non-linearity when the fractional polynomial method was applied to the naïve association even for a relatively small measurement error (e.g. attenuation factor = 0.8).

In this chapter, the aim was to extend the measurement error correction methods presented in Chapter 5, where the corrected linear and quadratic models (i.e. nested models) were compared, to a more systematic selection of the best transformation of the error-prone exposure (i.e. fractional polynomials). A simulation study similar to that performed in Chapter 5 was performed to evaluate the remaining bias in the corrected association curve fits and as well as the power to detect any linearity. RC was performed using the method from Strawbridge [49]; though, an important drawback of this is that it relies on an assumption that the true exposure conditional on the error-prone measures is log-normally distributed. The novel hybrid methods (MCMC-RC and INLA-RC), by design, are extended easily to this new context. Finally, I present a fully Bayesian method based on Bayesian variable selection methods applied to the fractional polynomial context with exposure measurement error for the first time.

## 6.2 FRACTIONAL POLYNOMIAL METHOD AND SHAPES OF ASSOCIATION

The fractional polynomial method stems from a generalized additive model (Equation 1.2) where the transformation of the error-prone predictor (i.e.  $f(X_i)$ ) is restricted to the standard transformations based on the Box-Tidwell set [132]. See Section 1.3.3 for greater detail.

For a first degree (FP1) model ( $m = 1$ ) where  $f(X_i)$  represents a single transformation of  $X_i$ , the possible transformations are in the set  $h(X_i, p) = \left\{ X_i^{-2}, X_i^{-1}, \frac{1}{\sqrt{X_i}}, \log(X_i), \sqrt{X_i}, X_i, X_i^2, X_i^3 \right\}$  represented by the parameter  $p \in S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$  (Equation 1.5). This set includes the linear model (i.e.  $p = 1$ ). For a second degree (FP2) model ( $m = 2$ ), wherein  $f(X_i)$  represents two transformations of  $X_i$ , any two transformations in the set of  $S$  may be chosen as  $p_1$  and  $p_2$  (i.e.  $\mathbf{p} = \{p_1, p_2\}$ ). When  $p_1 = p_2$ , the first transformation is  $X_i^{p_1}$  and the second  $X_i^{p_1} \log(X_i)$ , i.e.  $h(X_i, p_1) \log(X_i)$  (Equation 1.6). An FP1 model has a single regression coefficient for the exposure of interest,  $\beta_X$ , while an FP2 model has two,  $\beta_{X_1}$  and  $\beta_{X_2}$ , and the vector of regression coefficients for transformations of  $X_i$  is denoted  $\boldsymbol{\beta}_X$ . The quadratic model from Chapter 4 (Equation 4.1) can be represented as an FP2 model with  $\mathbf{p} = \{1, 2\}$ .

Typically, in applying the fractional polynomial method, the maximum desired degree is chosen *a priori* (i.e.  $m \leq 1$  or  $m \leq 2$ ). As discussed in Section 1.3.3, after fitting all models, the best FP1 and the best FP2 models are each selected as the model with the lowest deviance for that degree. The linear model is treated as a special case of FP1 models to be considered separately. The FP2 model is chosen (i.e. as the “chosen model”) if the deviance difference between that and the best FP1 model, the linear model, and null model are found to be significant (i.e. a p-value of  $< 0.05$  when compared to the  $\chi^2$ -distribution at an  $\alpha$ -level of 5%). Otherwise, the FP1 model is chosen if the deviance difference between that and the linear and null models are found to be significant. Greater detail is provided in Section 1.3.3.

In the fractional polynomial method, the SEs for the regression parameters or predicted outcome values estimated from the chosen model do not incorporate uncertainty due to model selection. However, it would be possible to obtain SEs which do include uncertainty due to model selection through bootstrapping.

A potential alternative to the comparison of the deviance for each model is the use of the AIC to compare models (Section 5.3) [8]. In this case, all models (null, linear, FP1, and FP2) would be fit and the model with the lowest AIC would be accepted as the final model. The use of AIC as compared to the selection procedure described above and in Section 1.3.3 (and other selection methods) is discussed by Royston and Sauerbrei in detail [8]. For brevity, in this chapter only the traditional fractional polynomial selection method is considered.

The three non-linear shapes of association from the previous chapter (Table 4.1) were employed in the Section 6.5 with a continuous outcome (i.e. using an identity link for  $g(\cdot)$  within a generalized additive model) to demonstrate the measurement error correction methods in the context of the fractional polynomial selection method in a simulation study. A linear association was also used to assess whether the type I error rate was at or below the nominal level; as discussed in Chapter 5, type I error rates are not expected to maintain exact frequentist probabilities for the Bayesian and hybrid methods. The weak quadratic association exemplifies a first degree polynomial while the J-shaped association exemplifies a second degree polynomial. The asymptotic association shape does not fit any polynomial function perfectly as is likely to be true most often for real data. In Section 6.6, the measurement error correction methods were also demonstrated in the context of the fractional polynomial method for logistic regression, i.e. a logit link.

### 6.3 NAÏVE ANALYSIS AND THE EFFECT OF MEASUREMENT ERROR

As in previous chapters, the naïve analysis uses a single error-prone measure  $W_1$  instead of the latent exposure  $X_i$  within the substantive model:

$$6.1 \quad g(E[Y_i|W_{1i}, \mathbf{Z}_i]) = \beta_0^* + f(W_{1i}) + \boldsymbol{\beta}_Z^{T*} \mathbf{Z}_i \quad i = 1, 2, \dots, n,$$

where  $f(W_{1i})$  is selected based on the fractional polynomial method (Equations 1.5 and 1.6) and is parameterized by  $\beta_W$  if  $m = 1$  or  $\beta_{W_1}$  and  $\beta_{W_2}$  if  $m = 2$ . Unlike the effect when a specific model is assumed, these coefficients are not necessarily simply biased estimates of  $\boldsymbol{\beta}_X$  because the transformation of  $W_{1i}$  in the chosen fractional polynomial model in the naïve analysis may be different, and even of a different degree, to the chosen model in the latent  $X$  analysis.

Assuming the classical error model (Equation 1.3;  $W = X + U$ ), the overdispersion of  $W$  results in an apparent association between the error-prone measure and the outcome more linear than the true association between the latent exposure and the outcome. The effect of the exposure is therefore overestimated for some exposure groups and underestimated for others (Figure 4.1). The specific effect of the addition of measurement error on the association shapes considered in this chapter were explored in Chapter 4. In the context of the fractional polynomial method, the reduced power to detect a non-linear model over a linear model may further exacerbate the flattening effect [8].

Mathematical expressions describing the effect of measurement error on a non-linear curve become increasingly cumbersome beyond those already described for a quadratic model in Chapter 4 and those described for a linear model in Chapter 2. Additionally, once derived, they would remain specific to an assumed model.

When the measurement error is more complex than the classical measurement error model, the resulting effect on the overall shape of the naïve association may be less predictable. The lognormal error model is described in the next section and is considered in a sensitivity analysis in Section 6.5. The application in Chapter 8 includes the use of the EC model.

## 6.4 CORRECTION METHODS IN THE CONTEXT OF FRACTIONAL POLYNOMIALS

### 6.4.1 Regression calibration (RC)

Strawbridge derived a method of RC for the fractional polynomial method by assuming that  $X|W, \mathbf{Z}$  is log-normally distributed [49]. A log-normal distribution was specified in lieu of a normal distribution because the required expectations cannot be derived analytically for negative values of  $p$  or for  $p = 0.5$  (e.g.  $X^{-2}$ ). This assumption cannot work for exposures with negative values, but the fractional polynomial method is likewise limited to exposures that take only positive values (Section 1.3.3). In order to make the assumption, a lognormal error model, extended from the classical error model wherein the error  $U'$  is added on the log scale, is defined as follows:

$$6.2 \quad \log(W_i) = \log(X_i) + U', \quad U' \sim N(0, \sigma_{U'}^2).$$

This error model is sometimes called multiplicative error as it can be re-expressed as  $W_i = X_i e^{U'}$  [5]. The attenuation factor for the lognormal error model is constructed by relating the error variance to the variance of the  $\log(X)$  conditional on any accurately measured covariates ( $\sigma_{\log(X)|\mathbf{Z}}^2$ ),  $\lambda' = \sigma_{\log(X)|\mathbf{Z}}^2 / (\sigma_{\log(X)|\mathbf{Z}}^2 + \sigma_{U'}^2)$ . It can then be shown that the distribution of  $X|W, \mathbf{Z}$  is lognormal:

$$6.3 \quad \log(X) | W, \mathbf{Z} \sim N(\lambda' \log(W_i) + (1 - \lambda')\mu', (1 - \lambda')\sigma_{\log(X)|\mathbf{Z}}^2),$$

where  $\mu'$  is the mean of  $\log(X)$  given  $\mathbf{Z}$ .

Using this distributional assumption, Strawbridge derived the expectations for each possible transformation of  $X$  for FP1 and FP2 [49]. Each expectation is obtained given a specific value for the parameter  $p$  (or for FP2,  $p_1$  or  $p_2$ ) and the error-prone measure,  $W_i$ :

$$6.4 \quad E[X_i^p | W_i, \mathbf{Z}_i] = k(p)W_i^{\lambda' p}$$

$$E[\log(X_i) | W_i, \mathbf{Z}_i] = \log(W_i^{\lambda'}) + (1 - \lambda')\mu'$$

$$E[X_i^p \log(X_i) | W_i, \mathbf{Z}_i] = k(p) \left[ W_i^{\lambda' p} \log(W_i^{\lambda'}) + (1 - \lambda')(\mu' + p\sigma_{\log(X)|\mathbf{Z}}^2)W_i^{\lambda' p} \right]$$

$$E[(\log(X))^2 | \mathbf{W}_i, \mathbf{Z}_i] = (1 - \lambda') [\sigma_{\log(X)|\mathbf{Z}}^2 + (1 - \lambda')\mu'] + \log(W_i^{\lambda'}) [\log(W_i^{\lambda'}) + 2(1 - \lambda')\mu']$$

$$k(p) = \exp\left((1 - \lambda') \left(\frac{\sigma_{\log(X)|\mathbf{Z}}^2 p^2}{2} + p\mu'\right)\right)$$

If a validation study has been performed, then both  $\mu'$  and  $\sigma_{\log(X)|\mathbf{Z}}^2$  can be estimated using the observed values of  $X$ , and  $\sigma_{U'}^2$  can be estimated as the difference between the estimated variance of  $\log(W)|\mathbf{Z}$  and  $\sigma_{\log(X)|\mathbf{Z}}^2$ . If only a replicate study is available, then  $\mu'$  can still be estimated from the observed measures (i.e.  $E[\log(X) | \mathbf{Z}] = E[\log(W) | \mathbf{Z}]$ ) and  $\sigma_{\log(X)|\mathbf{Z}}^2$  can be estimated from the covariance of the observed measures. For a replicate study,  $\sigma_{U'}^2$  is estimated as half the variance of the difference between  $\log(W_1)$  and  $\log(W_2)$ .

To apply the fractional polynomial method with RC-corrected expectations, each FP1 and FP2 model is fit by replacing the transformation of  $X$  with the appropriate expectation of that transformation as follows (i.e. replacing Equations 1.5 and 1.6 in the naïve method):

$$6.5 \quad f(X_i)_{\text{RC1}} = \beta_X E[h(X_i, p) | \mathbf{W}_i, \mathbf{Z}_i],$$

$$\text{where } E[h(X_i, p) | \mathbf{W}_i, \mathbf{Z}_i] = \begin{cases} E[X_i^p | \mathbf{W}_i, \mathbf{Z}_i], & p \neq 0 \\ E[\log(X_i) | \mathbf{W}_i, \mathbf{Z}_i], & p = 0 \end{cases},$$

$$6.6 \quad f(X_i)_{\text{RC2}} = \beta_{X_1} E[h(X_i, p_1) | \mathbf{W}_i, \mathbf{Z}_i] + \beta_{X_2} E[h'(X_i, p_1, p_2) | \mathbf{W}_i, \mathbf{Z}_i], \quad p_1 \leq p_2,$$

$$\text{where } E[h'(X_i, p_1, p_2) | \mathbf{W}_i, \mathbf{Z}_i] = \begin{cases} E[X_i^{p_2} | \mathbf{W}_i, \mathbf{Z}_i], & p_1 \neq p_2 \\ E[(\log(X_i))^2 | \mathbf{W}_i, \mathbf{Z}_i], & p_1 = p_2 \end{cases}.$$

After fitting each model using the estimated expectations, model selection is performed as detailed in Section 6.2.

The SEs for the predicted outcome values from the RC-corrected chosen model can be estimated by bootstrapping [75]. Because RC remains a relatively fast method in combination with fractional polynomials, the model-selection process may also be incorporated by bootstrapping. This was done in a sensitivity analysis (Appendix H).

#### 6.4.2 Bayesian regression calibration

In Bayesian regression calibration, posterior samples of  $f(X_i | \mathbf{W}_i, \mathbf{Z}_i; \boldsymbol{\alpha}, \boldsymbol{\pi})$  are drawn, where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\pi}$  represent the parameters of the exposure model and the error model, respectively (Section 2.3; Section 4.4.4). These posterior samples, denoted  $\tilde{X}_i$ , may be drawn from the stationary

distribution after an MCMC algorithm has converged (MCMC-RC) or generated from the distribution estimated by the INLA method (INLA-RC) (Section 4.4.4). The implementation of the first part of the method (the Bayesian part) does not differ in any way from the previous implementation in Chapters 4 and 5 (i.e. in the context of linear and quadratic models) because the substantive model is not incorporated into the sampling.

The posterior draws  $\tilde{X}_i$  can be transformed using  $h(\tilde{X}_i, p)$  and  $h'(\tilde{X}_i, p_1, p_2)$  (Equations 1.5 and 1.6). The expectations  $E[h(X_i, p)|\mathbf{W}_i, \mathbf{Z}_i]$  are then estimated using the mean over many posterior draws  $\tilde{X}_{ik}$  ( $k = 1, \dots, K$ ) where  $\tilde{X}_{ik}$  denotes the  $k$ th posterior draw for individual  $i$ :

$$6.7 \quad \hat{E}[h(\tilde{X}_i, p)|\mathbf{W}_i, \mathbf{Z}_i] = (1/K) \sum_{k=1}^K h(\tilde{X}_{ik}, p).$$

Likewise,  $\hat{E}[h'(X_i, p_1, p_2)|\mathbf{W}_i, \mathbf{Z}_i]$  is estimated from the mean of all  $h'(\tilde{X}_{ik}, p_1, p_2)$ . Having obtained  $\hat{E}[h(\tilde{X}_i, p)|\mathbf{W}_i, \mathbf{Z}_i]$  and  $\hat{E}[h'(X_i, p_1, p_2)|\mathbf{W}_i, \mathbf{Z}_i]$  for all powers  $p$  in the set of  $\mathbf{S}$ , all FP1 and FP2 models may be fit by replacing the appropriate expectations in Equations 6.5 and 6.6. Once the models have been fit, the fractional polynomial model selection procedure can be implemented as described in Section 6.2.

As stated in the previous chapter, the SEs of the predicted outcome values from the chosen model will not incorporate uncertainty due to measurement error in the estimation of the expectations of  $h(\tilde{X}_i, p)$  and  $h'(\tilde{X}_i, p_1, p_2)$ . Furthermore, in employing MCMC-RC with the fractional polynomial method, uncertainty due to model selection is also not included in the SEs. However, for single applications, it may be possible to use bootstrapping with INLA-RC to estimate the SEs thereby incorporating uncertainty due to both model selection and measurement error.

### 6.4.3 Bayesian transformation selection (BTS)

The relatively simple spike and slab prior used for this approach when applied to the selection of the quadratic model versus a linear model in Section 5.4.2 was observed to have poor mixing. Moreover, Kuo and Mallick demonstrated that the method cannot not distinguish between highly correlated variables such as is the case for the transformations of the error-prone exposure [60]. Stochastic search variable selection (SSVS), introduced by George and McCulloch in 1993, is another method of Bayesian variable selection which uses a variation of the spike and slab prior [124]. SSVS has previously been shown [120] to have better mixing properties than the Kuo and Mallick method when a fixed value is specified for the hyperparameter  $\pi$ . Improved mixing results in a better representation of the true probability of each model from the MCMC sampling. In this section, SSVS is adapted for BTS with measurement error correction.



### ***BTS for an FPI model***

The substantive model is a generalized linear model which includes all possible transformations of the error-prone exposure. SSVS stipulates that the prior distribution for  $\beta_{X_k}$  is a Gaussian mixture distribution conditional on  $I_{X_k}$ . The prior specified for  $I_{X_k}$  is still a Bernoulli distribution with probability  $\pi$ . If  $I_{X_k} = 0$ , the prior for  $\beta_{X_k}$  takes an extremely narrow (“spike”) density distributed around zero; if  $I_{X_k} = 1$ , the prior for  $\beta_{X_k}$  takes the more typical Gaussian density for a regression coefficient. The result is that, instead of a probability mass spike directly over zero, the spike is an extremely narrow density distribution around zero. The substantive model and the associated priors are then:

$$\begin{aligned} 6.8 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) &= \beta_0 + \beta_{X_1}X_i^{-2} + \beta_{X_2}X_i^{-1} + \dots + \beta_{X_7}X_i^2 + \beta_{X_8}X_i^3 + \boldsymbol{\beta}_Z^T \mathbf{Z}_i \\ &= \beta_0 + \boldsymbol{\beta}_X \mathbf{h}(X_i, \mathbf{p}) + \boldsymbol{\beta}_Z^T \mathbf{Z}_i \\ \beta_{X_k} | I_{X_k} &\sim (1 - I_{X_k})N(0, \sigma_a^2) + I_{X_k}N(0, \sigma_b^2), \\ I_{X_k} &\sim \text{Bern}(\pi). \end{aligned}$$

The variance  $\sigma_b^2$  defines the Gaussian distribution when the variable (or transformation of  $X$ ) should be included in the model, whereas the variance  $\sigma_a^2$  defines the spike distribution when the regression coefficient is virtually zero. A value of 3,600 was proposed by O’hara and Sillanpää for  $\sigma_a^2$  [120]; however, increasing this variance increases the width of the “spike” and therefore the probability that  $I_{X_k}$  will change from 1 to 0. The best value of this tuning parameter may be specific to the data.

The model in Equation 6.8 allows for the possibility of several transformations of  $X$ , if more than one of the  $I_{X_k}$  terms is equal to 1. Because a more complex model including more transformations of the exposure has greater flexibility to fit the data, the prior distribution for  $\mathbf{I}_X$  must be specified to induce sparsity in order to select more parsimonious models.

As discussed previously in Section 4.4.2, appropriate scaling of the variable transformations is essential for correct prior specification. Scaling can ensure that the variance of each transformation is near 1. For this implementation of the BTS, scaling was also important for appropriate mixing (i.e. change of  $I_{X_k}$  between 0 and 1). The value for  $\sigma_b^2$ , the variance of the prior for  $\beta_{X_k}$  when in the “slab”, must be small enough so that values of  $\beta_{X_k}$  sampled while  $I_{X_k}$  was 0 does not prevent the move to 1. For example, a variance of 3 will ensure nearly all prior values of  $\beta_{X_k}$  are drawn from the range -10 to 10; a reasonable range for a logOR and for the regression coefficient with a continuous outcome if that outcome is also scaled.

**Table 6.1** Example of a results table summarizing the frequency of each model sampled using Bayesian transformation selection (BTS).

Model	$I_{X_1}$	$I_{X_2}$	$I_{X_3}$	$I_{X_4}$	$I_{X_5}$	$I_{X_6}$	$I_{X_7}$	$I_{X_8}$	$p$	Freq.	Transformations of $X$
M1	0	0	0	0	0	0	1	0	2	0.35	$X^2$
M2 or L	0	0	0	0	0	1	0	0	1	0.34	$X$
M3	0	0	0	0	1	0	0	0	0.5	0.21	$X^{0.5}$
M4	0	0	0	0	1	0	0	1	{0.5, 3}	0.05	$X^{0.5} + X^3$
M5	0	0	0	1	0	0	1	1	{0, 2, 3}	0.05	$\log(X) + X^2 + X^3$

The collinearity between the different transformations of  $X$  in Equation 6.8 is extremely high posing a substantial challenge to discerning a parsimonious model. Specifically, I found that the transformations  $X^{-2}$  and  $X^3$  have a tendency to be over selected because they have the least correlation with each other and with the other transformations. Centering of independent variables can be used to reduce apparent collinearity between variables in Bayesian regression and is therefore essential in this context [55]. Both centering and scaling are performed after transformation of the variable. When the variable is entirely latent, as is the case when only a replicate study is available, the appropriate scaling and centering factors can be approximated based on the observed measures. In the implementation in this thesis, the mean and variance of  $W_{i2}$  was used to perform the scaling and centering.

As in Section 5.4, the BF can be used to assess whether the model found to have the highest posterior probability, denoted  $\hat{I}_{M1}$ , has sufficient evidence (Box 5.1) to be accepted over the linear model (i.e.  $I_L = \{0,0,0,0,0,1,0,0\}$ ) [127]. Additionally, the BF can be used to assess whether  $\hat{I}_{M1}$  has significantly more evidence in its favor than alternative models with the second-, third-, or fourth-highest posterior probability, i.e.  $\hat{I}_{M2}, \hat{I}_{M3}, \hat{I}_{M4}$ .

The MCMC samples of the BTS method are best represented in a frequency table such as that exemplified in Table 6.1 [60,120]. This table demonstrates what sampling frequencies might resemble when the best FP1 model is either  $p = 2$  (a model with only a squared term;  $\hat{I}_{M1}$ ) or  $p = 1$  (the linear model;  $\hat{I}_L$ ) [60,127]. The BF between  $\hat{I}_{M1}$  and  $\hat{I}_L$  is only 1.03 indicating there is only very weak evidence (Box 5.1) for a non-linear model.

The 95% CrIs for the regression coefficients estimated by this method include uncertainty due to model selection as well as from estimation of the parameters of the measurement error and exposure models. This was the only method where uncertainty from model selection was incorporated.

### ***BTS for an FP2 model***

The ‘‘FP1’’ BTS model can be extended to FP2 by incorporating the eight  $h(X_i, p) \log(X_i)$  terms in the linear predictor in addition to the  $h(X_i, p)$  terms where  $p$  is still drawn from  $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ . The FP1 model in Equation 6.8 can already accommodate any FP2

model wherein  $p_1 \neq p_2$ . The first eight terms in the FP2 model are therefore identical to the eight terms of the FP1 model except that the regression coefficients for these terms are denoted  $\beta_{X_{1,k}}$  for  $k = 1, 2, \dots, 8$ . The second set of eight terms represent  $h(X_i, p) \log(X_i)$  and the regression coefficients for these are denoted  $\beta_{X_{2,k}}$ . The  $h(X_i, p) \log(X_i)$  terms for a specific  $p$  should be eligible for inclusion only if  $h(X_i, p)$  is included, i.e. if  $\beta_{X_{1,k}} = 0$ , then  $\beta_{X_{2,k}} = 0$ . This is enforced in the ‘‘FP2’’ BTS model by specifying that  $\beta_{X_{2,k}}$  is dependent on both  $I_{X_{1,k}}$  and  $I_{X_{2,k}}$ . The resulting substantive model and regression coefficient priors are:

$$\begin{aligned}
6.9 \quad g(E[Y_i|X_i, \mathbf{Z}_i]) &= \beta_0 + \beta_{X_{1,1}} X_i^{-2} + \beta_{X_{1,2}} X_i^{-1} + \dots + \beta_{X_{1,7}} X_i^2 + \beta_{X_{1,8}} X_i^3 + \\
&\quad \beta_{X_{2,1}} X_i^{-2} \log(X) + \beta_{X_{2,2}} X_i^{-1} \log(X) + \dots + \beta_{X_{2,8}} X_i^3 \log(X) + \beta_{\mathbf{Z}}^T \mathbf{Z}_i \\
&= \beta_0 + \beta_{X_1} \mathbf{h}(X_i, \mathbf{p}) + \beta_{X_2} \mathbf{h}(X_i, \mathbf{p}) \log(X) + \beta_{\mathbf{Z}}^T \mathbf{Z}_i, \\
\beta_{X_{1,k}} | I_{X_{1,k}} &\sim (1 - I_{X_{1,k}}) N(0, \sigma_a^2) + I_{X_{1,k}} N(0, \sigma_b^2), \\
I_{X_{1,k}} &\sim \text{Bern}(\pi) \\
\beta_{X_{2,k}} | I_{X_{2,k}}, I_{X_{1,k}} &\sim (1 - I_{X_{2,k}} | I_{X_{1,k}}) N(0, \sigma_a^2) + (I_{X_{2,k}} | I_{X_{1,k}}) N(0, \sigma_b^2), \\
I_{X_{2,k}} &\sim I_{X_{1,k}} \text{Bern}(\pi).
\end{aligned}$$

This model does not ensure a second degree model is selected; in fact, both  $\beta_{X_{1,k}}$  and  $\beta_{X_{2,k}}$  may be scalars or vectors in the chosen model. The model can then be represented as  $\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2\}$  where  $\mathbf{p}_2$  is limited to a subset of  $\mathbf{p}_1$ .

## 6.5 SIMULATIONS STUDY

### 6.5.1 Simulation study set up

The simulation scenarios used to assess the selection and fit of the quadratic model in Chapters 4 and 5 were employed again in this section to assess the measurement error correction methods in the context of the fractional polynomial method. In summary, 200 simulations were generated for each study setting (i.e. validation study with  $\sigma_{\tilde{Y}}^2 = 0.5$ , validation study with  $\sigma_{\tilde{Y}}^2 = 1$ , and replicate study with  $\sigma_{\tilde{Y}}^2 = 1$ ) and association shape (i.e. linear, weak quadratic, asymptotic, and J-shaped) ( $200 \times 3 \times 4$ ). Details can be found in Box 5.2. For the ‘‘FP2’’ BTS method, only 100 simulations were run for each validation study scenario.

Because the RC method assumes a lognormal  $X|W$  relationship, an additional sensitivity analysis was performed to evaluate the methods when data were generated from the lognormal error model as described in Section 6.4.1. In this case, the RC method was correctly specified whereas the

hybrid and Bayesian methods were misspecified (for all other simulations in this section, RC was misspecified and the hybrid and Bayesian methods correctly specified). In the lognormal error setting, the latent exposure,  $X_i$ , was drawn from a normal distribution with a mean of 10 and a variance of 1 (i.e.  $\sigma_X^2 = 1$ ) and the outcome was generated as described for each association in Table 4.1. The error prone measure,  $W_1$ , was generated according to the lognormal error model (Equation 6.2;  $\log(W) = \log(X) + U'$ ) with  $\sigma_{U'}^2 = 0.05$  which is roughly equivalent to the variance of the  $\log(X)$ . For this sensitivity analysis, a validation study was assumed in which the latent  $X$  observed for 30% of the participants. For this setting, 100 simulations per association shape were performed ( $100 \times 4$ ).

### 6.5.2 Criteria for evaluating fractional polynomial curve fits

Each of the four correction methods outlined in this chapter were evaluated and compared to the performance of traditional fractional polynomial selection using the underlying  $X$  (i.e. latent  $X$  analysis) or the first error-prone measure,  $W_1$  (i.e. naïve analysis). To assess the power of each method to select a non-linear model or, for a linear association, the type I error in selecting a non-linear model, the proportion which rejected the null and linear models was assessed. For the non-linear associations, selection of the correct degree of the model was assessed by looking at the proportion of simulated data sets which selected an FP1 model versus an FP2 model. Finally, the frequency of each specific model, as represented by  $\mathbf{p}$ , was recorded and compared across methods.

In previous chapters, bias in the regression coefficients was used to assess measurement error correction given the model. However, given that different models may be chosen and be a good fit, for some of the associations, mean bias of the regression coefficients was not meaningful in this context. The performance of the methods was instead assessed by comparing the estimated curve fit with the true curve using the MISE (Section 4.5.2; Equation 4.9) for each simulation. The average MISE across simulations was then estimated. Additionally, for each method and scenario, the average curve fit was estimated as the point-wise mean of the predicted outcome between exposure values  $X = 7$  to 13 in steps of 0.1. The average confidence or credible bounds for each method were likewise estimated.

Each method was evaluated where the maximum (frequentist and hybrid methods) or desired maximum (Bayesian method) degree of the model was 1 ( $m \leq 1$ ) and where the maximum or desired maximum was 2 ( $m \leq 2$ ) allowing for models incorporating  $h(X, \mathbf{p}) \log(X)$ .

For the fully Bayesian method, in addition to the criteria above, for each simulation the estimated marginal posterior probability of each sampled model was assessed as well as the BF comparing highest probability model ( $\hat{\mathbf{I}}_{M1}$ ) to the linear model ( $\mathbf{I}_L$ ) and to the second-highest probability

model ( $\hat{I}_{M2}$ ). The average posterior probability of  $\hat{I}_{M1}$  was obtained although the form of the model may have been different in each simulation. The median of each BF across simulations was also obtained as well as the proportion of simulations for which sufficient evidence was observed to select  $\hat{I}_{M1}$  over the linear model using cut off values of BF = 1, 3, and 10.

### 6.5.3 Method implementation

Example code for simulated data sets and implementation of each method can be found in Appendix C.

For each simulation the fractional polynomial method was applied to the latent  $X$  using the ‘mfp’ package in R [133]. Likewise, for the naïve analysis,  $W_1$  was fit to the fractional polynomial method in place of  $X$ .

For standard RC, 1,000 bootstraps were used to estimate point-wise basic bootstrap 95% intervals for outcome estimates across the range of  $X$  from 7 to 13 in steps of 0.1 [77]. Model selection was incorporated into the bootstrap estimates in a sensitivity analysis for only the setting with a validation study and  $\sigma_{ij}^2 = 1$ . Results can be seen in Appendix C. Otherwise, RC 95% CIs included only uncertainty from measurement error and estimation of the regression parameters.

The model and priors stipulated for MCMC-RC were identical to those outlined in Section 4.5.3 given that the substantive model was not included and no change was required. Burn-in was 40,000 iterations for validation study settings and 70,000 iterations in replicate study settings with 10,000 samples drawn from each of three chains for inference. As in the previous chapter, Gelman and Rubin’s convergence diagnostic,  $\hat{R}$  (Section 2.4), was applied to the last 10,000 burn-in samples to assess convergence, and the ESS was assessed from the final samples.

The INLA portion of the INLA-RC method was implemented exactly as described in Section 4.5.3.

For BTS, the error-prone measures  $W$  and any observed values of  $X$  were standardized (i.e. centered and scaled) before entry into the model. For each iteration of the MCMC sampler, after estimation of the latent  $X_i$  values, the values were re-scaled to the original  $X_i$  scale and centered on the appropriate mean. After this re-scaling, the  $X_i$  values were transformed (i.e.  $h(X_i, p)$ ) then the transformed values were standardized again using estimated standard deviations and means based on the transformations of either the observed measures of  $X$  (validation study) or the second error-prone measure  $W_{i2}$  (replicate study). The outcome was also scaled using its standard deviation before entry into the model. Regression coefficient estimates were back-transformed to the original scale at each iteration for ease of inference.

Prior specification for the variance parameters ( $\sigma_X^2$ ,  $\sigma_U^2$ ,  $\sigma_{Y|X}^2$ ) for BTS was similar to other Bayesian implementations in this thesis; each was assigned an inverse gamma prior,  $IG(1, 1)$ . However, other priors were specified more strongly in this chapter in order to improve mixing and convergence for model selection [28,120]. The prior for the expectation of  $X$ ,  $\mu_X$ , was  $N(0, 100)$  and the prior for the substantive model intercept,  $\beta_0$ , was  $N(0, 100)$ . The prior for all  $\beta_{X_k}$  coefficients when in the “slab” was  $N(0, 3)$  resulting in values between -10 and 10 (i.e.  $\sigma_b^2 = 3$ ) and when in the “spike” was  $N(0, 1/3600)$ . The prior for each  $I_{X_k}$ , or for FP2 BTS  $I_{X_{1,k}}$  and  $I_{X_{2,k}}$ , determining whether  $\beta_{X_k}$  was drawn from the spike or the slab, was a Bernoulli distribution with probability 1/32.

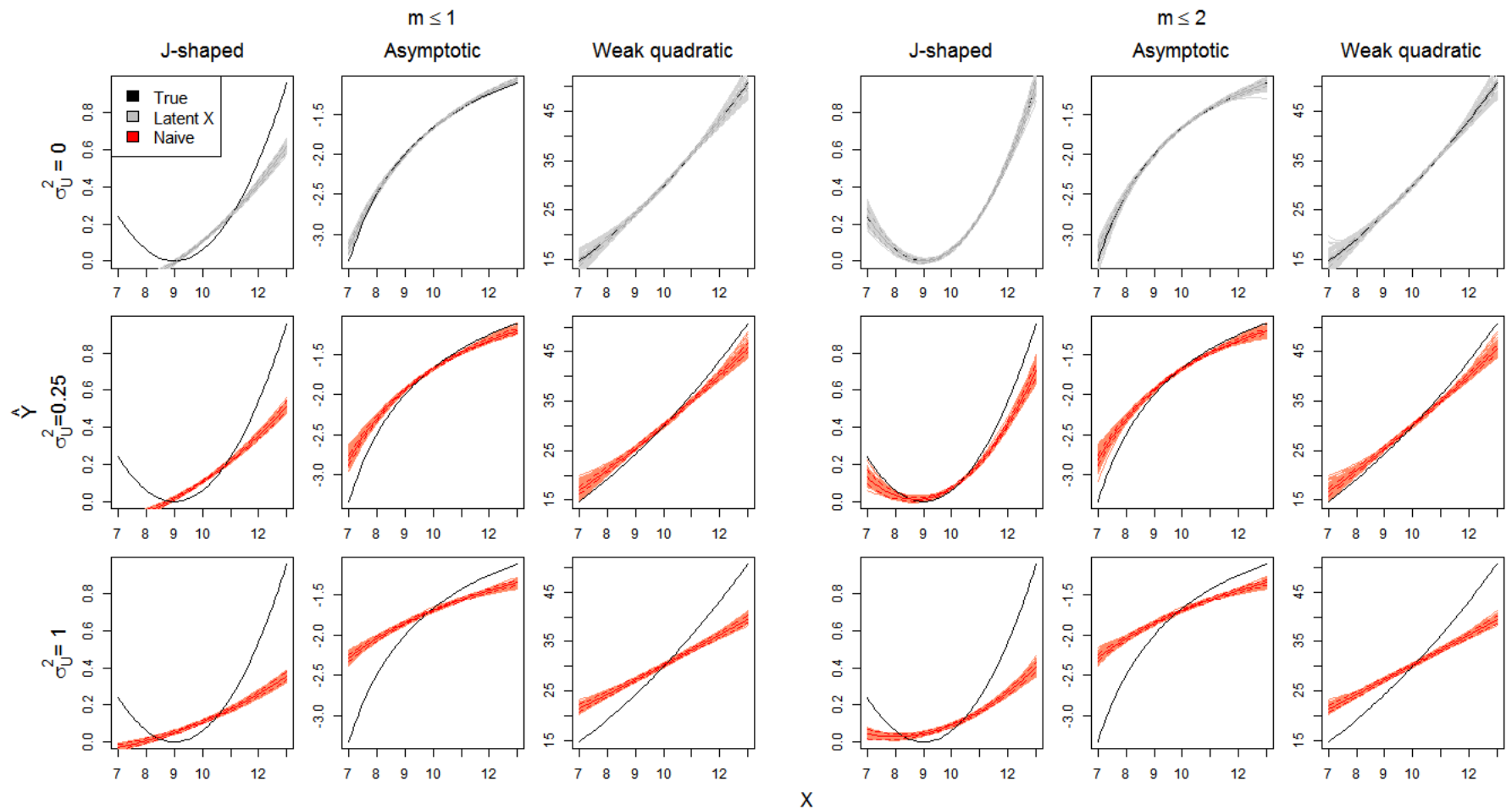
There were five chains run for each simulation for BTS with a burn-in of 40,000 for a validation study and 100,000 for a replicate study; an additional burn-in of 10,000 iterations was run for the weak quadratic and linear associations. After burn-in, 10,000 samples per chain were collected for inference. While  $\hat{R}$  was estimated for the final 10,000 burn-in samples, it is unreliable when multiple models are being sampled [28]. The ESS was assessed from the final samples.

#### 6.5.4 Model selection using the latent $X$

The proportion of simulations which selected the linear model, selected the FP1 model, or selected the FP2 model as well as the average MISE for the fractional polynomial method when applied with the latent  $X$  (either  $m \leq 1$  or  $m \leq 2$ ) can be found in the top lines of Tables 6.2-6.5. The chosen curve fits for each simulation can be found in the top row of Figure 6.1. The null model was rejected for all simulations.

For  $m \leq 2$ , the fractional polynomial method using the latent  $X$  selected a curve very close to the true curve for all association shapes (Figure 6.1). When  $m \leq 1$  and the underlying true curve was a non-monotonic curve over the relevant range of  $X$  (i.e. the J-shaped association), then the fractional polynomial method produced a very poor fit even with the latent  $X$ . This may be observed in the first column of Figure 6.1. However, designation of  $m \leq 1$  resulted in greater power to select non-linear models for the weak quadratic association than did  $m \leq 2$ . The loss of power to detect non-linearity when accepting a higher maximum degree was accompanied by a reduction in type I error associated with selecting a non-linear model when the true association was linear (Table 6.4) [8].

For the asymptotic and J-shaped association, use of the fractional polynomial selection method resulted in lower average MISE values than did specifying the quadratic model in Chapter 4 (Tables 6.3-6.4 and Table 4.3). While the model used to generate the J-shaped association was an FP2 model with  $\mathbf{p} = \{1, 2\}$ , the fractional polynomial method with  $m \leq 2$  resulted in a variety of FP2 models being selected across the simulations (Figure 6.2). The most common model, chosen



**Figure 6.1** The chosen curve fits from the application of the fractional polynomial method with the latent  $X$  and naïve analysis for each of 200 simulations of sample size 2000 (latent  $X$  in gray lines, naïve in red lines). Maximum degrees of 1 ( $m \leq 1$ ) and 2 ( $m \leq 2$ ) were applied. The black line represents the true curve used to generate the simulated data. Darker solid gray/red line is the mean chosen curve fit (the point-wise mean of all curve fits), and the dashed gray/red lines are the mean point-wise 95% confidence bounds. Measurement error variance ( $\sigma_U^2$ ) in the naïve analysis was 0.25, or 1.

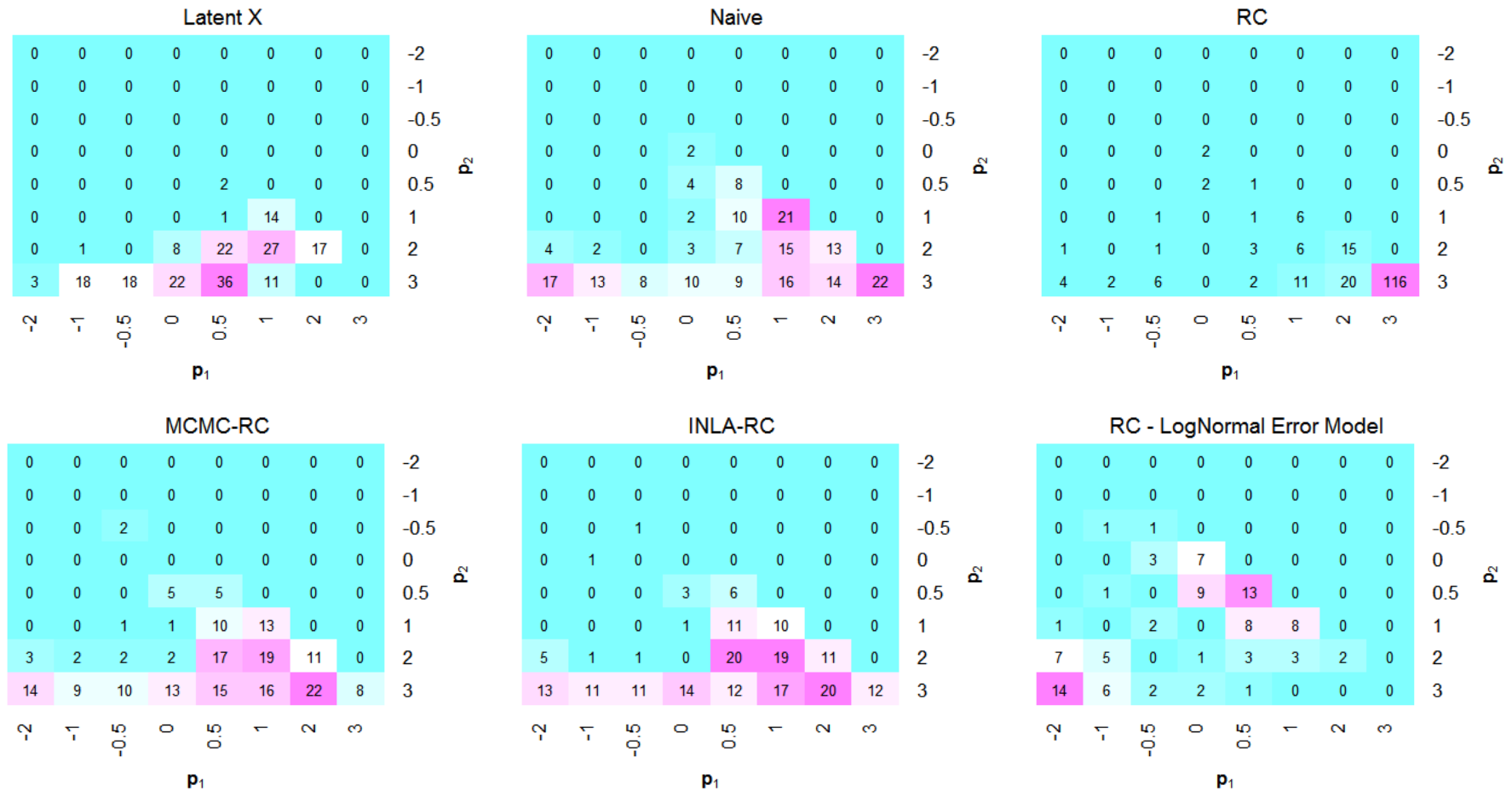
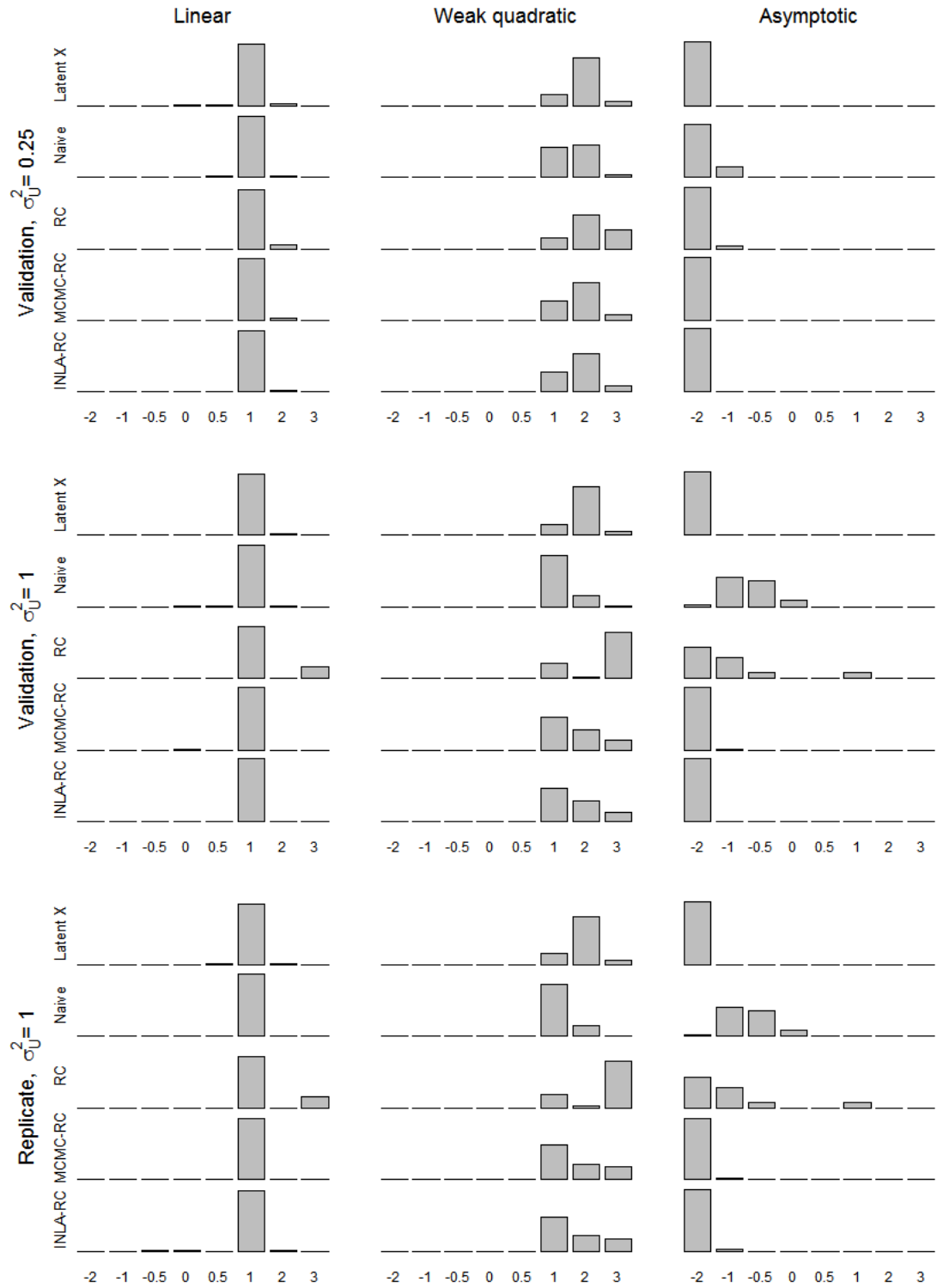


Figure 6.2 Heatmaps for second degree fractional polynomial model selection for the J-shaped association as represented by  $\mathbf{p} = \{p_1, p_2\}$ . Each map shows number of simulations which selected each model by method used for fitting or correction, i.e. the latent exposure analysis (Latent X), the naïve analysis (Naïve), correction with regression calibration (RC), correction with Markov chain Monte Carlo implementation of Bayesian RC (MCMC-RC), or correction with integrated nested Laplace approximations implementation of Bayesian RC (INLA-RC). Simulations assume a validation study in 30% of the participants and are based on the classical error model (first 5 maps;  $W = X + U$ ) with measurement error variance of 1 or the lognormal error model (last map;  $\log(W) = \log(X) + U'$ ) with measurement error variance of 0.05.





**Figure 6.3** Barplots for first degree fractional polynomial model selection for the asymptotic, weak quadratic, and linear associations as represented by the parameter  $p$ . Each plot shows the number of simulations which selected each model by method used for fitting or correction, i.e. the latent exposure analysis (Latent X), the naïve analysis (Naïve), correction with regression calibration (RC), correction with Markov chain Monte Carlo implementation of Bayesian RC (MCMC-RC), or correction with integrated nested Laplace approximations implementation of Bayesian RC (INLA-RC). Simulations are based on the classical error model with a validation study in 30% of the participants and measurement error variance ( $\sigma_U^2$ )  $\frac{1}{4}$  the variance of  $X$ , a validation study and  $\sigma_U^2$  equal to the variance of  $X$ , or a replicate study with replicate measures for all participants and  $\sigma_U^2$  equal to the variance of  $X$ .

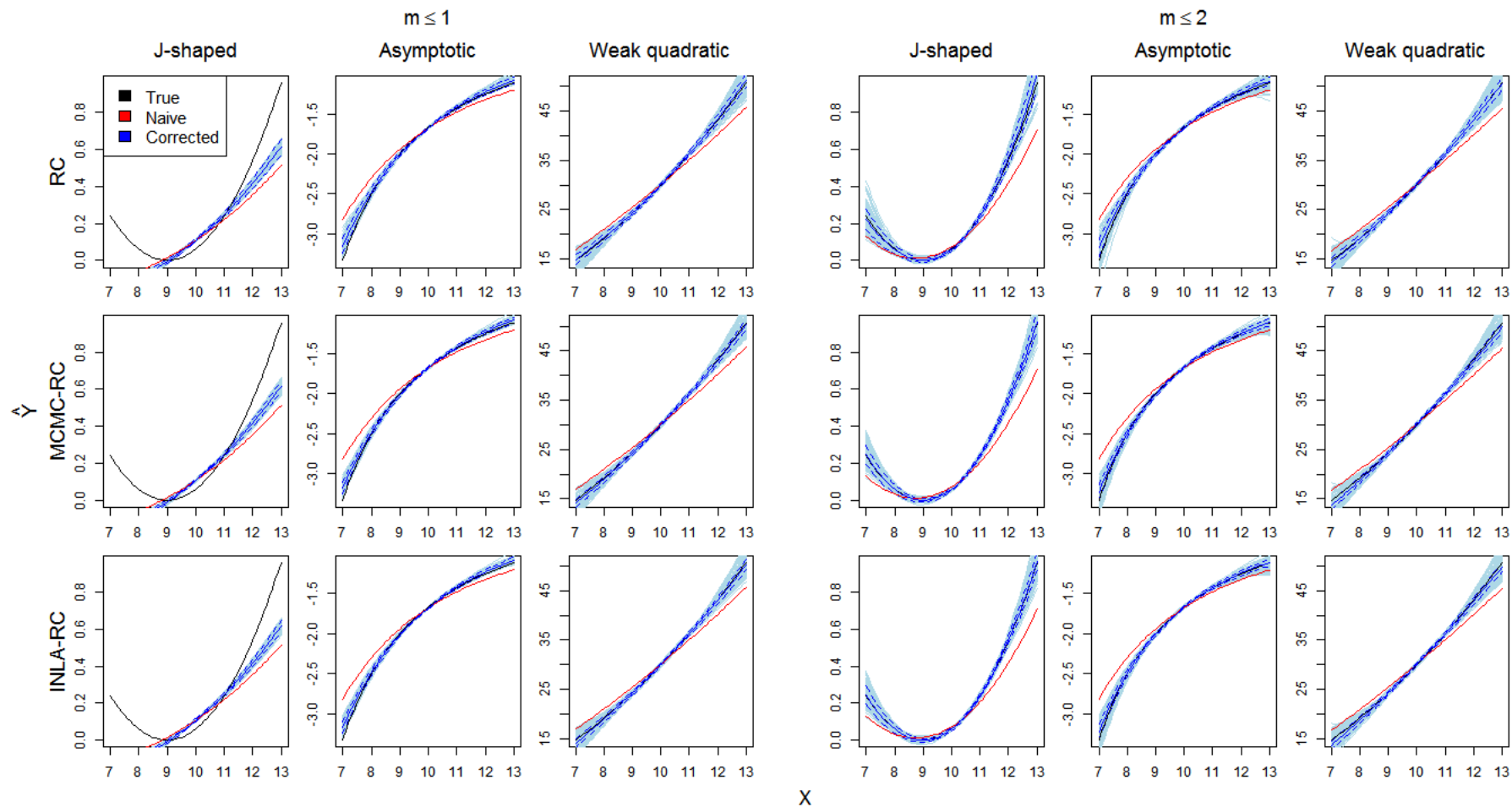
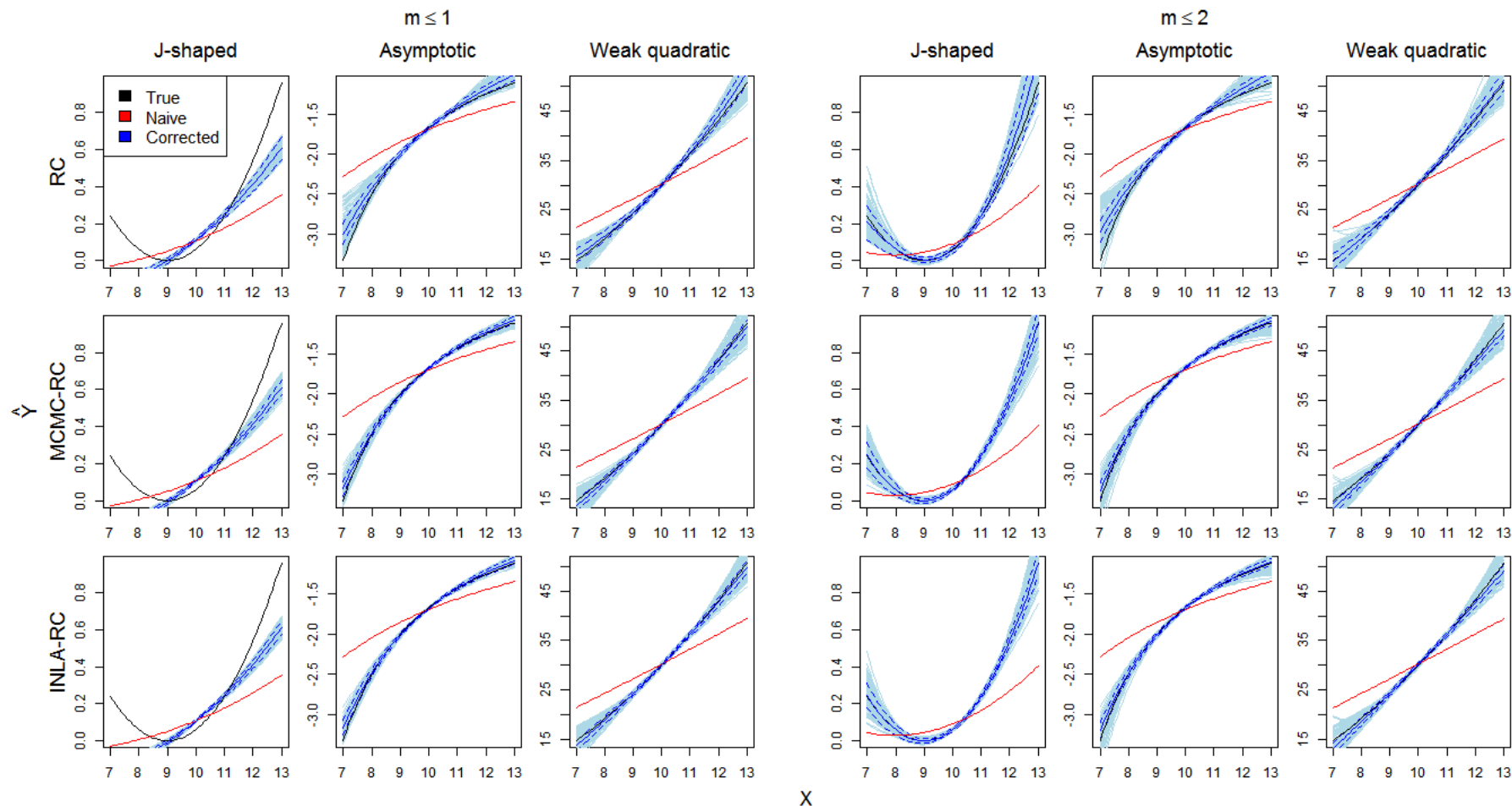
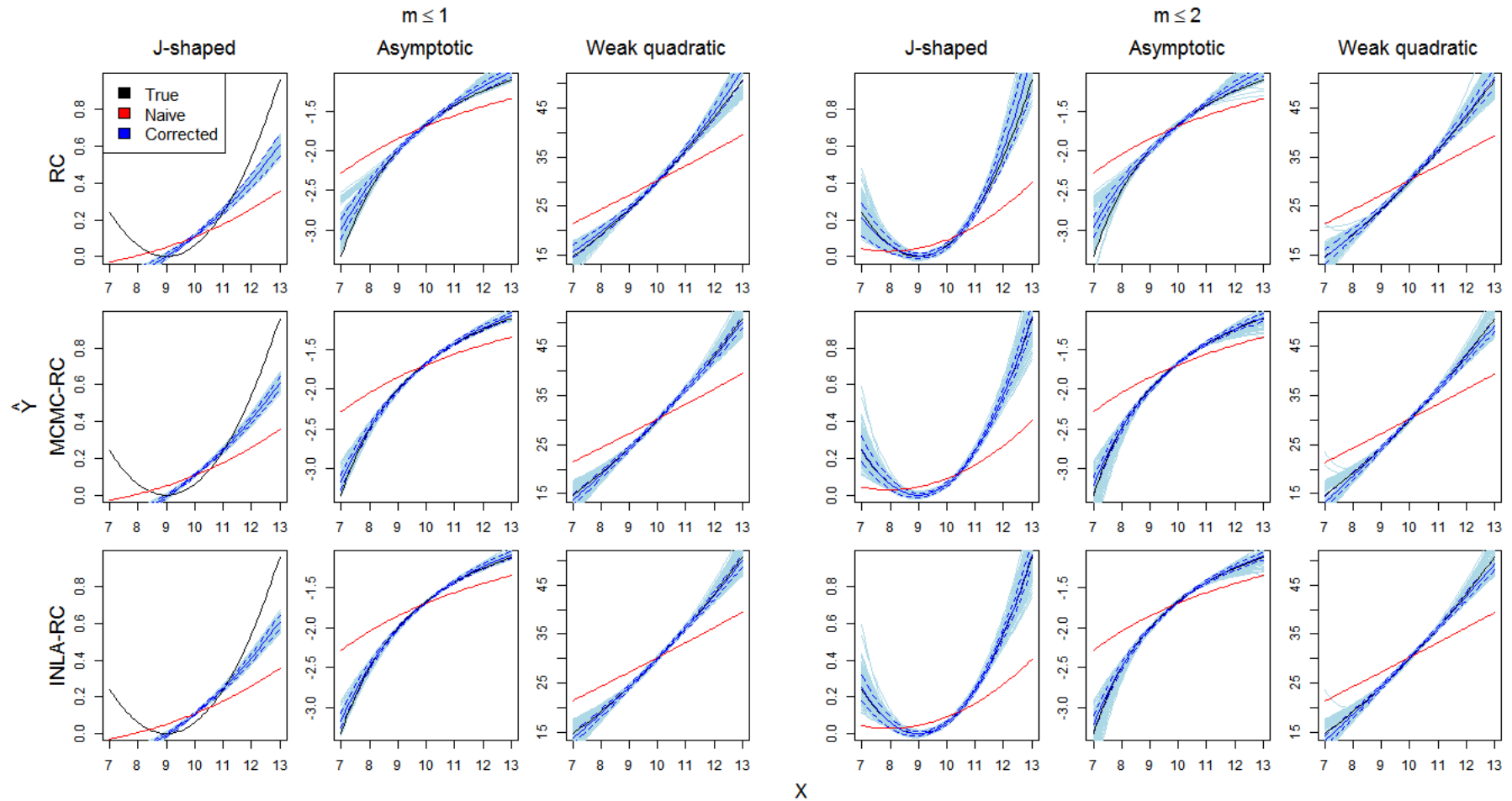


Figure 6.4 The chosen curve fits from the fractional polynomial method with a validation study and classical measurement error variance  $\frac{1}{4}$  the variance of  $X$  for each of 200 simulations of sample size 2000 after measurement error correction using regression calibration (RC), the Markov chain Monte Carlo implementation of Bayesian RC (MCMC-RC), or the integrated nested Laplace approximations implementation of Bayesian RC (INLA-RC) (blue lines). Maximum degrees of 1 ( $m \leq 1$ ) and 2 ( $m \leq 2$ ) were applied. The black line represents the true curve used to generate the simulated data. The red line is the mean naïve curve fit applying the fractional polynomial method with a single error-prone measure. Darker solid blue line is the mean chosen curve fit (the point-wise mean of all curve fits), and the dashed blue lines are the mean point-wise 95% confidence bounds as determined by each method.



**Figure 6.5** The chosen curve fits from the fractional polynomial method with a validation study and classical measurement error variance equal to the variance of  $X$  for each of 200 simulations of sample size 2000 after measurement error correction using regression calibration (RC), the Markov chain Monte Carlo implementation of Bayesian RC (MCMC-RC), or the integrated nested Laplace approximations implementation of Bayesian RC (INLA-RC) (blue lines). Maximum degrees of 1 ( $m \leq 1$ ) and 2 ( $m \leq 2$ ) were applied. The black line represents the true curve used to generate the simulated data. The red line is the mean naïve curve fit applying the fractional polynomial method with a single error-prone measure. The darker solid blue line is the mean chosen curve fit (the point-wise mean of all curve fits), and the dashed blue lines are the mean point-wise 95% confidence bounds as determined by each method.



*Figure 6.6 The chosen curve fits from the fractional polynomial method with a replicate study and classical measurement error variance equal to the variance of  $X$  for each of 200 simulations of sample size 2000 after measurement error correction using regression calibration (RC), the Markov chain Monte Carlo implementation of Bayesian RC (MCMC-RC), or the integrated nested Laplace approximations implementation of Bayesian RC (INLA-RC) (blue lines). Maximum degrees of 1 ( $m \leq 1$ ) and 2 ( $m \leq 2$ ) were applied. The black line represents the true curve used to generate the simulated data. The red line is the mean naïve curve fit applying the fractional polynomial method with a single error-prone measure. The darker solid blue line is the mean chosen curve fit (the point-wise mean of all curve fits), and the dashed blue lines are the mean point-wise 95% confidence bounds as determined by each method.*

**Table 6.2 Simulation model selection and fit for the J-shaped association.** For the latent X analysis, the naïve analysis, and each correction method, the proportion of the 200 simulations of sample size 2000 which selected the linear model, a first degree fractional polynomial (FP) model other than linear (FP1), or a second degree FP model (FP2) as well as the average mean integrated square error (MISE) for each maximum degree of the model  $m \leq 1$  or  $m \leq 2$  is displayed for each of three settings with classical error: measurement error variance ( $\sigma_U^2$ ) equal to 0.25 with a validation study;  $\sigma_U^2 = 1$  with a validation study; and  $\sigma_U^2 = 1$  with a replicate study. RC: regression calibration; MCMC-RC: Bayesian RC implemented with Markov chain Monte Carlo sampling; INLA-RC: Bayesian RC implemented with integrated nested Laplace approximations; BTS: Bayesian transformation selection performed in MCMC specifying a threshold of a Bayes factor of 10 as enough evidence of non-linearity.

		$m \leq 1$				$m \leq 2$			
		Prop. Linear	Prop. FP1	Prop. FP2	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Average MISE
Latent X		0.00	1.00	NA	0.0274	0.00	0.00	1.00	0.0002
Naïve, $\sigma_U^2 = 0.25$		0.00	1.00	NA	0.0315	0.00	0.00	1.00	0.0086
Naïve, $\sigma_U^2 = 1$		0.00	1.00	NA	0.0464	0.00	0.00	1.00	0.0390
Validation $\sigma_U^2 = 0.25$	RC	0.00	1.00	NA	0.0272	0.00	0.00	1.00	0.0008
	MCMC-RC	0.00	1.00	NA	0.0274	0.00	0.00	1.00	0.0004
	INLA-RC	0.00	1.00	NA	0.0274	0.00	0.00	1.00	0.0004
	BTS	0.00	0.00	1.00	0.0003	0.00	0.00	1.00	0.0006
Validation $\sigma_U^2 = 1$	RC	0.00	1.00	NA	0.0268	0.00	0.00	1.00	0.0034
	MCMC-RC	0.00	1.00	NA	0.0276	0.00	0.00	1.00	0.0008
	INLA-RC	0.00	1.00	NA	0.0276	0.00	0.00	1.00	0.0007
	BTS	0.00	0.00	1.00	0.0004	0.00	0.00	1.00	0.0009
Replicate $\sigma_U^2 = 1$	RC	0.00	1.00	NA	0.0267	0.00	0.00	1.00	0.0034
	MCMC-RC	0.00	1.00	NA	0.0276	0.00	0.00	1.00	0.0012
	INLA-RC	0.00	1.00	NA	0.0276	0.00	0.00	1.00	0.0012
	BTS	0.00	0.00	0.35	8.06				

**Table 6.3 Simulation model selection and fit for the asymptotic association.** For the latent X analysis, the naïve analysis, and each correction method, the proportion of the 200 simulations of sample size 2000 which selected the linear model, a first degree fractional polynomial (FP) model other than linear (FP1), or a second degree FP model (FP2) as well as the average mean integrated square error (MISE) for each maximum degree of the model  $m \leq 1$  or  $m \leq 2$  is displayed for each of three settings with classical error: measurement error variance ( $\sigma_U^2$ ) equal to 0.25 with a validation study;  $\sigma_U^2 = 1$  with a validation study; and  $\sigma_U^2 = 1$  with a replicate study. RC: regression calibration; MCMC-RC: Bayesian RC implemented with Markov chain Monte Carlo sampling; INLA-RC: Bayesian RC implemented with integrated nested Laplace approximations; BTS: Bayesian transformation selection performed in MCMC specifying a threshold of a Bayes factor of 10 as enough evidence of non-linearity.

		$m \leq 1$				$m \leq 2$			
		Prop. Linear	Prop. FP1	Prop. FP2	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Average MISE
Latent X		0.00	1.00	NA	0.0017	0.00	0.61	0.39	0.0014
Naïve, $\sigma_U^2 = 0.25$		0.00	1.00	NA	0.0271	0.00	0.97	0.03	0.0271
Naïve, $\sigma_U^2 = 1$		0.00	1.00	NA	0.1283	0.01	0.97	0.02	0.1284
Validation $\sigma_U^2 = 0.25$	RC	0.00	1.00	NA	0.0025	0.00	0.91	0.10	0.0025
	MCMC-RC	0.00	1.00	NA	0.0018	0.00	0.72	0.28	0.0018
	INLA-RC	0.00	1.00	NA	0.0018	0.00	0.73	0.28	0.0018
	BTS	0.00	0.88	0.12	0.0089	0.00	0.72	0.28	0.1085
Validation $\sigma_U^2 = 1$	RC	0.09	0.91	NA	0.0123	0.24	0.75	0.02	0.0174
	MCMC-RC	0.00	1.00	NA	0.0027	0.00	0.89	0.11	0.0028
	INLA-RC	0.00	1.00	NA	0.0023	0.00	0.87	0.13	0.0025
	BTS	0.00	0.91	0.09	0.0081	0.00	0.75	0.25	0.1117
Replicate $\sigma_U^2 = 1$	RC	0.09	0.91	NA	0.0125	0.24	0.75	0.02	0.0176
	MCMC-RC	0.00	1.00	NA	0.0024	0.00	0.87	0.13	0.0032
	INLA-RC	0.00	1.00	NA	0.0024	0.00	0.87	0.13	0.0032
	BTS	0.00	0.02	0.98	5.92				

**Table 6.4 Simulation model selection and fit for the weak quadratic association.** For the latent X analysis, the naïve analysis, and each correction method, the proportion of the 200 simulations of sample size 2000 which selected the linear model, a first degree fractional polynomial (FP) model other than linear (FP1), or a second degree FP model (FP2) as well as the average mean integrated square error (MISE) for each maximum degree of the model  $m \leq 1$  or  $m \leq 2$  is displayed for each of three settings with classical error: measurement error variance ( $\sigma_U^2$ ) equal to 0.25 with a validation study;  $\sigma_U^2 = 1$  with a validation study; and  $\sigma_U^2 = 1$  with a replicate study. RC: regression calibration; MCMC-RC: Bayesian RC implemented with Markov chain Monte Carlo sampling; INLA-RC: Bayesian RC implemented with integrated nested Laplace approximations; BTS: Bayesian transformation selection performed in MCMC specifying a threshold of a Bayes factor of 10 as enough evidence of non-linearity.

		$m \leq 1$				$m \leq 2$			
		Prop. Linear	Prop. FP1	Prop. FP2	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Average MISE
Latent X		0.18	0.83	NA	0.34	0.28	0.71	0.02	0.48
Naïve, $\sigma_U^2 = 0.25$		0.47	0.54	NA	5.04	0.61	0.38	0.02	5.21
Naïve, $\sigma_U^2 = 1$		0.82	0.19	NA	28.8	0.93	0.08	0.00	28.9
Validation $\sigma_U^2 = 0.25$	RC	0.18	0.83	NA	0.65	0.42	0.57	0.01	0.91
	MCMC-RC	0.31	0.69	NA	0.55	0.50	0.49	0.01	0.78
	INLA-RC	0.31	0.69	NA	0.55	0.50	0.49	0.01	0.77
	BTS	0.44	0.56	0.00	1.51	0.39	0.57	0.04	14.5
Validation $\sigma_U^2 = 1$	RC	0.24	0.76	NA	1.76	0.54	0.46	0.01	1.64
	MCMC-RC	0.52	0.48	NA	1.10	0.79	0.20	0.01	1.41
	INLA-RC	0.53	0.48	NA	0.94	0.80	0.19	0.01	1.26
	BTS	0.47	0.54	0.00	1.48	0.45	0.49	0.06	15.7
Replicate $\sigma_U^2 = 1$	RC	0.23	0.78	NA	1.69	0.53	0.46	0.01	1.58
	MCMC-RC	0.55	0.45	NA	1.02	0.79	0.20	0.01	1.29
	INLA-RC	0.55	0.46	NA	1.02	0.79	0.21	0.01	1.28
	BTS	0.09	0.15	0.67	7496				

**Table 6.5 Simulation model selection and fit for the linear association.** For the latent X analysis, the naïve analysis, and each correction method, the proportion of the 200 simulations of sample size 2000 which selected the linear model, a first degree fractional polynomial (FP) model other than linear (FP1), or a second degree FP model (FP2) as well as the average mean integrated square error (MISE) for each maximum degree of the model  $m \leq 1$  or  $m \leq 2$  is displayed for each of three settings with classical error: measurement error variance ( $\sigma_U^2$ ) equal to 0.25 with a validation study;  $\sigma_U^2 = 1$  with a validation study; and  $\sigma_U^2 = 1$  with a replicate study. RC: regression calibration; MCMC-RC: Bayesian RC implemented with Markov chain Monte Carlo sampling; INLA-RC: Bayesian RC implemented with integrated nested Laplace approximations; BTS: Bayesian transformation selection performed in MCMC specifying a threshold of a Bayes factor of 10 as enough evidence of non-linearity.

		$m \leq 1$				$m \leq 2$			
		Prop. Linear	Prop. FP1	Prop. FP2	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Average MISE
Latent X		0.96	0.04	NA	0.12	0.99	0.01	0.01	0.09
Naïve, $\sigma_U^2 = 0.25$		0.96	0.04	NA	4.52	0.98	0.01	0.01	4.51
Naïve, $\sigma_U^2 = 1$		0.98	0.02	NA	27.8	0.99	0.01	0.00	27.8
Validation $\sigma_U^2 = 0.25$	RC	0.94	0.07	NA	0.22	0.98	0.03	0.00	0.17
	MCMC-RC	0.97	0.03	NA	0.15	1.00	0.01	0.00	0.13
	INLA-RC	0.97	0.03	NA	0.15	1.00	0.01	0.00	0.12
	BTS	0.89	0.12	0.00	0.37	0.88	0.11	0.01	6.71
Validation $\sigma_U^2 = 1$	RC	0.82	0.18	NA	1.34	0.94	0.07	0.00	0.74
	MCMC-RC	1.00	0.01	NA	0.37	1.00	0.00	0.00	0.36
	INLA-RC	0.99	0.01	NA	0.23	1.00	0.00	0.00	0.22
	BTS	0.91	0.10	0.00	0.44	0.86	0.13	0.01	5.01
Replicate $\sigma_U^2 = 1$	RC	0.82	0.19	NA	1.33	0.93	0.07	0.00	0.70
	MCMC-RC	0.97	0.04	NA	0.26	0.99	0.01	0.00	0.21
	INLA-RC	0.96	0.04	NA	0.26	0.99	0.01	0.00	0.21
	BTS	0.08	0.16	0.61	8132				



in 18% of simulations, had  $\mathbf{p} = \{0.5, 3\}$ . The second and third most common models had  $\mathbf{p} = \{1, 2\}$  (13.5%) and  $\mathbf{p} = \{0.5, 2\}$  (11%). As the model used to generate the asymptotic association was not consistent with any fractional polynomial model, it was unsurprising that the chosen fit varied between degrees. As can be seen in the top line of the bar plot in Figure 6.3, the FP1 model choice was consistently  $p = -2$  whereas the FP2 model choice most commonly had  $\mathbf{p} = \{-2, -2\}$  with this set of powers being chosen for 78% of the simulations where the FP2 model was selected.

For the weak quadratic association, where the true curve is described by an FP1 model, use of the fractional polynomial method resulted in a higher average MISE than did specifying the quadratic model (Table 6.4 and Table 4.3). Specifying  $m \leq 2$  also resulted in a higher average MISE for this association shape.

In this simulation study, the type I error rates (Table 6.5) were observed to be lower for the latent  $X$  analysis when  $m \leq 2$ . Type I error rates were expected to be lower than the nominal 5% when  $m \leq 2$  when applying the fractional polynomial method [8,134].

### 6.5.5 Model selection using the naïve analysis

The proportion of simulations which selected the linear model, selected the FP1 model, or selected the FP2 model as well as the average MISE for the naïve analysis can be found in the second and third lines of Tables 6.2-6.5. The chosen curve fits for each simulation with the addition of classical measurement error with variance either  $\frac{1}{4}$  the variance of  $X$  or equal to the variance of  $X$  can be found in the second and third rows of Figure 6.1, respectively. The null model was rejected for all simulations.

The addition of measurement error did not lead to any change in the selection of non-linear models for the J-shaped association. However, for  $m \leq 2$ , the average MISE was 43-fold that of the latent  $X$  for  $\sigma_U^2 = 0.25$  and 195-fold for  $\sigma_U^2 = 1$  indicating a much poorer fit in the naïve analysis. In the right half of Figure 6.1, one can see that although the naïve analysis may select an FP2 model, the fit is flattened leading to an underestimation of the association between the exposure and the outcome at the tails and an overestimation between approximately  $X = 8.5$  and 10. The specific FP2 models selected are more varied than those selected for the latent  $X$  as can be seen in Figure 6.2.

For the asymptotic association, the addition of measurement error (whether  $\sigma_U^2 = 0.25$  or  $\sigma_U^2 = 1$ ) resulted in the selection of FP1 models for 97% of simulations. While still non-linear, the rise in the average MISE (16-fold the average MISE of the latent  $X$  for  $\sigma_U^2 = 0.25$  and approximately 76-fold for  $\sigma_U^2 = 1$ ) indicated that these FP1 models were a poorer fit for the underlying curve. For  $\sigma_U^2 = 0.25$ , while the majority select  $p = -2$  (83%), 16.5% selected  $p = -1$ . For  $\sigma_U^2 = 1$ ,

the largest proportion selected  $p = -1$  (46%) followed by  $p = -0.5$  (40.5%) with only 3% selecting  $p = -2$ . These differences in model selection account in part for the much higher average MISE in the naïve analysis.

For the weak quadratic association, the addition of measurement error led to a proportional loss of power to select a non-linear model. Where 83% of simulations selected an FP1 model ( $m \leq 1$ ) in the latent  $X$  analysis, with  $\sigma_{\eta}^2 = 1$  in the naïve analysis only 19% of simulations selected an FP1 model. When the FP1 model was selected, the most common model was  $p = 2$  as expected but when  $\sigma_{\eta}^2 = 0.25$  one simulation selected  $p = 3$  and when  $\sigma_{\eta}^2 = 1$  six simulations selected  $p = 3$  (Figure 6.3).

The naïve analysis did not result in higher type I error rate for the linear association (Table 6.5). However, attenuation did occur in the curve fit as observed in Chapter 3. The resulting average MISE for the naïve analysis was 56-fold and 348-fold that of the latent  $X$  analysis.

### **6.5.6 Results: Model selection using RC and Bayesian RC**

Overall, all three of these correction methods (RC, MCMC-RC, and INLA-RC) were able to dramatically improve model selection and fit over the naïve analysis. The curve fits of the chosen model for each simulation as well as the average chosen curve fit and the average 95% confidence bounds are shown in Figures 6.4-6.7. The proportion of simulations that selected the linear model, selected the FP1 model, or selected the FP2 model as well as the average MISE for both  $m \leq 1$  and  $m \leq 2$  can be found for each method, shape, measurement error variance, and setting in Tables 6.2-6.5.

Because the three correction methods performed very similarly to each other when  $\sigma_{\eta}^2 = 0.25$ , the focus in presenting the results will be on the more challenging setting where  $\sigma_{\eta}^2 = 1$ . Additionally, performance of the three RC-based correction methods was very similar whether a validation study on 30% of the participants or a replicate study on 100% of the participants was available; therefore, results for the replicate study simulations are reported only where they significantly diverge.

Because the simulation study set up is identical to Chapter 4 and the MCMC part of the implementation of MCMC-RC does not change, convergence diagnostics are not given in this chapter as they do not differ.

#### ***J-shaped association***

All three correction methods selected only non-linear models in all simulations (Table 6.2). When  $m \leq 2$ , FP2 models were selected in all simulations.

For  $m \leq 1$ , the corrected average MISE values were at or very near the average MISE for the latent  $X$  analysis (Table 6.2); however, none of these estimated curves were a good fit because an FP1 model was not appropriate (Figure 6.3-6.5). For  $m \leq 2$ , while all methods resulted in an estimated curve close to the true shape, the average MISE of the Bayesian RC methods was roughly 4-fold lower than the average MISE of RC indicating a much better fit. On the right side of Figure 6.5 ( $m \leq 2$ ), one can see greater variability in the chosen RC curve fits. The average RC curve fit has a slightly greater deviation from the true model for at the lower range of exposure values compared to MCMC-RC and INLA-RC. Some bias is expected for RC because it assumes a lognormal error model whereas the data were generated using a classical error model.

The specific models selected when  $\sigma_U^2 = 1$  can be seen in the heat maps in Figure 6.2. RC stands apart with 58% of models selected as  $\mathbf{p} = \{3, 3\}$ . The MCMC-RC heat map and INLA-RC heat map subjectively resemble a compromise between the latent  $X$  and naïve analysis heat maps. The RC model selection, which resembles neither the latent  $X$  nor the naïve analyses, was likely related to the misspecification of the model for RC in this simulation study.

RC also had the widest 95% confidence bounds; however, this was expected as they incorporated uncertainty due to estimation of the measurement error and exposure model parameters whereas the hybrid methods confidence bounds did not.

#### ***Asymptotic association***

All three correction methods resulted in chosen curves very near the true curve (Figure 6.5; Table 6.3). Bias was reduced dramatically over the naïve analysis as indicated by the average MISE being reduced by 10- to 50-fold. The average MISE for RC was 5-fold higher than the average MISE values of the hybrid methods when restricted to  $m \leq 1$  and 7-fold higher when  $m \leq 2$ . The specific model choice when the chosen model was FP1 is shown in the right-most column of Figure 6.3. MCMC-RC and INLA-RC resulted in the selection of a model that has  $p = -2$  in 99.5% of simulations. RC resulted in the selection of a model that has  $p = -2$  in only 54% of simulations. In fact, RC appeared to select a larger range of models than even the naïve analysis. RC also had the greatest variability in the chosen curve fits (Figure 6.5).

#### ***Weak quadratic association***

The loss of power to detect non-linearity when specifying  $m \leq 2$  when the shape was derived from an FP1 model is illustrated with this shape of association (Table 6.4). All three methods had a reduction in power to detect a non-linear model of more than 25% when  $m \leq 2$  was specified.

RC was observed to have the greatest power to detect a non-linear model when applied to this association shape (Table 6.4), but RC often selected the wrong powers leading to a poorer fit than the hybrid methods. The specific model choice when the chosen model was FP1 is shown in the

center column of Figure 6.3. In simulations wherein the application of RC resulted in the selection of an FP1 model, 95% had power  $p = 3$  instead of  $p = 2$ . By comparison, when the hybrid methods resulted in the selection of FP1 models, 31-32% of these had power  $p = 3$ . Consequently, the average MISE was higher for RC than for MCMC-RC and INLA-RC.

INLA-RC had the largest proportion of simulations that selected the linear model (Table 6.4). However, INLA-RC also had the lowest average MISE (approximately 10% lower than MCMC-RC and 25% lower than RC).

### ***Linear association***

The type I error rates, i.e. selection of a non-linear model for the linear association, for MCMC-RC and INLA-RC were observed to be below the nominal level and below the type I error rate observed for the latent  $X$  analysis in all settings. The use of prior information in estimating the parameters of the measurement error and exposure models should result in the reduction of the type I error rate as long as the models are correctly specified.

RC was observed to have a type I error rate of 18% for  $m \leq 1$  where the nominal rate was expected within the fraction polynomial method. However, this implementation of RC assumed the lognormal error model and the data were generated using the classical error model.

### **6.5.7 Results: Lognormal error sensitivity analysis**

In the primary simulation study, RC was observed to have greater power but poorer fit due to the specific models selected. Furthermore, type I error was inflated. This sensitivity analysis assessed whether this was due to misspecification of the model or inherent in the application of RC in this setting.

In the simulation with lognormal error, the proportions of simulations in which the selection of FP1 or FP2 models after RC was applied were much more similar to those observed for MCMC-RC and INLA-RC (Table 6.6). In Figure 6.2, the heatmap of the powers associated with the FP2 models selected for the J-shaped association by RC in this setting was subjectively much more similar to the models selected in the latent  $X$  and naïve analyses. Gains in power to detect non-linearity over that seen for the naïve analysis could be observed for RC.

In this simulation study, the naïve analysis had an extremely elevated type I error rate indicating that the hypothesis testing inherent in the fractional polynomial method was not valid in the presence of lognormal error. After application of RC, the type I error was still elevated but reduced from the naïve analysis.

The hybrid methods performed rather well in this setting wherein the measurement error model was, for these methods, misspecified. For all association shapes, the hybrid methods still resulted

**Table 6.6 Simulation model selection and fit when the error is lognormally related to the exposure.** For the latent X analysis, the naïve analysis, and each correction method, the proportion of the 200 simulations of sample size 2000 which selected the linear model, a first degree fractional polynomial (FP) model other than linear (FP1), or a second degree FP model (FP2) as well as the average mean integrated square error (MISE) for each maximum degree of the model  $m \leq 1$  or  $m \leq 2$  is displayed for the setting where error was generated from a lognormal error model with measurement error variance ( $\sigma_U^2 = 0.05$ ). RC: regression calibration; MCMC-RC: Bayesian RC implemented with Markov chain Monte Carlo sampling; INLA-RC: Bayesian RC implemented with integrated nested Laplace approximations; Prop. = proportion.

		$m \leq 1$			$m \leq 2$			
		Prop. Linear	Prop. FP1	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Average MISE
J-shaped association	Latent X	0	1.00	0.0273	0	0	1.00	0.0002
	Naïve	0	1.00	0.0323	0	0	1.00	0.0119
	RC	0	1.00	0.0278	0	0	1.00	0.0017
	MCMC-RC	0	1.00	0.0277	0	0	1.00	0.0010
	INLA-RC	0	1.00	0.0277	0	0	1.00	0.0010
Asymptotic association	Latent X	0	1.00	0.0018	0	0.63	0.37	0.0014
	Naïve	0	1.00	0.0208	0	0.68	0.32	0.0192
	RC	0	1.00	0.0017	0	0.31	0.69	0.0037
	MCMC-RC	0	1.00	0.0014	0	0.37	0.63	0.0028
	INLA-RC	0	1.00	0.0014	0	0.37	0.63	0.0028
Weak quadratic association	Latent X	0.16	0.84	0.34	0.28	0.72	0	0.46
	Naïve	0.92	0.08	5.82	0.98	0.01	0.01	5.89
	RC	0.85	0.15	1.06	0.95	0.05	0	1.16
	MCMC-RC	0.72	0.28	0.94	0.91	0.09	0	1.15
	INLA-RC	0.72	0.28	0.93	0.91	0.09	0	1.13
Linear association	Latent X	0.99	0.01	0.08	0.99	0.01	0	0.08
	Naïve	0.59	0.41	4.83	0.81	0.18	0.01	4.80
	RC	0.75	0.25	0.45	0.90	0.10	0	0.30
	MCMC-RC	0.75	0.25	0.40	0.93	0.07	0	0.21
	INLA-RC	0.76	0.24	0.40	0.93	0.07	0	0.22

in dramatically lowered average MISE values compared to the naïve analysis. The average MISE values were slightly lower than the correctly-specified RC average MISE values. Similarly, the hybrid methods had slightly larger gains in power to detect non-linearity than the correctly-specified RC. Finally, while type I error rates were still elevated after application of the hybrid methods, they were lower than the naïve analysis and slightly lower than RC.

### 6.5.8 Results: Model selection using BTS

The performance of the BTS method presented in this chapter (Section 6.4.3) compared alongside the other methods of measurement error correction can be found in Tables 6.2-6.5. In these tables, the non-linear model was selected over the linear model only if the BF was at least 10 (strong evidence). This information is repeated in Tables 6.7-6.8 for a cut-off value of  $BF \leq 3$  (moderate evidence) alongside a further breakdown of the proportion of simulations that met each evidence category.

**Table 6.7 Simulation model selection using Bayesian transformation selection (BTS) results including only fractional polynomial terms  $h(X, \mathbf{p})$  (“FP1”) for four shapes of association and three settings with classical error: measurement error variance ( $\sigma_U^2$ ) equal to 0.25 with a validation study;  $\sigma_U^2 = 1$  with a validation study; and  $\sigma_U^2 = 1$  with a replicate study. 200 simulations of sample size 2000 were performed for each shape and setting. MISE = mean integrated square error estimated from exposure value 7 to 13; BF = Bayes factor; FP1 = first degree fractional polynomial model excluding the linear model; FP2 = second degree fractional polynomial model;  $\hat{\mathbf{I}}_{M1}$  = model found to have the highest posterior probability;  $\hat{\mathbf{I}}_{M2}$  = model found to have the second highest posterior probability;  $\mathbf{I}_L$  = the linear model; Prop. = proportion; Freq. = frequency; and Prob. = probability. The average MISE assumes a linear model whenever the BF < 3. \*Indicates that not all data sets would run within the MCMC model: 2 for  $\sigma_U^2 = 0.25$  and 1 for  $\sigma_U^2 = 1$ .**

	Most freq. $\mathbf{p}$	Mean posterior prob. of $\mathbf{I}_{M1}$	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Median BF $\hat{\mathbf{I}}_{M1}$ vs $\hat{\mathbf{I}}_L$	$\hat{\mathbf{I}}_{M1}$ vs $\hat{\mathbf{I}}_L$ BF $\geq 3$	$\hat{\mathbf{I}}_{M1}$ vs $\hat{\mathbf{I}}_L$ BF $\geq 10$	$\hat{\mathbf{I}}_{M1}$ vs $\hat{\mathbf{I}}_L$ BF $\geq 100$	Median BF $\hat{\mathbf{I}}_{M1}$ vs $\hat{\mathbf{I}}_{M2}$
J-shaped association											
Validation study, $\sigma_U^2 = 0.25$	-0.5, 3	0.44	0.0003	0.00	0.00	1.00	>100	1	1	1	1.9
Validation study, $\sigma_U^2 = 1$	-2, 3	0.43	0.0004	0.00	0.00	1.00	>100	1	1	1	1.9
Replicate study, $\sigma_U^2 = 1$	3	0.38	8.06	0.00	0.00	0.35	>100	1	1	1	1.4
Asymptotic association											
Validation study, $\sigma_U^2 = 0.25^*$	-2	0.61	0.0089	0.00	0.88	0.12	>100	1	1	1	1.8
Validation study, $\sigma_U^2 = 1^*$	-2	0.63	0.0081	0.00	0.91	0.09	>100	1	1	1	1.9
Replicate study, $\sigma_U^2 = 1$	0, 3	0.91	5.92	0.00	0.02	0.98	>100	0.97	0.94	0.93	36.1
Weak quadratic association											
Validation study, $\sigma_U^2 = 0.25$	2	0.60	1.14	0.32	0.69	0.00	78.5	0.69	0.56	0.5	1.8
Validation study, $\sigma_U^2 = 1$	2	0.62	1.09	0.31	0.70	0.00	18.4	0.70	0.54	0.43	2.0
Replicate study, $\sigma_U^2 = 1$	1, 2	0.27	7530	0.06	0.15	0.69	>100	0.93	0.91	0.87	1.1
Linear association											
Validation study, $\sigma_U^2 = 0.25$	1	0.55	0.42	0.83	0.16	0.01	1.01	0.16	0.12	0.11	1.7
Validation study, $\sigma_U^2 = 1$	1	0.53	0.54	0.80	0.20	0.01	1.00	0.20	0.10	0.09	1.5
Replicate study, $\sigma_U^2 = 1$	0.5, 1	0.28	8184	0.05	0.16	0.64	>100	0.94	0.91	0.90	1.1

**Table 6.8 Simulation model selection using Bayesian transformation selection (BTS) results including fractional polynomial terms  $h(X, \mathbf{p})$  and  $h(X, \mathbf{p})\log(X)$  (“FP2”) for four shapes of association and three settings with classical error: measurement error variance ( $\sigma_U^2$ ) equal to 0.25 with a validation study;  $\sigma_U^2 = 1$  with a validation study; and  $\sigma_U^2 = 1$  with a replicate study. 200 simulations of sample size 2000 were performed for each shape and setting. MISE = mean integrated square error estimated from exposure value 7 to 13; BF = Bayes factor; FP1 = first degree fractional polynomial model excluding the linear model; FP2 = second degree fractional polynomial model;  $\hat{I}_{M1}$  = model found to have the highest posterior probability;  $\hat{I}_{M2}$  = model found to have the second highest posterior probability;  $I_L$  = the linear model; Prop. = proportion; Freq. = frequency; and Prob. = probability. The average MISE assumes a linear model whenever the  $BF < 3$ .**

	Most freq. $\mathbf{p}$	Mean posterior prob. of $I_{M1}$	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Median BF $\hat{I}_{M1}$ vs $\hat{I}_L$	$\hat{I}_{M1}$ vs $\hat{I}_L$ BF $\geq 3$	$\hat{I}_{M1}$ vs $\hat{I}_L$ BF $\geq 10$	$\hat{I}_{M1}$ vs $\hat{I}_L$ BF $\geq 100$	Median BF $\hat{I}_{M1}$ vs $\hat{I}_{M2}$
J-shaped association											
Validation study, $\sigma_U^2 = 0.25$	-0.5, 3	0.40	0.0064	0.00	0.00	1.00	>100	1.00	1.00	1.00	1.6
Validation study, $\sigma_U^2 = 1$	-0.5, 3	0.41	0.0009	0.00	0.00	1.00	>100	1.00	1.00	1.00	1.9
Asymptotic association											
Validation study, $\sigma_U^2 = 0.25$	-2	0.54	0.1085	0.00	0.72	0.28	>100	1.00	1.00	1.00	1.9
Validation study, $\sigma_U^2 = 1$	-2	0.52	0.1117	0.00	0.75	0.25	>100	1.00	1.00	1.00	1.5
Weak quadratic association											
Validation study, $\sigma_U^2 = 0.25$	2	0.56	12.8	0.28	0.67	0.05	>100	0.72	0.61	0.58	1.6
Validation study, $\sigma_U^2 = 1$	2	0.55	13.7	0.33	0.61	0.06	29.4	0.67	0.55	0.46	1.9
Linear association											
Validation study, $\sigma_U^2 = 0.25$	1	0.55	6.37	0.79	0.19	0.02	1.00	0.21	0.12	0.12	1.9
Validation study, $\sigma_U^2 = 1$	1	0.52	5.32	0.74	0.24	0.02	1.51	0.26	0.14	0.14	1.6

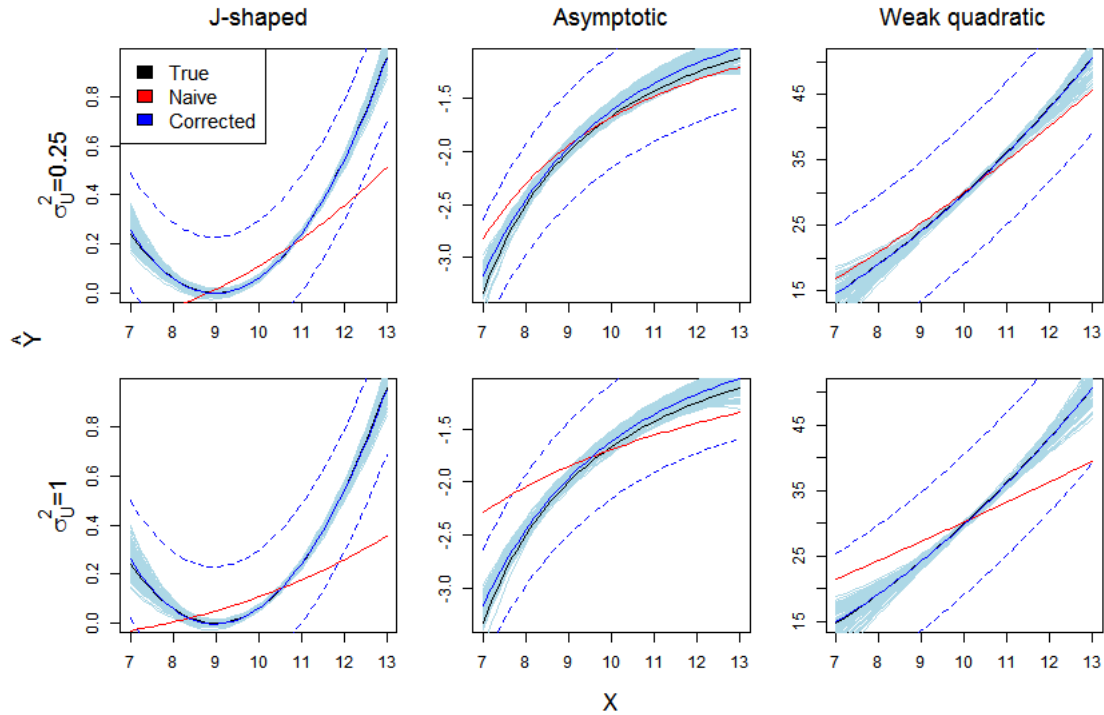


Figure 6.7 The curve fits after application of Bayesian transformation selection (BTS) including only fractional polynomial terms  $h(X, \mathbf{p})$  (“FP1”) for each of 200 simulations with a validation study in 30% in 2000 participants. The black line represents the true curve used to generate the simulated data. The red line is the mean naïve curve fit applying the fractional polynomial method with a single error-prone measure. Darker solid blue line is the mean chosen curve fit (the point-wise mean of all curve fits), and the dashed blue lines are the mean point-wise 95% confidence bounds as determined by each method. Measurement error variance ( $\sigma_U^2$ ) 0.25, or 1.

In general, the BTS method presented in this chapter was able to reduce bias dramatically over the naïve analysis when a validation study was present (Figure 6.7; Tables 6.2-6.5). The use of prior information and the full joint likelihood also resulted in greater power to detect non-linearity and selection of the correct degree of the fractional polynomial model. The gains in power were observed whether one required moderate evidence ( $BF \geq 3$ ; Tables 6.7-6.8) or strong evidence ( $BF \geq 10$ ; Tables 6.2-6.5)

In Figure 6.7, the 95% credible bounds can be seen to be very wide but keeping the same shape as the curve fits. This is due to extremely wide 95% CrIs for the intercept; the 95% CrIs for the regression coefficients are much smaller as might be expected given the strong priors.

The “FP2” extension had higher average MISE values than the “FP1” version as would be expected given the additional parameters in the model without additional supporting data. None of the simulated associations in this thesis contained any of the  $h(X, \mathbf{p}) \log(X)$  terms; such a model would require the “FP2” extension.

Due to the extreme collinearity between transformations of the latent variable, there was a tendency to over-select from the ends of the range of  $\mathbf{S}$  in this method (Figure 6.8); ie. to select models such as with  $\mathbf{p} = \{-2, 3\}$  because they had the lowest correlation. As observed for the J-shaped association, this did not result in a poorer fit because second degree models had a greater



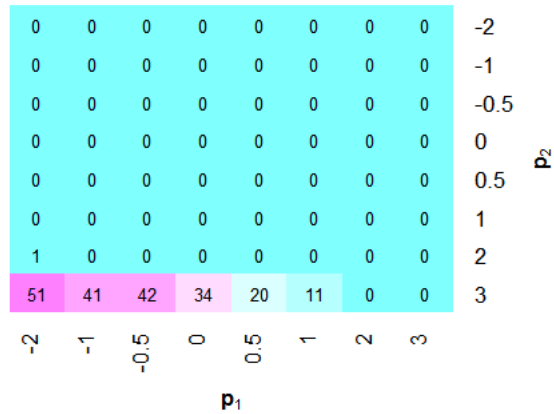


Figure 6.8 Heatmap for model selection for the J-shaped association applying Bayesian transformation selection including only fractional polynomial terms  $h(X, p)$  (“FP1”). Model selection represented by  $\mathbf{p} = \{p_1, p_2\}$ . Performed in 200 simulations with sample size 2000 each wherein a validation study was performed on 30% of the participants and classical measurement error wherein the variance is equal to the variance of the exposure.

range of flexibility. The model chosen most frequently had  $\mathbf{p} = \{-0.5, 3\}$  for  $\sigma_U^2 = 0.25$  (27%) and  $\mathbf{p} = \{-2, 3\}$  for  $\sigma_U^2 = 1$  (26%). The average MISE values for the J-shaped association for either  $\sigma_U^2 = 0.25$  or  $\sigma_U^2 = 1$  were nearly as low as the average MISE for the latent  $X$  analysis with  $m \leq 2$  (Table 6.2).

For the asymptotic association, the proportions of each degree of model (linear, FP1, or FP2) were most similar using this method to the latent  $X$  analysis with  $m \leq 2$  (Table 6.3). The most frequent model selected (80% of simulations) had  $p = -2$ .

For the weak quadratic association, when a validation study was present and  $\sigma_U^2 = 1$ , application of “FP1” BTS with measurement error correction selected a non-linear model in 35% more simulations than the naïve analysis and 6% more simulations than the Bayesian RC methods. Unlike RC, the non-linear model selected had  $\mathbf{p} = 2$  for all simulations. The average MISE in this scenario was 20-fold lower than the average MISE for the naïve analysis, 0.5-fold higher than the average MISE values for the Bayesian RC methods, and lower than the average MISE for RC. The increased bias and variability over the Bayesian RC methods was due to variability in estimation of the regression coefficients rather than variability in the model selection.

In the setting of a replicate study, BTS did not perform well for any association shape. When a replicate study was present, the average MISE after application of BTS was >100-fold higher than the average MISE of the naïve method for all association shapes. For the J-shaped association, usually best fit by a second degree model, a higher degree model was chosen most frequently (65%). For the asymptotic, weak quadratic, and linear associations, usually best fit by first degree models, a second degree model was chosen most frequently (98, 67, and 61%, respectively). For this reason, the “FP2” implementation of BTS was not run for a replicate study.

A replicate study on all participants resulted in a likelihood in this context (where  $\sigma_U^2 = 1$ ) that is less empirically identifiable than a validation study on 30% of the participants at the same measurement error variance. The parameters in the measurement error and exposure models were adapted to suit any model selection leading to an over selection of more complex models and bias in the resulting estimates. This phenomenon was termed model “feedback” [135]; feedback has been a known drawback of “modular” Bayesian models [129,136,137].

The phenomenon of “feedback” was also evidenced in the high type I error rates observed. Even when requiring strong evidence ( $BF \geq 10$ ; Table 6.5) in order to select the non-linear model with the highest posterior probability over the linear model, a non-linear model was selected in 10-13% of simulations. If requiring only moderate evidence, the empirical type I error rate was 21-26%.

## 6.6 LOGISTIC REGRESSION

RC, MCMC-RC, and INLA-RC require an approximation for logistic regression. Furthermore, the specific samplers employed by JAGS for MCMC are different for logistic regression and may affect the efficiency of the BTS method in sampling from the models. Therefore, in this section these methods of measurement error correction presented in this chapter in conjunction with the fractional polynomial method were applied in a simulation study wherein the outcome was binary and the substantive model was logistic regression. Power is of the greatest concern in this context as measurement error reduces power to detect non-linear associations, the fractional polynomial method has relatively low power to detect non-linearity, and power is further reduced given a binary outcome.

### 6.6.1 Simulation study set up

The latent exposure,  $X$ , and the error-prone measure,  $W_1$ , are generated as described in Section 5.6.1 and shown again here in Box 6.1.

The J-shaped association from Section 5.6.1 was generated for this simulation study without alteration,

#### Box 6.1 Simulation study design

Simulations per scenario: 100

True exposure:  $X_i \sim N(10, 1)$

Classical error model:

$$W_{ij} = X_i + U_{ij}, \quad U_{ij} \sim N(0, \sigma_U^2), \quad j = 1, 2$$

Outcome models:

J-shaped association,  $N = 2000$ :

$$P(Y_i = 1|X_i) \sim \text{Bern}(-1.8 + 0.2(X - 9)^2)$$

Weak quadratic association,  $N = 5000$ :

$$P(Y_i = 1|X_i) \sim \text{Bern}(-7.8 + 0.06 * X^2)$$

Linear association,  $N = 5000$ :

$$P(Y_i = 1|X_i) \sim \text{Bern}(-7.66 + 0.62 * X)$$

Setting: Validation study in 30% of participants,  $\sigma_U^2 = 0.25$

i.e.  $N = 2000$ . However, the MCMC sampler was very slow for BTS with logistic regression and large sample sizes so for reasons of expediency, the weak quadratic association was strengthened and the sample size reduced to  $N = 5000$ . The linear association was likewise altered to have a sample size of  $N = 5000$ . The linear predictors for each association shape are shown in Box 6.1.

Only a validation study with  $\sigma_U^2 = 0.25$  was considered in this simulation study and 100 simulations were run for each association shape ( $3 \times 100$ ). The RC, MCMC-RC, INLA-RC, and “FP1” BTS methods of measurement error correction and model selection described in this chapter were applied to each simulated data set.

The same criteria for evaluation of the performance as described in Section 6.5.2 were applied to the results of this simulation study.

### 6.6.2 Implementation of methods

All methods were implemented as described in Section 6.5.3 with changes as described below. No changes for RC or the hybrid methods were required other than the use of the logistic model as the substantive model within the fractional polynomial method.

For the implementation of the “FP1” BTS model, in addition to the use of a logistic model, no scaling of the binary outcome was applied. Each of five chains were run for a burn-in of 100,000 iterations. 10,000 samples were collected from each chain for inference.

### 6.6.3 Fitting the latent $X$ and naïve models

For the J-shaped association, all simulations selected the FP1 model when  $m \leq 1$  and 95% selected the FP2 model (four selected the FP1 model) for the latent  $X$  analysis (Table 6.9). The FP1 model was not a good fit as the average MISE for  $m \leq 1$  was 0.320 while the average MISE for  $m \leq 2$  was 0.095. The most frequent model selected when  $m \leq 2$  was  $\mathbf{p} = \{3, 3\}$  (42%).

In the naïve analysis, with the addition of measurement error with  $\sigma_U^2 = 0.25$ , simulations with the J-shaped association selected the linear model for  $m \leq 1$  or  $m \leq 2$  in 8% and 11% of simulations, respectively (Table 6.9). This highlights the loss of power to detect the non-linear association attributable to measurement error. The average MISE for  $m \leq 2$  was 0.175 for the naïve analysis. The model  $\mathbf{p} = \{3, 3\}$  was still the most frequently selected, although only for 31% of simulations.

When the latent  $X$  analysis was applied in the scenario of the weak quadratic association specifying  $m \leq 1$ , 27% of simulations selected FP1 and 73% selected the linear model. For  $m \leq 2$ , 11% selected FP1 and 88% selected the linear model. The most frequent FP1 model selected was  $p = 2$  (69% of FP1 models selected) and the average MISE was 0.068.

**Table 6.9 Simulation model selection and fit for logistic regression models.** For the latent  $X$  analysis, the naïve analysis, and each correction method, the proportion of the 100 simulations of sample size ( $N$ ) indicated which selected the linear model, a first degree fractional polynomial (FP) model other than linear (FP1), or a second degree FP model (FP2) as well as the average mean integrated square error (MISE) for each maximum degree of the model  $m \leq 1$  or  $m \leq 2$  is displayed for the setting where error was generated from a classical error model with measurement error variance  $\sigma_{\epsilon}^2 = 0.25$ . RC: regression calibration; MCMC-RC: Bayesian RC implemented with Markov chain Monte Carlo sampling; INLA-RC: Bayesian RC implemented with integrated nested Laplace approximations; BTS: Bayesian transformation selection performed in MCMC specifying a threshold of a Bayes factor of 10 as enough evidence of non-linearity; Prop. = proportion.

		$m \leq 1$				$m \leq 2$			
		Prop. Linear	Prop. FP1	Prop. FP2	Average MISE	Prop. Linear	Prop. FP1	Prop. FP2	Average MISE
J-shaped association N=2000	Latent X	0.00	1.00	NA	0.320	0.01	0.04	0.95	0.095
	Naïve	0.08	0.92	NA	0.348	0.11	0.16	0.73	0.175
	RC	0.08	0.92	NA	0.323	0.09	0.10	0.81	0.127
	MCMC-RC	0.06	0.94	NA	0.324	0.06	0.06	0.88	0.094
	INLA-RC	0.06	0.94	NA	0.324	0.06	0.06	0.88	0.093
	BTS	0.10	0.23	0.67	0.116				
Weak quadratic association N=5000	Latent X	0.73	0.27	NA	0.068	0.88	0.11	0.01	0.075
	Naïve	0.81	0.19	NA	0.232	0.92	0.08	0.00	0.224
	RC	0.76	0.24	NA	0.061	0.90	0.10	0.00	0.046
	MCMC-RC	0.80	0.20	NA	0.054	0.91	0.09	0.00	0.050
	INLA-RC	0.80	0.20	NA	0.054	0.91	0.09	0.00	0.050
	BTS	0.58	0.42	0.00	0.081				
Linear association N=5000	Latent X	0.96	0.04	NA	0.010	1.00	0.00	0.00	0.007
	Naïve	0.96	0.04	NA	0.062	1.00	0.00	0.00	0.061
	RC	0.97	0.03	NA	0.010	1.00	0.00	0.00	0.009
	MCMC-RC	0.95	0.05	NA	0.012	1.00	0.00	0.00	0.008
	INLA-RC	0.94	0.06	NA	0.013	1.00	0.00	0.00	0.008
	BTS	0.68	0.32	0.00	0.061				

The addition of measurement error lead to only a small increase in the selection of the linear model (81%) over the non-linear FP1 models (19%) for the weak quadratic association when  $m \leq 1$ . However, the average MISE was increased almost 4-fold. Results for the weak quadratic association were quite similar when  $m \leq 2$  was specified.

When applied to the scenario with a linear association, both the latent  $X$  and naïve analyses had the type I error rate around the nominal level when  $m \leq 1$  (4%) and near zero when  $m \leq 2$ .

#### 6.6.4 Results after application of correction methods

The proportion of simulations which selected the linear model, selected an FP1 model, or selected an FP2 model as well as the average MISE for both  $m \leq 1$  and  $m \leq 2$  can be found for each association shape and method in Table 6.9.

Overall, RC, MCMC-RC, and INLA-RC methods corrected for exposure measurement error and recovered much of the power lost for both non-linear association shapes (Table 6.9). The type I

error observed for the linear association was near the nominal rate. These three methods performed much more similarly to each other in this setting than was observed in the simulation study with linear regression. That is, the power gains and average MISE values were relatively similar over the scenarios tested.

In this setting, BTS did not perform well for any association shape. For the J-shaped association, BTS selected the linear model more frequently than any other method (10% of simulations) except the naïve analysis (11% of simulations) (Table 6.9). For the weak quadratic association, BTS resulted in selection of an FP1 model in a larger proportion of simulations than for the latent  $X$  analysis. Similarly, when the underlying association was linear, the use of BTS resulted in selection of a non-linear model in 32% of simulations.

The poor performance of BTS in this setting was similar to the performance when the outcome was continuous and only a replicate study was available with high measurement error variance. This lack of discrimination between models was due to Bayesian feedback [135].

## 6.7 SUMMARY

In this chapter, I have extended the hybrid and fully Bayesian methods introduced in Chapters 4 and 5 for use in the context of the fractional polynomial method.

For the frequentist or hybrid methods, designating  $m \leq 1$  resulted in slightly greater power to detect a non-linear model than did designating  $m \leq 2$ . For the fully Bayesian method, the “FP1” implementation allowed for the detection of some but not all of the traditional FP2 models, i.e. those which did not incorporate  $h(X, p)\log(X)$ .

The standard frequentist method, RC, was shown to reduce bias from exposure measurement error even when misspecified. However, while the implementation of RC improved power to detect non-linear models, the wrong models were selected resulting in some remaining bias as shown by the average MISE. Furthermore, RC was observed to have inflated type I error. Both these attributes stem from the fact that the data were generated using a different error model than that specified in this implementation of RC. RC required the assumption that  $X|W$  is log-normally distributed, i.e. the lognormal error model. While performance improved when the data were generated under this measurement error model, the restriction to this error-model represents a significant drawback to the use of the method. As in previous chapters, RC was also observed to have greater variability in the estimated curves than the Bayesian RC methods.

The hybrid methods, MCMC-RC and INLA-RC, corrected for bias from measurement error as well or better than standard RC in all settings. Additionally, they were shown to improve power to detect the non-linear models and to frequently select the model most similar to that used to

generate the data. The hybrid methods were also shown to have nominal type I error rates when the model was correctly specified. Unlike RC, these methods did not incorporate uncertainty due to measurement error in the final model estimates; however, the 95% confidence bounds observed in Figures 5.4-5.7 for the hybrid methods are not substantially different from those observed for RC.

Notably, when data was generated from the lognormal error model, leaving RC correctly specified and the Bayesian RC methods misspecified, the Bayesian RC methods still resulted in greater gains in power and lower average MISE values than RC. The type I error rates were similar for the correctly-specified RC and the misspecified hybrid methods. Furthermore, adapting the Bayesian RC methods to specify the lognormal error model instead of the classical error model would be a straightforward extension to implement for these methods.

The fully Bayesian method, BTS using SSVS, shows the potential for full integration of both model selection and measurement error correction. When a validation study was present and the outcome was linear, use of BTS for model selection while correcting for classical measurement error led to reduction in bias on average, but had wider variability across simulations than the RC methods. Furthermore, the 95% credible bounds were very wide due to uncertainty in the intercept as the method allowed for alternate model selection. However, when there was insufficient information in the likelihood, use of SSVS resulted in over selection of more complex models and in extreme cases, such as the setting in this chapter with a replicate study, resulted in uninformative posterior distributions due to model feedback. Model feedback also resulted in inflated type I error in all settings.

## **7 APPLICATION TO ALCOHOL AND MORTALITY IN THE EPIC-NORFOLK STUDY**

---

### **7.1 AIMS & OVERVIEW**

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 [24–26], using data from the EPIC-Norfolk study [68,69]. The association between alcohol intake and all-cause mortality has previously been found to be J-shaped, but whether this is the true underlying association shape is a subject of debate [138]. 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. The model representing this compound error is termed the ‘episodic consumers (EC) model’ [17,19,59,64,65]. The EC model has been fitted using both maximum likelihood [17,65] and Bayesian methods [19,59] but has not previously been used in the setting of the fractional polynomial method.

In this application, I aimed to make use of replicate study data to correct for complex measurement error while fitting a potentially non-linear association. The only methods that have performed well in this setting in the simulation studies in Chapter 6 were the Bayesian RC methods, MCMC-RC and INLA-RC. Therefore, these methods were applied to the estimation of the association between usual alcohol intake and all-cause mortality using data from the EPIC-Norfolk study.

### **7.2 EPIC-NORFOLK COHORT**

The EPIC-Norfolk study recruited 25,636 eligible individuals, aged 39 to 79 years, between 1993 and 1998 from the population of Norfolk, UK as a part of the European Prospective Investigation into Cancer and Nutrition [70]. Alcohol intake was recorded using 7-day diet diaries (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 [68,70,71]. I focused 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 [25,139,140]. Ethnicity was not included as a potential confounder as less than

*Table 7.1 Descriptive statistics for the covariates in the EPIC-Norfolk population.*

<b>Covariates (C)</b>	<b>#</b>	<b>% of Total</b>		
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%		
	<b>Median</b>	<b>Mean</b>	<b>IQR</b>	<b>Range</b>
Mean alcohol intake FFQ (g/day; 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 <sup>2</sup> )	26	26	24 – 28	15 – 59

0.5% of the EPIC-Norfolk population were non-white. The mean of 1-2 recorded FFQ values (g/day) and the baseline body mass index (BMI; kg/m<sup>2</sup>) were used as predictive covariates in the measurement error model in addition to the potential confounders from the outcome model.

Those who had no FFQ recorded, had missing baseline values (baseline 7DD, outcome model covariates, or predictive covariates), or recorded less than seven days in their baseline 7DD were excluded totaling 3,588 individuals. A further 2,570 individuals were excluded as they reported prior serious illness, including cancer, diabetes, stroke or heart attack, which may change alcohol intake behavior (i.e. abstinence or reduced drinking) while representing a higher risk of mortality [139,141]. Analyses are based on 19,478 individuals. In the first 15 years of follow-up there were a total of 3,085 deaths.

Table 7.1 displays a summary of the covariates used to predict the usual alcohol intake and the potential confounders in the outcome regression model. There was a higher proportion of women than men across an age spectrum from 39 to 79 at baseline in the EPIC-Norfolk population. Only 11% were current smokers while 41% were previous smokers. The average BMI was 26, but the range of 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



*Table 7.2 Descriptive statistics for the error-prone exposures in the EPIC-Norfolk population 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 naïve, MCMC-RC, and INLA-RC methods.*

<b>Error-prone measures &amp; estimates of usual alcohol intake</b>	<b>No. observed (% of total)</b>	<b>No. not consuming alcohol in 7DD (% of those observed)</b>	<b>Median (g/day)</b>	<b>Mean (g/day)</b>	<b>IQR (g/day)</b>	<b>Range (g/day)</b>
<b>Consumers Only</b>						
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
<b>ALL</b>						
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
MCMC fitted $E(\tilde{X} \mathbf{R}, \mathbf{C}, \boldsymbol{\theta})$	19,478	NA	6.1	11.2	1.3 – 15.9	0.9 – 142.8
INLA fitted $E(\tilde{X} \mathbf{R}, \mathbf{C}, \boldsymbol{\theta})$	19,478	NA	5.9	11.2	1.3 – 15.9	1.0 – 147.0

recall time of the FFQ, the percentage of alcohol consumers recorded by this instrument was higher than that recorded in the 7DDs; however, the mean and median usual alcohol intake determined by the FFQs were 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 7.2 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 g/day and 4.6 g/day, respectively). These observations were preserved when exclusively describing consumers of alcohol.

### 7.3 THE EPISODIC CONSUMERS MODEL

Measures of alcohol intake from short term assessments such as the 7DD are known to be subject to excess zero measurements due to episodic consumers as well as to other sources of measurement error. I therefore used a measurement error model that accounts for episodic consumers. The EC model [17–19,59] extends the classical model to accommodate excess zero measurements in addition to classical measurement error. The model for  $R_{ij}$ , where  $R_{ij}$  is the  $j^{\text{th}}$  measurement for the  $i^{\text{th}}$  person, 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. The first component is modelled 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,  $\mathbf{u}_i = (u_{1i}, u_{2i})$ :

$$7.1 \quad P(R_{ij} > 0 | \mathbf{u}_i, \mathbf{C}_i) = H(\gamma_{1,0} + \boldsymbol{\gamma}_{1,C} \mathbf{C}_i + u_{1i})$$

$$7.2 \quad R_{ij} | (R_{ij} > 0), \mathbf{u}_i, \mathbf{C}_i \sim N(\gamma_{2,0} + \boldsymbol{\gamma}_{2,C} \mathbf{C}_i + u_{2i}, \sigma_\varepsilon^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where  $H(\cdot) = \exp(\cdot) / (1 + \exp(\cdot))$ ,  $\boldsymbol{\gamma}_1$  and  $\boldsymbol{\gamma}_2$  are vectors of regression coefficients,  $\varepsilon_{ij}$  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 7.1 and 7.2, but typically they are and I assumed that here.  $\mathbf{C}_i$  generally includes covariates in the outcome regression model as well as additional covariates related to  $X_i$  given  $\mathbf{R}_i$  [17]. It is assumed that the measurement errors,  $\varepsilon_{ij}$ , have constant variance,  $\sigma_\varepsilon^2$ , 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. The distribution of  $\mathbf{u}_i$  is therefore modeled as a bivariate normal distribution with variances  $\sigma_{u1}^2$  and  $\sigma_{u2}^2$  and correlation  $\rho$ :

$$7.3 \quad \mathbf{u}_i \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u1}^2 & \rho\sigma_{u1}\sigma_{u2} \\ \rho\sigma_{u1}\sigma_{u2} & \sigma_{u2}^2 \end{pmatrix} \right\}$$

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:

$$7.4 \quad X_i = E(R_{ij} | \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}) = P(R_{ij} > 0 | \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}) E(R_{ij} | R_{ij} > 0, \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta})$$

where  $\boldsymbol{\theta} = \{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \sigma_{u1}^2, \sigma_{u2}^2, \rho, \sigma_\varepsilon^2\}$  is the set of all model parameters. The EC model as shown in Equations 7.1 and 7.2 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 [142],  $g(x, \lambda) = (x^\lambda - 1) / \lambda$ , where  $\lambda = 0$  denotes the log transformation [17,18,64]. 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

$$7.5 \quad X_i = E(R_{ij} | \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}) = P(R_{ij} > 0 | \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}) E\{g^{-1}(R_{ij}^*, \hat{\lambda}) | R_{ij} > 0, \mathbf{u}_i, \mathbf{C}_i; \boldsymbol{\theta}\}$$

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 the model in Equation 7.4 or

7.5 are  $\hat{X}_i(\boldsymbol{\theta}) = E(X_i | \mathbf{R}_i, \mathbf{C}_i, \boldsymbol{\theta})$  and these form the basis of the Bayesian RC measurement error correction methods.

Kipnis et al. and Keogh and White used maximum likelihood estimation to estimate the parameters of the EC model and hence to obtain  $\hat{X}_i(\hat{\boldsymbol{\theta}})$  [17,18]. 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 [52] due to lack of empirical identifiability in a finite sample set. The maximum likelihood estimation of  $\hat{X}_i(\hat{\boldsymbol{\theta}})$  failed to converge for the EPIC-Norfolk data set in this application.

The EC model can alternatively be fitted using a Bayesian approach [19,52]. I used JAGS [62] via R package ‘rjags’ [95] to implement the MCMC sampling from the EC model, and I used the INLA software (available at [www.r-inla.org](http://www.r-inla.org)) accessed through an R package for the INLA implementation of the EC model. Code for each implementation of the model can be found in Appendix D. Previous Bayesian implementations of the EC model have used MATLAB (MATLAB and Statistics Toolbox, The MathWorks, Inc) [19,52].

#### 7.4 FITTING THE EC MODEL

The EC model was fitted to the 7DD measures of alcohol intake, using from one 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 [142]. The best transformation parameter,  $\hat{\lambda}$ , for the Box-Cox transformation was found by finding the value of  $\lambda$  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  $\hat{\lambda}_{BMI} = 0$ ) as well as being centered and scaled in order to improve efficiency in the MCMC sampler [85]. I used the same priors as used by Zhang et al and Bhadra et al [19,52]; further detail can be found in the appendices to those papers. All  $\gamma$  coefficients in the EC model were given normal priors with mean zero and variance 100. Due to the scaling of  $R_{ij}$ , the variance  $\sigma_{\epsilon}^2$  was able to be given a uniform prior between 0 and 3 in the MCMC-RC implementation. This differs from the inverse Gamma prior used for the variances in the simulation studies in previous

chapters. The inverse gamma prior is commonly used for variances as it allows for an infinite positive range of values unlike the uniform prior, hence that was selected in simulation studies where the focus was not on the EC model. For the INLA-RC implementation  $\sigma_{\varepsilon}^2$  was given an inverse Gamma prior,  $IG(1, 1)$ , as in previous chapters because INLA cannot easily accommodate the uniform prior.

The covariance matrix for the person-specific random effect,  $\mathbf{u}_i$ , was given an inverse Wishart prior, which is a multivariate generalization of the scaled  $\chi^2$ -distribution and is conjugate to the multivariate normal distribution [143]. The inverse Wishart prior is parameterized by the degrees of freedom,  $df$ , and an expectation matrix. Specifying the  $df$  as greater 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, 5 degrees of freedom (for a matrix of order 2) was specified, and for the expectation matrix, an expectation of the variances on the diagonal equal to 1 and the correlation equal to 0.5 was specified [19,52].

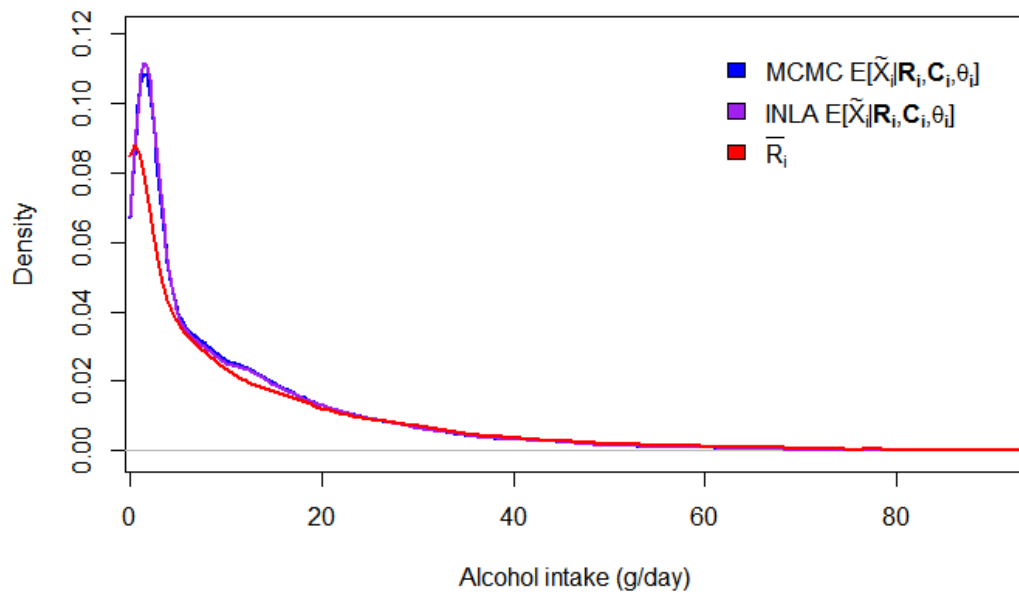
Approximately 250,000 burn-in samples were required before the MCMC chains were believed to have reached the stationary distribution as determined by the Geweke diagnostic [144]. An additional 500,000 samples were collected after burn-in on each of three chains for inference, keeping only every 200th sample to reduce burden on computer memory.

The posterior expected values of the EC model parameters after fitting with MCMC and with INLA can be found in Table 7.3. The estimated posterior means were very similar between the two methods. The regression coefficients for the logistic model describing the probability of alcohol consumption ( $\boldsymbol{\gamma}_1$ ) had more variation between the two methods than did the regression coefficients for the linear regression model describing the amount of alcohol consumed ( $\boldsymbol{\gamma}_2$ ). This is not unexpected given that INLA is approximate for logistic regression. Additionally, the MCMC estimated variance of the measurement error ( $\sigma_{\varepsilon}^2$ ) was observed to be more than twice as high as the INLA estimated  $\sigma_{\varepsilon}^2$  where a different prior was specified. Given the MCMC estimation, the  $\sigma_{\varepsilon}^2$  was approximately 50% the variance of the estimated exposure (when Box-Cox transformed and scaled); whereas given the INLA estimation, the  $\sigma_{\varepsilon}^2$  was approximately 20% the variance of the estimated exposure. However, the resulting estimated densities of  $\hat{X}_i(\boldsymbol{\theta})$  were very similar (Figure 7.1; Table 7.2) and distinct from the naïve distribution constructed from the mean of observed measures of alcohol intake, denoted  $\bar{R}_i$ .

The peak of the distribution of  $\hat{X}_i(\hat{\boldsymbol{\theta}})$  estimated by either Bayesian method was observed to be higher than for  $\bar{R}_i$ , which is expected because individuals with low usual consumption are likely

Table 7.3 The fitted EC model using either MCMC sampling or INLA approximation. Posterior means and standard deviation (SD) of each regression coefficient and model parameter. “ref” indicates reference category. \*\*\* indicates p-value < 0.001; \*\* p-value < 0.01; \* p-value < 0.05.

		MCMC EC model				INLA EC model			
		Logistic regression for probability of consumption		Linear regression for amount consumed		Logistic regression for probability of consumption		Linear regression for amount consumed	
<b>Intercept</b>		$\gamma_{1,0}$	3.598 (0.209)	$\gamma_{2,0}$	1.166 (0.025)	$\gamma_{1,0}$	3.187 (0.199)	$\gamma_{2,0}$	1.159 (0.026)
<b>Age (Age<sup>3</sup> for the outcome models)</b>		$\gamma_{1,3}$	-0.237 (0.028)	$\gamma_{2,3}$	-0.043 (0.004)	$\gamma_{1,3}$	-0.200 (0.033)	$\gamma_{2,3}$	-0.042 (0.004)
<b>Sex</b>	Male	ref		ref		ref		ref	
	Female	$\gamma_{1,4}$	0.077 (0.075)	$\gamma_{2,4}$	-0.051 (0.009)	$\gamma_{1,4}$	0.023 (0.069)	$\gamma_{2,4}$	-0.054 (0.009)
<b>Smoking status</b>	Current smoker	ref		ref		ref		ref	
	Former smoker	$\gamma_{1,5}$	0.152 (0.122)	$\gamma_{2,5}$	-0.064 (0.014)	$\gamma_{1,5}$	0.0134 (0.112)	$\gamma_{2,5}$	-0.065 (0.014)
	Never smoker	$\gamma_{1,6}$	-0.123 (0.102)	$\gamma_{2,6}$	-0.133 (0.014)	$\gamma_{1,6}$	-0.090 (0.107)	$\gamma_{2,6}$	-0.132 (0.014)
<b>Education</b>	No qualifications	ref		ref		ref		ref	
	O-levels	$\gamma_{1,7}$	0.202 (0.099)	$\gamma_{2,7}$	0.017 (0.014)	$\gamma_{1,7}$	0.250 (0.111)	$\gamma_{2,7}$	0.021 (0.015)
	A-levels	$\gamma_{1,8}$	0.314 (0.089)	$\gamma_{2,8}$	0.036 (0.010)	$\gamma_{1,8}$	0.364 (0.074)	$\gamma_{2,8}$	0.039 (0.010)
	Bachelors degree or higher	$\gamma_{1,9}$	0.283 (0.150)	$\gamma_{2,9}$	0.048 (0.016)	$\gamma_{1,9}$	0.284 (0.124)	$\gamma_{2,9}$	0.051 (0.015)
<b>Social class</b>	Non-skilled work	ref		ref		ref		ref	
	Semi-skilled work	$\gamma_{1,10}$	0.116 (0.170)	$\gamma_{2,10}$	-0.001 (0.025)	$\gamma_{1,10}$	0.228 (0.166)	$\gamma_{2,10}$	0.006 (0.025)
	Skilled work (manual or non-manual)	$\gamma_{1,11}$	0.283 (0.148)	$\gamma_{2,11}$	0.003 (0.023)	$\gamma_{1,11}$	0.348 (0.155)	$\gamma_{2,11}$	0.008 (0.024)
	Manager	$\gamma_{1,12}$	0.596 (0.134)	$\gamma_{2,12}$	0.074 (0.023)	$\gamma_{1,12}$	0.617 (0.161)	$\gamma_{2,12}$	0.079 (0.024)
	Professional	$\gamma_{1,13}$	0.541 (0.149)	$\gamma_{2,13}$	0.057 (0.027)	$\gamma_{1,13}$	0.512 (0.206)	$\gamma_{2,13}$	0.060 (0.029)
<b>Mean FFQ</b>		$\gamma_{1,1}$	4.499 (0.146)	$\gamma_{2,1}$	0.873 (0.005)	$\gamma_{1,1}$	4.152 (0.109)	$\gamma_{2,1}$	0.873 (0.005)
<b>Baseline BMI</b>		$\gamma_{1,2}$	-0.048 (0.032)	$\gamma_{2,2}$	0.013 (0.004)	$\gamma_{1,2}$	-0.026 (0.030)	$\gamma_{2,2}$	0.014 (0.004)
<b>Correlation between random effects</b>		$\rho$	0.937 (0.005)			$\rho$	0.968 (0.003)		
<b>Precision (<math>\tau_{U_1}^2</math>) or variance (<math>\sigma_{U_1}^2</math>) of random effect in probability of consumption</b>		$\tau_{U_1}^2$	0.135	$\sigma_{U_1}^2$	7.426 (0.521)	$\tau_{U_1}^2$	0.176 (0.012)	$\sigma_{U_1}^2$	5.692
<b>Precision (<math>\tau_{U_2}^2</math>) or variance (<math>\sigma_{U_2}^2</math>) of random effect in amount of consumption</b>		$\tau_{U_2}^2$	5.71	$\sigma_{U_2}^2$	0.175 (0.521)	$\tau_{U_2}^2$	5.580 (0.121)	$\sigma_{U_2}^2$	0.179
<b>Precision (<math>\tau_{\epsilon}^2</math>) or variance (<math>\sigma_{\epsilon}^2</math>) of measurement error variance</b>		$\tau_{\epsilon}^2$	2.29	$\sigma_{\epsilon}^2$	0.435 (0.003)	$\tau_{\epsilon}^2$	5.424 (0.084)	$\sigma_{\epsilon}^2$	0.184



*Figure 7.1 Kernel density plot of empirical distribution of usual alcohol intake across individuals in the EPIC-Norfolk study based on the naïve estimation,  $\bar{R}_i$ , and the expectation of  $X$  estimated,  $E(\tilde{X}_i | \mathbf{R}_i, \mathbf{C}_i, \boldsymbol{\theta})$ , from the MCMC samples of the EC model and INLA samples of the EC model.*

to have reported zero consumption in the 7DD. From MCMC, the minimum  $\hat{X}_i(\hat{\boldsymbol{\theta}})$  for any individual was 0.9 g/day while the maximum was 142.8 g/day (Table 7.2), and from INLA the range was 1.0g/day to 147.0 g/day. Due to the minimum, a product of assuming all members of our population have some probability of drinking, inference can be made down to only approximately 1g/day.

As would be expected, an increase in the mean FFQ 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 were 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 were both associated with a higher probability of drinking and greater amounts consumed.

## 7.5 RESULTS: APPLICATION

Two analyses were performed. 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, I adjusted for additional potential confounders, smoking status, social class, and educational attainment (the “fully adjusted outcome model”). The naïve analysis was performed using  $\bar{R}_i$ , the mean of

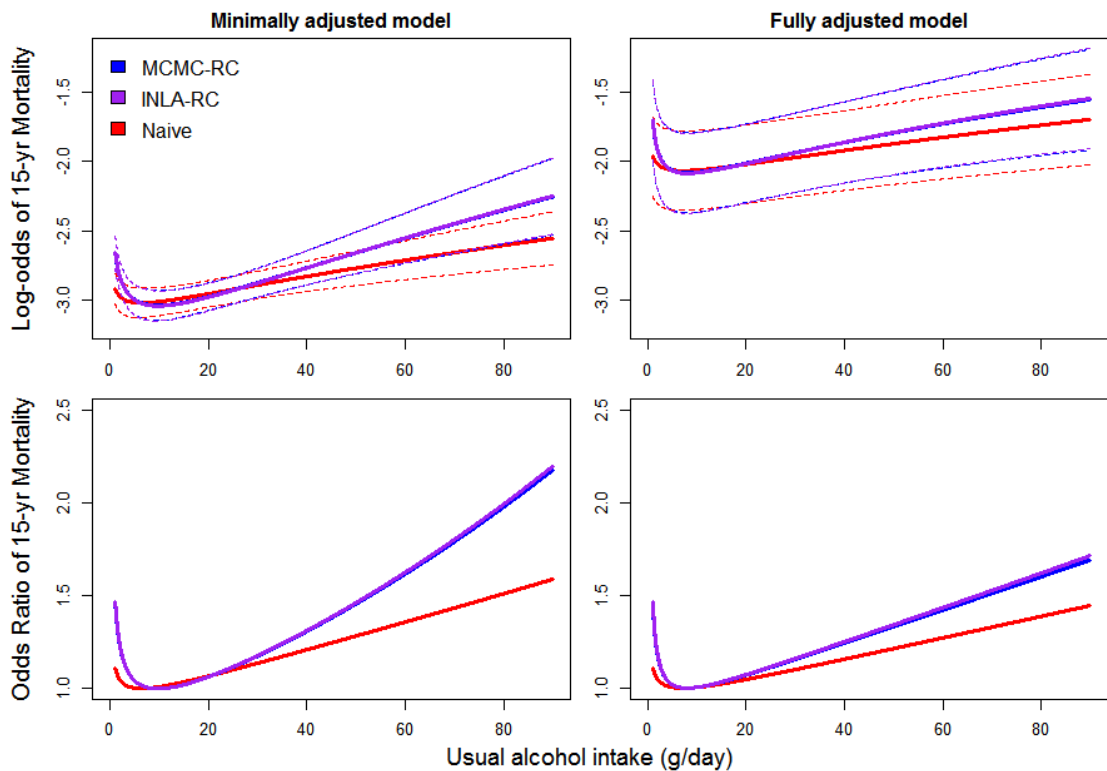
**Table 7.4 Results on the association between alcohol intake and all-cause mortality in EPIC-Norfolk.** Best-fitting curves: the p-value comparing the deviance of the best-fitting FP1 model to the best-fitting FP2 model, the values of  $\hat{\mathbf{p}}$  defining the chosen transformations, the coefficients 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
<b>Naïve model,</b> $\bar{R}$	Minimally adjusted	0.0015	0	0.5	-4.47	-0.16	0.13	6.3	1.11
	Fully Adjusted	0.0214	0	0.5	-3.51	-0.14	0.11	7.1	1.11
<b>MCMC-RC</b>	Minimally adjusted	0.0002	0	0.5	-4.35	-0.42	0.27	9.7	1.46
	Fully Adjusted	0.0090	-0.5	0.5	-4.15	0.90	0.11	8.0	1.45
<b>INLA-RC</b>	Minimally adjusted	0.0002	0	0.5	-4.35	-0.42	0.27	9.7	1.47
	Fully Adjusted	0.0083	-0.5	0.5	-4.17	0.92	0.12	8.0	1.47

measures  $R_i$ , in the fractional polynomial method in place of  $X$ . MCMC-RC and INLA-RC were applied as described in Chapter 6 using the sampled values of the exposure, denoted  $\tilde{X}_i$ .

For the naïve analysis, the fractional polynomial method was performed using the ‘mfp’ package available in R [133,145]. The naïve measure,  $\bar{R}_i$ , included zeros and the fractional polynomial method cannot directly accommodate non-positive data points; therefore, the method routinely adds the constant 0.1 to every individual’s measure in order to make all points positive [8]. 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 was not an issue in the MCMC-RC and INLA-RC analyses, because all of the posterior sampled values  $\tilde{X}_i$  are positive.

All methods found that the best-fitting FP1 curve had power  $\hat{p} = -2$  for both the minimally adjusted and fully adjusted models. These best-fitting FP1 curves provided a better fit than both the linear and null models (p-value  $< 10^{-8}$ ). All methods also found that the best-fitting FP2 model provided a significantly better fit than the best-fitting FP1 model, but they differed on the powers selected in the FP2 model (Table 7.4). For both the minimally adjusted and fully adjusted models, the naïve method selected  $\hat{\mathbf{p}} = \{0, 0.5\}$ . MCMC-RC and INLA-RC selected  $\hat{\mathbf{p}} = \{0, 0.5\}$  for the minimally adjusted and  $\hat{\mathbf{p}} = \{-0.5, 0.5\}$  for the fully adjusted curve. The resulting curves are shown in Figure 7.2. All three curves had in common the characteristic J-shape. The MCMC-RC and INLA-RC curves were nearly identical. The 95% confidence bounds associated with the selected model fit by each method also appear in Figure 7.2. These confidence bounds do not incorporate the uncertainty due to measurement error or model uncertainty.



**Figure 7.2** 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 odds ratio (OR) compared with 8 g/day (1 unit of alcohol).

The naïve curve was the flatter across the range of alcohol intake (Figure 7.2). This was what is expected based on the simulation study results. The higher risk associated with lower alcohol consumption was tempered in the naïve model compared to the MCMC-RC and INLA-RC measurement error corrected curves. The higher risk of mortality associated with higher alcohol intake was also modest compared to the corrected curves. The nadir at which the risk is lowest was 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 and INLA-RC curves were 9.7 g/day and 8.0 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 (Table 7.4).

There were no great differences for most of the estimated coefficients for the covariates between the naïve and corrected models (Table 7.5). MCMC-RC and INLA-RC have virtually identical coefficients. The intercept was higher for the naïve minimally adjusted model than the corrected minimally adjusted models whereas the intercept was lower for the naïve fully adjusted model than the corrected fully adjusted models. Being female had a slightly more minimal effect on mortality in both of the naïve models compared to the corrected models.



*Table 7.5 Regression coefficients for the non-error prone covariates in the selected minimally adjusted and fully adjusted outcome models fitted using the naïve analysis, MCMC-RC, or INLA-RC method. “ref” indicates reference category. \*\*\* indicates p-value < 0.001; \*\* p-value < 0.01; \* p-value < 0.05.*

			Minimally adjusted outcome model			Fully adjusted outcome model		
			Naïve	MCMC-RC	INLA-RC	Naïve	MCMC-RC	INLA-RC
<b>Intercept</b>		$\beta_0$	-4.837	-4.346	-4.347	-3.838	-4.152	-4.173
<b>Age (Age<sup>3</sup> for the outcome models)</b>		$\beta_{Z,1}$	0.114***	0.114***	0.114***	0.116***	0.115***	0.115***
<b>Sex</b>	Male	ref						
	Female	$\beta_{Z,2}$	-0.570***	-0.582***	-0.582***	-0.527***	-0.538***	-0.539***
<b>Smoking status</b>	Current smoker	ref						
	Former smoker	$\beta_{Z,3}$				-0.706***	-0.706***	-0.706***
	Never smoker	$\beta_{Z,4}$				-0.961***	-0.969***	-0.970***
<b>Education</b>	No qualifications	ref						
	O-levels	$\beta_{Z,5}$				-0.010	-0.007	-0.007
	A-levels	$\beta_{Z,6}$				-0.070	-0.069	-0.069
	Bachelors degree or higher	$\beta_{Z,7}$				-0.136	-0.133	-0.133
<b>Social class</b>	Non-skilled work	ref						
	Semi-skilled work	$\beta_{Z,8}$				-0.254*	-0.254*	-0.254*
	Skilled work (manual or non-manual)	$\beta_{Z,9}$				-0.181	-0.180	-0.179
	Manager	$\beta_{Z,10}$				-0.234*	-0.234*	-0.234*
	Professional	$\beta_{Z,11}$				-0.491***	-0.487**	-0.486**

## 7.6 SUMMARY

This work was motivated by an application to investigate the association between alcohol intake and all-cause mortality in the EPIC-Norfolk cohort in which 7-day diary measures of alcohol intake are subject to measurement error due to episodic consumers. This association is prone to a complex form of measurement error captured in the EC model and is believed to be non-linear in nature. In the EPIC-Norfolk cohort, only replicate measures are available to aid in estimation of the measurement error model parameters. To correct for measurement error in this context, I proposed the Bayesian RC method in Chapter 4 and extended it for use with the fractional polynomial method in Chapter 6. The application of this correction method resulted in a stronger and more non-linear association between alcohol intake and all-cause mortality than the naïve approach.

The minimum risk, of particular public health importance, was found in this chapter to be similar to that found in other studies. 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 and INLA-RC models. Di Castelnuovo found that the range of 1-2 drinks per day in women and 2-4 drinks per day in men had an inverse association with all-cause

mortality [24]. 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 [146]. The J-shaped association between alcohol and all-cause mortality remains a topic of much debate [138]. I adjusted for a number of potential confounders in the analysis, including socioeconomic indicators; however, 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.

This analysis was limited in the scope of inference that can be made from the estimated association. 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, I also cannot make inference about non-drinkers.

Another limitation 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 for this application.

An extension to the EC model has been developed to include non-drinkers [18,52]. The EC model as presented in this chapter 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 recorded 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 [147].

Bayesian RC has potential due to its flexibility. However, given that the uncertainty in estimation of the measurement error and exposure parameters is not included in the SEs for the substantive model, it may be considered a partial solution. A more accurate variance could in theory be generated by bootstrapping with INLA-RC. For MCMC-RC, such bootstrapping would be unwieldy. 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 chapter, the model selection and the binary outcome model are the greatest sources of uncertainty. A fully Bayesian model which combines both measurement error correction and model selection will be poorly identifiable in the settings outlined in this chapter; areas for further research in this regard will be outlined in the next chapter.

## 8 DISCUSSION

---

### 8.1 AIMS & OVERVIEW

In this thesis, I proposed to explore the use of Bayesian methods, including MI and INLA, to address measurement error. It was hypothesized that Bayesian methods may have an advantage in this area due to the “modularity” of Bayesian analysis and the ease with which complex measurement error may be accommodated. Additionally, the use of sensible prior information in Bayesian analysis can be used to improve power to detect non-linearity. Bayesian methods for measurement error correction have not previously been compared extensively, particularly for non-linear associations. Furthermore, there have been no publications to date which explore Bayesian methods for measurement error correction in the context of polynomial model selection.

The aims of this thesis were:

- 1) To investigate the performance of the Bayesian measurement error correction methods in exposure-outcome relationships where the error-prone exposure is untransformed using simulation studies;
- 2) To extend and adapt, as necessary, promising correction methods for use in a outcome regression model which incorporates the addition of a squared term for the error-prone exposure, and to assess the performance of model comparison tests between this model and the linear model;
- 3) To extend and evaluate promising methods of measurement error correction for use when the exposure is transformed as in the fractional polynomial method, and to evaluate each method in terms of curve fit, power, and validity of hypothesis testing;
- 4) To apply the most promising methods to the estimation of the non-linear association between usual alcohol intake and all-cause mortality using data from the EPIC-Norfolk cohort where replicate measures are available in a subset of the study population and the outcome is binary.

In order to address these aims, I have presented a version of MI for measurement error correction extending the previously published SMCFCFS method for missing data suitable for use with a validation study or a replicate study [46]. I have presented the use of INLA for measurement error correction with a replicate study, as previously published by Muff et al [29], and extended it for use with a validation study. I have carried out an extensive simulation study comparing these methods to the use of RC and Bayesian analysis using MCMC in Chapter 3 where the error-prone exposure is untransformed in either a linear or logistic regression model. INLA has not previously been compared with alternative methods in simulations. While MI has previously been suggested for measurement error correction, earlier work has focused on simple settings and not on substantive-model-compatible imputation models.

In order to assess the use of Bayesian analysis for measurement error correction in the context of model selection, I started with the simple example of selecting between a linear model and a quadratic model. The problem was approached using three strategies: 1) performing a full Bayesian analysis for each specified substantive model (linear and quadratic) and determining the best fit using an “information criterion”; 2) use of Bayesian analysis for the prediction of the expectations of the untransformed and transformed error-prone exposure for use with RC; and 3) a unified Bayesian analysis wherein a joint model incorporates both measurement error correction and model selection and fit. For strategies 2 and 3, I proposed and assessed two novel approaches – Bayesian RC and BTS with measurement error correction, respectively. These were then extended as appropriate for use with the fractional polynomial method for model selection.

Within each of these strategies, the methods may be further refined and compared. For example, alternative methods of BTS available in the literature might be employed in the joint model with measurement error correction (Section 8.6). However, I believe that in selecting one or two examples of each strategy and comparing them to one another, the larger strengths and weaknesses of each approach in various scenarios have been assessed.

## **8.2 SUMMARY OF RESULTS**

### **8.2.1 Reference method: regression calibration**

RC is straightforward to implement and performs well in many circumstances. However, the method requires an approximation in non-linear outcome models, such as logistic regression. In Chapter 3 it was observed that for larger ORs ( $OR = 4.1$ ) this approximation results in biased estimates and undercoverage in the 95% CIs when using the delta method. This is consistent with other publications which found ORs over 3.0 to be biased [20]. Moreover, even for settings where RC is unbiased, the effect estimates and curve fits after application of RC are more variable than for the other methods as demonstrated by the high RMSE and MISE estimates as well as the plots of the curve fits (Chapter 3-6). In model selection between a quadratic and linear model, RC recovers none of the power lost to measurement error (Chapter 5). However, some power was gained within the fractional polynomial method when the data were generated according to the model assumed by the extension of RC (Chapter 6). When the measurement error model was not that assumed by this extension of RC, the resulting curve fit is an improvement over the naïve analysis but type I error is inflated.

There are many scenarios to which RC cannot be easily adapted. In Chapter 6, it was highlighted that even the classical measurement error model poses an analytical challenge for estimation of the maximum likelihood estimates required for RC in the context of the fractional polynomial method. Specification of an alternative error model was required. In the applications in Chapter

7, I used the EC model which has previously been published for use with RC when the error-prone exposure is untransformed in the substantive model [17,19,59,64,65]. It has been previously noted that this model does not converge for all data sets using numerical integration [52]. I found this to be the case for the EPIC-Norfolk data set. In order to use RC with the EC model with the fractional polynomial method, an additional 8 or 16 conditional expectations have to be obtained. Closed-form solutions for these conditional expectations may not exist.

Despite the above drawbacks to the method, RC can be recommended for use in diverse settings. As was shown in Chapters 3-5, assuming classical measurement error variance and only  $\frac{1}{4}$  the variance of  $X$ , all methods resulted in curve fits which reduce bias for either linear or logistic regression. The ease of use of RC makes widespread implementation by researchers more likely in simple additive error settings.

### **8.2.2 Use of multiple imputation for measurement error correction**

Use of MI for measurement error correction alongside the already routine use of MI for missing data is appealing as it increases the likelihood of adoption by researchers. When a validation study is available, an exposure subject to additive measurement error can be treated as if it were missing data using  $W$  as an auxiliary variable. In Chapter 3, I demonstrated the use of MI in this context in simulation studies using the SMC-FCS method of MI. Use of MI successfully corrected for bias from measurement error and resulted in efficiency gains over RC due to the incorporation of the information from the outcome in the imputation via rejection sampling. Because the priors used in this implementation of MI are flat, no additional gains in efficiency are derived.

The use of SMC-FCS for correction of measurement error was demonstrated for an untransformed functional form of the error-prone exposure in the substantive model (Chapter 3), but the principle should work well for any specified functional form of the error-prone exposure as described by Bartlett et al [46]. SMC-FCS has the additional advantage over RC of not needing to rely on approximations when the outcome model is non-linear, e.g. logistic regression.

A modified SMC-FCS method was also successful in correcting bias from measurement error in the setting of a replicate study provided there was sufficient information in the likelihood, i.e. the likelihood was empirically identifiable. In this case, some added efficiency was derived from the use of proper priors for the variance parameters. In scenarios with small sample sizes and high measurement error the use of flat priors for the regression coefficients resulted in uninformative posteriors, i.e. extreme variability in the estimates of  $\hat{\beta}_X$ . While, in theory, proper priors could be introduced with some effort for all parameters of the substantive model, Carpenter and Kenward pointed out that in practice MI relies on a “short cut” to obtain the mean and variance of the substantive model regression parameters without which there is little to be gained over a fully Bayesian analysis [98].

MI does not necessarily extend easily to accommodate more complex error models. Even in the case of a validation study, a specific extension may be required to accommodate multiplicative error such as the lognormal error model (Section 6.4.1). The EC model presented in Chapter 7, for example, relies on numerical integration for maximum likelihood estimates which would need to be integrated within the MI algorithm.

Further work has been done in the field of missing data to explore the use of MI with the fractional polynomial method [148]. It is likely that these methods could also be applied to accommodate additive measurement error when a validation study is available by treating the error-prone exposure as missing.

### **8.2.3 Use of INLA for measurement error correction**

INLA promises the advantages of Bayesian analysis, such as the inclusion of prior information, but at a fraction of the computation time. In Chapter 3, I showed that for an error-prone exposure that is untransformed in the substantive model, INLA resulted in unbiased estimates of the regression coefficient  $\beta_x$  in most scenarios, excepting those with both very large measurement error variance and large effect size. There was however, moderate to severe undercoverage of 95% CrIs in many scenarios. Use of bootstrapping to estimate 95% CIs resulted in similar undercoverage. INLA relies on fast approximations to estimate the posterior probability of the parameters of the model. For larger measurement error or when there is a small validation study, these approximations may not be sufficient for good coverage and correction of bias.

INLA was much more efficient than RC and MI in most scenarios. This is largely due to the inclusion of prior information even when such prior information is very weakly informative.

Importantly, INLA cannot at present accommodate a transformation of a latent variable. Therefore, this method was not used in subsequent chapters.

### **8.2.4 Bayesian regression calibration**

The principle of RC remains true whether or not the expectations of the latent error-prone exposure are estimated via maximum likelihood estimation or a Bayesian posterior distribution. In Chapter 4, I proposed the Bayesian RC approach, in which the expectations required for implementing RC are obtained from posterior samples estimated by either MCMC or INLA.

Use of Bayesian analysis to estimate the expectations of the latent error-prone exposure has several important advantages. Firstly, the extension to estimation of transformations of the latent error-prone exposure is straightforward whereas it is a substantial undertaking in most settings for maximum likelihood estimation. Secondly, the use of prior information for the parameters of the measurement error and exposure models results in improved efficiency in estimation of the regression coefficients in the substantive model and improved power in selection of non-linear

models. Assuming a correctly specified model (measurement error and exposure models), the use of prior information also results in type I error rates below the nominal level. Conversely, it would be expected that a misspecified model could lead to inflated type I error rates. However, in Chapter 6, when MCMC-RC and INLA-RC were applied with the measurement error model misspecified (classical measurement error was specified while the data were generated under the lognormal error model), type I error was still substantially lower than for the naïve analysis.

Bayesian RC requires the same approximations for non-linear outcome models as frequentist RC. In Chapter 5, bias was observed in the regression coefficient estimates after application of each of the three RC-type methods (RC, MCMC-RC, and INLA-RC) when the substantive model was logistic regression.

The 95% CIs estimated by MCMC-RC and INLA-RC as implemented for the simulation studies in this thesis do not include uncertainty due to the estimation of the parameters of the measurement error and exposure models. Nor do they include the uncertainty from model selection. As a result, undercoverage is expected in many scenarios. This was observed for the quadratic model fit of the J-shaped association with a continuous outcome in Chapter 4. The undercoverage was most severe when only a replicate study was available. When the outcome was binary and a validation study was available, there did not appear to be undercoverage (within the limitations of the smaller simulation study); it is anticipated that the sampling variability inherent in the parameter estimates from the logistic regression will dwarf the uncertainty due to measurement error.

The use of INLA for the Bayesian estimation in Bayesian RC (INLA-RC) resulted in slightly better average curve fits in some of the simulated scenarios compared to the use of MCMC sampling for estimation (MCMC-RC). The INLA-RC method is exact for classical measurement error and a normally distributed  $X|Z$  and does not require consideration of burn-in, convergence, or autocorrelation as does MCMC-RC. Additionally, as demonstrated in a sensitivity analysis in Chapter 4, bootstrapping can be performed with INLA-RC to obtain 95% CIs with better coverage properties. Model selection could also be incorporated into bootstrap samples in the context of the fractional polynomial method.

The number of samples drawn from the approximated posterior distribution within INLA-RC was set at 3,000 somewhat arbitrarily. This number was just under the lowest ESS observed for any of the MCMC applications in this thesis. As the marginal posterior of  $X$  for these simulation studies should be a simple normal distribution and the Bayesian RC method requires only accurate estimation of the expectation (and not the variance), 3,000 samples were sufficient. Where more complex posterior distributions are anticipated, such as in the application in Chapter 7 where

10,000 samples were drawn, a sufficient number of samples should be drawn to represent the posterior.

### **8.2.5 Use of Bayesian MCMC modelling for a known functional form of the exposure**

In Chapters 3 and 4, I demonstrated that when the functional form of the error-prone exposure in the substantive model is specified, Bayesian MCMC can be reliably used to correct for bias from measurement error with good coverage properties, in agreement with the literature [1,37]. Unlike INLA, implementation of the Bayesian model using MCMC does not rely on any approximations.

A fully Bayesian model benefits from the prior information in order to recover power lost to measurement error. When very weak prior distributions, i.e.  $\beta_X \sim N(0, 1 \times 10^6)$ , were specified and the likelihood was empirically unidentifiable, the posterior distribution was seen to be less informative, i.e. less efficient (Chapter 3), though the nominal coverage was maintained. Such very weak priors result in slow convergence especially when used alongside logistic regression. Use of such weak prior distributions is often unnecessary except to obtain estimates with frequentist properties [37]. For regression coefficients, a Gaussian distribution with smaller variance will result in reasonable posterior values when the likelihood is empirically uninformative; for variances, scaling the variables can enable users to estimate realistic ranges for the prior information. Gelman et al [87] and Carroll et al [1] recommend the use of a uniform prior between 0 and 3. This uniform prior was used in the application of the EC model for the measurement error variance and is common to other publications of the EC model using Bayesian analysis [19,52].

### **8.2.6 Use of Bayesian MCMC modelling for an unknown functional form of the exposure**

In Chapter 5, I showed that use of hypothesis testing, similar to a Wald test, using the HPD 95% CrIs was very effective in improving power to detect the quadratic association while preserving nominal type I error rates. However, this method is limited because it cannot be applied when the models proposed are not nested. Therefore, the following alternative methods of model comparison were explored.

#### ***Model selection using DIC***

Model selection using DIC involves performing a Bayesian analysis specifying the functional form of the error-prone exposure in the substantive model for each potential functional form. When comparing the many potential models, such as those required for the fractional polynomial method, this approach is computationally quite intensive [120].

In Chapter 5, it was observed that the use of DIC was unreliable in recovering power lost to measurement error when selecting between a quadratic model and a linear model for a quadratic association. In scenarios with higher measurement error, the DIC had improved power to select



the quadratic model over the linear model for a weak quadratic association; however, for the lower measurement error, use of DIC with Bayesian analysis resulted in dramatically lower power than the naïve analysis. Furthermore, when only a replicate study was available, the use of DIC for model selection resulted in markedly inflated empirical type I error rates. While it is not expected that Bayesian methods maintain nominal frequencies for type I error, due to the introduction of prior information, overselection of the more complex model is not a desirable property. Without a sufficient proportion of the true exposure observed, the posterior marginal distribution of the latent exposure will be dominated by the proposed substantive model and the samples of the latent value will be selected primarily to fit that model; this phenomenon is commonly termed “feedback” [129,135].

### ***Model selection using Bayesian transformation selection (BTS)***

While the frequentist fractional polynomial method is rooted in systematic hypothesis testing, BTS relies on variable reduction via a prior that encourages sparsity. This method is a unified analysis which incorporates model selection into the joint model by transforming the problem into one of parameter estimation [120]. The specific approach I implemented is the incorporation of a “spike and slab” prior for each proposed regression term.

When the simpler Kuo and Mallick method [60], wherein the “spike” is a probability mass over zero, was applied in the context of selection of the quadratic model over the linear model while correcting for measurement error (Chapter 5), the expected gains in power from a Bayesian analysis were not observed. This was due in part to poor mixing.

In Chapter 6, a more flexible “spike and slab” prior, wherein the “spike” is drawn from an extremely narrow Gaussian distribution (SSVS) [124], was used for BTS in the context of the fractional polynomial method while correcting for measurement error. Improved mixing was observed. I presented an “FP1” implementation which has the capability to select any number of first degree fractional polynomial terms as well as an “FP2” implementation which includes all second degree fractional polynomial terms as well.

When a validation study was present, the use of BTS alongside measurement error correction within a Bayesian model resulted in substantial gains in power over the naïve analysis (Chapter 6). This was true for both the “FP1” implementation and the “FP2” implementation. However, there was greater variability in the model fit resulting in a much higher average MISE, particularly for the “FP2” implementation which incorporated greater flexibility in the model. In most cases, the average MISE was still less than the naïve analysis, but the incorporation of uncertainty due to model selection and measurement error lead to greater bias than seen in the other Bayesian methods and occasionally greater bias than RC. Furthermore, the 95% CrIs are very wide due to uncertainty in the estimation of the intercept.

When only a replicate study was present in simulations, the use of BTS resulted in uninformative posterior estimates of the regression coefficients (Chapter 6). Likely, with a smaller measurement error variance proportional to the variance of  $X$  the method would work as well in this setting as it did for the validation setting. However, given the measurement error equal to the variance of  $X$ , there was not enough information in the likelihood or strong enough prior information for an identifiable posterior in this setting.

In the literature, it has been noted that a joint model approach to measurement error correction and fitting of a substantive model is more sensitive to misspecification of either model [135,149]. In this thesis, I did not perform any simulations for this method with a misspecified error model.

### ***Bayesian model averaging***

After the application of BTS, the final model, i.e. the model powers and regression coefficient estimates, may be arrived at in a few ways. If the priors for the binary indicator variables ( $I_{X_k}$ ) have been specified to be informative, then the model with the highest marginal posterior probability may be selected as the final model. If the priors are not intended to favor any particular model, the strength of the evidence in favor of one model over another may be assessed using Bayes factors (BFs) [127] as demonstrated in Chapters 5 and 6.

An approach not used in this thesis is Bayesian model averaging. This method may be most appropriate when there are several similarly high probability models [150]. The posterior mean values of the regression coefficients of the substantive model ( $\beta_j$ ) are weighted by the posterior probability of that model. These weighted values are then summed across all models to be included to arrive at the Bayesian model averaged regression coefficients. If all models are included in the model averaging, then the estimated regression coefficients are equal to the marginal posterior mean of the  $\hat{\beta}_j$ 's. The advantage of this approach in the context of Bayesian variable selection is a better fit when several models are likely; the disadvantage is that it does not resolve into a more parsimonious model than any of the individual models. The latter is why the method was not included in the context of the fractional polynomial method in Chapter 6. When applied to the simulation studies in Chapter 5 (not shown) after the application of BTS, the resulting quadratic models had bias in both the regression coefficients and the turning points.

## **8.3 BAYESIAN “FEEDBACK”**

Bayesian feedback occurs in Bayesian joint models when estimation of the parameters of the outcome model affects estimation of the parameters in subsidiary models, such as the measurement error and exposure models. This concept was explored by Lunn et al. in the context of pharmacokinetic/pharmacodynamic models [135]; more recently, Plummer [129] illustrated

the problem of feedback specifically for the Bayesian measurement error example, using the conditional independence structure of Richardson and Gilks [35].

### **8.3.1 Bayesian “feedback” in the measurement error correction methods presented**

The challenge of this thesis was finding methods that would perform well with a replicate study, high measurement error, or a binary outcome. These attributes may lead to a likelihood that is empirically unidentifiable (or a flat likelihood). It was felt that the use of prior information in Bayesian analysis would enable the model to overcome these shortcomings to arrive at informative posterior estimates. However, without strong priors for parameters in the measurement error and exposure models or a strong likelihood, feedback resulted in uninformative posteriors (or a flat posterior).

Feedback has been observed in a number of the methods implemented in this thesis. In Chapter 3, the use of the SMC-FCS implementation of MI ensured that imputations were drawn from the full joint posterior distribution. In doing so, when there was insufficient support from the likelihood, the feedback resulted in erratic parameter estimates. If MI had been performed without inclusion of the outcome or consideration of the form of the substantive model, more precise estimates might have been obtained but it is hard to say exactly what posterior distribution they would represent and whether they would be unbiased.

In Chapter 5, DIC was not useful as a means of discerning the best model. This was because by specifying either the quadratic or linear models, the parameters of the substantive model played a role in determining the parameter estimates in the measurement error and exposure models. As a result, the plausibility of each form of the substantive model was inflated.

The most severe effects of feedback were observed in Chapter 6 in the context of trying to achieve the flexibility of the fractional polynomial method while relying on a replicate study for information about the measurement error. The use of a joint model (BTS with measurement error correction) in this setting resulted in an uninformative posterior (i.e. erratic curve estimates). However, even when a validation study was present, the use of the joint model of this type resulted in higher empirical type I error as a result of feedback.

It is important to note that Bayesian feedback is not necessarily a drawback. In Chapter 4, when the substantive model was specified to be a quadratic model, the feedback of this information to estimating the parameter estimates of the subsidiary models (measurement error and exposure) is an element of Bayesian analysis which improves the estimates of the association resulting in unbiased and more efficient estimates [149].

### **8.3.2 Future directions: Potential methods for addressing Bayesian “feedback”**

The Bayesian RC methods presented in this thesis separate the estimation of the measurement error and exposure model parameters from the estimation of the substantive model parameters. The substantive model includes only a one-dimensional summary measure of the measurement error and exposure models. This approach is not unlike propensity score analysis wherein a logistic regression model is used to estimate the binary probability of being in the treated or untreated group in an observational study; a one-dimensional summary measure of that probability can be adjusted for in the exposure-outcome model [136]. This may be referred to broadly as “sequential analysis”. In 2009, McCandless introduced a Bayesian propensity score analysis approach which used the joint likelihood of the propensity score model and the outcome model for estimates of causal inference [151]. The feedback of information was noted in that publication as a feature resulting in more precise estimate of causal effect. However, Zigler et al demonstrated via simulation study that the Bayesian joint estimation of propensity scores and causal effects can result in biased estimates of the causal effects [136].

Stronger priors for the parameters of the subsidiary models and weaker priors for the parameters of the substantive model may improve estimation when the substantive model is unknown by effectively weighting the subsidiary models over the substantive model. In the implementations of BTS in Chapters 5 and 6, I have done the opposite by introducing stronger priors in the substantive model (in order to improve convergence for those methods). To my knowledge, there is no direct manner of introducing a prior that weights the faith in each model.

The most common solution to Bayesian “feedback” when the effects are undesirable is to “cut” the flow of information between the subsidiary models and the outcome model [135,137]. Posterior estimates that result from a cut model no longer represent draws from the full joint posterior; it is undefined what density they do represent, and therefore, these estimates are only approximately Bayesian.

The option to create a cut in the Bayesian model is available in OpenBUGS [93] and WinBUGS [94] but not in the JAGS software [62]. Plummer has demonstrated that the use of cuts as they are implemented in OpenBUGS results in different posterior distributions depending on the sampler used [129]. As an alternative, Plummer proposes an improved cut algorithm (“tempered cut algorithm”).

It has been suggested that MI would also achieve a separation of the joint model in a fashion similar to cut models were the outcome to be excluded from the imputation model [129]. For missing data, it has been shown that exclusion of the outcome from the imputation model results in greater bias than MI with the outcome included [103]. However, in that study no difference in the standard errors or the coverage was observed and both methods resulted in much less bias

than complete case analysis. For correction of additive measurement error where a validation study is available, it is possible that by using  $W$  as a (highly correlated) auxiliary variable for  $X$  the bias from ignoring the outcome would be minimal. Additionally, if the measurement error model were correctly specified, MI with a replicate study available which ignores the outcome would be similarly feasible. Further work should be done to compare 1) ignoring the outcome entirely in the imputation model to 2) incorporating the outcome into the imputation model in a way that incorrectly represents the relationship between exposure and outcome, i.e. assumes a linear relationship when, in fact, they are non-linearly related.

## 8.4 FURTHER EXTENSIONS

### 8.4.1 Extensions to multiple covariates measured with error

In this thesis, I have focused on the setting wherein a single exposure is measured with error. However, there will be many applications wherein there are multiple exposures measured with error. Even with classical measurement error in linear regression, this scenario can result in bias in any direction if the error is ignored [1,5]. The Bayesian methods outlined in this thesis should all extend, in theory, to accommodate multiple exposures measured with error.

Extension of RC to multiple covariates measured with error would depend on the correlation between the errors [1]. For specific dietary applications, there exists an extension of the EC model to a bivariate model applicable to the estimation of both dietary intake and overall caloric consumption [65]. It is unlikely that RC would be easily extended to many covariates measured with error if there exists a complex correlation structure.

If an additive error model is specified for each covariate measured with error and validation data are available for each, then MI with fully conditional specification (or chained equations) can be used without alteration to its use in the missing data context by treating each covariates as subject to missingness.

Muff et al demonstrated the application of INLA for measurement error correction for two covariates measured with error given a complex measurement error model (a mix of classical error and Berkson error, i.e.  $X = W + U$ ) [152]. It has therefore been demonstrated that this approach can be used for multiple covariates measured with error.

Bayesian analysis with MCMC could undoubtedly accommodate multiple covariates measured with error when the functional form of each latent covariate is known. When the functional form is unknown, the method of Bayesian variable selection that I extended to BTS with measurement error, SSVS, has been previously extended for use in the multivariate setting [61].

The Bayesian RC methods may most easily accommodate multiple covariates measured with error whatever the error structure between them. In fact, this approach is most similar to the two-stage analysis often used in environmental epidemiology wherein multiple data sources may be used in the first stage to estimate exposures and these estimated exposures are in the second stage used in the exposure-outcome model [149].

#### **8.4.2 Extensions to survival analysis**

All methods presented in this thesis could be extended for use in survival analysis. The general application of RC in the context of survival models has been previously asserted [72,73]; an approximation is required. The Bayesian implementations of RC require no further extension from what was presented in Section 4.4.4 and 6.4.2 to be applied in the context of survival analysis but they require the same approximation as frequentist RC.

When applied with the Cox proportional hazards model, the fractional polynomial method has been found to have a slightly elevated type I error rate in simulation studies [153]. Given this and the need for an approximation, any future work which combines survival analysis, RC (including Bayesian RC), and fractional polynomials should include an evaluation of the type I error.

Muff et al demonstrated the use of INLA for measurement error correction for an untransformed error-prone exposure with a survival outcome model [152]. INLA can accommodate survival outcomes through either the Cox proportional hazards model or the Weibull model [154].

While Bayesian analysis may be performed with the Cox proportional hazards model, this has been found to result in very slow convergence of the model [37]. Fully parametric models, such as the Weibull model, may be used as an alternative and results in faster convergence.

### **8.5 LIMITATIONS**

The shapes of association simulated in this thesis did not include features found to be challenging in the fractional polynomial method. The shapes were either monotonic in the exposure range or had a single maxima or minima. These relatively simple association shapes are examples of the types of shapes best fit by fractional polynomials [155]. More complex non-linear exposure-outcome relationships may require the use of alternative non-parametric methods. Additionally, the range of exposure values in the simulation studies did not include zeros or negative values. While negative values can be transformed to be in the positive range by the addition of a uniform value, the specific choice of value may result in quite disparate model selection introducing additional analytical challenge [8].

Due to the time involved in the convergence of MCMC methods, particularly those incorporating logistic regression, simulation studies in this thesis were limited to 100 – 400 simulated data sets.

This limits the extent to which bias and coverage may be assessed. The MC errors for bias were relatively small compared to the bias being addressed by measurement error correction. The MC error for coverage were significantly wider and only moderate to severe under or overcoverage could be distinguished.

In this thesis, only very limited exploration of the effects of misspecification of the error model were explored in a sensitivity analysis for a subset of the methods. While no sensitivity analysis was performed for BTS with measurement error correction wherein the measurement error model was misspecified, it has previously been shown that misspecification of the subsidiary model in a Bayesian joint modelling approach results in more biased estimates than sequential approaches [149,156].

## 8.6 FUTURE DIRECTIONS

Given the utility of polynomial models, I would suggest that further research into a fully Bayesian solution focus on the problem of model feedback. While the Bayesian RC methods proposed in this thesis were shown to improve power and reduce bias, they are, as one reviewer commented on my manuscript (Appendix H), a partial solution. In Section 8.3.2 I have outlined a number of possible solutions which may be pursued.

This thesis aimed to develop Bayesian methods of measurement error correction in the context of model selection. However, Bayesian polynomial model selection is itself an interesting pursuit. O'hara and Sillanpää published a review of Bayesian variable selection methods in which they divided the approaches used into three broad categories [120]: 1) indicator model selection; 2) adaptive shrinkage; and 3) model space approach. In this thesis, I focused on indicator model selection methods. In future work, a model space approach called the reversible jump MCMC, or some variation thereof, is promising in application to BTS.

In the model space approach, given a number of predictor variables,  $V$ , from  $j = 1, 2, \dots, J$ , at each iteration of the MCMC chain the removal or addition of a variable  $V_j$  is proposed. The number of variables included,  $N_V$  is therefore an additional parameter of the model and can be assigned a prior distribution to encourage sparsity. Furthermore, the method may be adapted to specify a maximum value for  $N_V$ . The ability to more directly specify the maximum degree of the model is appealing for its parallel to the frequentist fractional polynomial method.

Bové Sebanas published a more elegant model space algorithm for fractional polynomial model selection in high-dimensional analysis [55]. This method requires specialized programming of the sampler whereas I was primarily interested in methods that could be implemented in widely available software such as JAGS.

Hamiltonian Monte Carlo is an alternative sampling algorithm to MCMC which allow for more efficient sampling [82,157]. This includes better mixing for models that change states or have highly correlated variables. In further pursuit of fully Bayesian methods which may accommodate both measurement error correction and model selection, this alternative sampling method may improve estimates or at least improve speed and adaptability to higher dimensional data sets. This sampling algorithm is accessible through the Stan software [158].

## 8.7 CONCLUSIONS

In this thesis, I have proposed and assessed several Bayesian strategies for correcting for exposure measurement error in extensive simulation studies. These include the use of MI and INLA as well as two novel Bayesian approaches: Bayesian RC and the use of BTS with measurement error correction. All methods were compared to RC, arguably the most common method of measurement error correction, throughout this work.

RC is simple to implement and reduces bias in many scenarios, but it does not recover power lost to measurement error and results in estimates with large variability. Where the substantive model is specified, a fully Bayesian analysis using MCMC is a powerful approach which results in unbiased effect estimates with good coverage properties with much greater precision than RC. Where the functional form of the exposure is not specified, Bayesian analysis is subject to model feedback resulting in uninformative posterior estimates in the setting of a replicate study. INLA is a fast tool for finding Bayesian estimates, but the approximations inherent in the method resulted in undercoverage in some scenarios. Furthermore, INLA cannot accommodate a transformation of the latent exposure in the substantive model. MI using SMC-FCS was shown to be a reliable method when a validation study was available and the latent exposure could be treated as if it were missing data. However, when relying on a replicate study for estimation of the measurement error and exposure model parameters, MI was also subject to model feedback.

The most flexible approach in this thesis was the use of Bayesian RC. This approach is relatively easy to implement and effectively cuts the influence of the substantive model parameters on the estimation of latent error-prone exposure. Complex measurement error may be accommodated and the use of prior information can improve power in model selection. Bayesian RC was used to estimate the association between alcohol intake and all-cause mortality, incorporating both complex measurement error model and a non-linear exposure-outcome association.

In this work I have demonstrated that Bayesian methods may be used to reduce bias from and recover power lost to measurement error without sacrificing efficiency. These methods fill an important niche in applicability to applied research.



## BIBLIOGRAPHY

---

1. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models*. Vol 39. 2nd ed. Chapman & Hall/CRC, 2006.
2. Fuller WA. *Measurement Error Models*. John Wiley & Sons, Ltd, 1987.
3. Buonaccorsi JP. *Measurement Error: Models, Methods, and Applications*. (N. Keiding, B.J.T. Morgan, C.K. Wikle P van der H, ed.). Chapman & Hall/CRC: Boca Raton, 2010.
4. Brakenhoff TB, van Smeden M, Visseren FL, Groenwold RHH. Random measurement error: why worry? An example of cardiovascular risk factors, *PLoS One* **2018**; **13(2)**, pp. e0192298.
5. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology, *Stat Med* **2014**; **33(12)**, pp. 2137-2155.
6. Kuha J, Temple J. Covariate measurement error in quadratic regression, *Int Stat Rev* **2003**; **71(1)**, pp. 131-150.
7. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology, *Int J Epidemiol* **1999**; **28**, pp. 964-974.
8. Royston P, Sauerbrei W. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. (Balding DJ, Cressie NA, Fitzmaurice GM, et al., eds.). John Wiley & Sons, Ltd, 2008.
9. Brakenhoff TB, Mitroiu M, Keogh RH, et al. Measurement error is often neglected in medical literature: a systematic review, *J Clin Epidemiol* **2018**; **98**, pp. 89-97.
10. Sauerbrei W, Abrahamowicz M, Altman DG, et al. Strengthening analytical thinking for observational studies: The STRATOS initiative, *Stat Med* **2014**; **33(30)**, pp. 5413-5432.
11. Wood SN. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, 2006.
12. Hastie T, Tibshirani R. Generalized additive models, *Stat Sci* **1986**; **1(3)**, pp. 297-318.
13. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling, *Appl Stat* **1994**; **43(3)**, pp. 429-467.
14. Abbas S, Linseisen J, Slinger T, et al. Serum 25-hydroxyvitamin D and risk of postmenopausal breast cancer - Results of a large case-control study, *Carcinogenesis* **2008**;

**29(1)**, pp. 93-99.

15. Cao S, Moineddin R, Urquia ML, et al. J-shapedness: an often missed, often miscalculated relation: the example of weight and mortality, *J Epidemiol Community Health* **2014**; **68(7)**, pp. 683-690.
16. Takagi H, Umemoto T. 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 meta-regression of large statin randomized trials, *JAMA Intern Med* **2013**; **173(11)**, pp. 1028-1029.
17. Kipnis V, Midthune D, Buckman DW, et al. Modeling data with excess zeros and measurement error: Application to evaluating relationships between episodically consumed foods and health outcomes, *Biometrics* **2009**; **65(4)**, pp. 1003-1010.
18. Keogh RH, White IR. Allowing for never and episodic consumers when correcting for error in food record measurements of dietary intake, *Biostatistics* **2011**; **12(4)**, pp. 624-636.
19. Zhang S, Krebs-Smith SM, Midthune D, et al. Fitting a bivariate measurement error model for episodically consumed dietary components., *Int J Biostat* **2011**; **7(1)**, pp. 1-32.
20. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error, *Stat Med* **1989**; **8(9)**, pp. 1051-1069.
21. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error, *Am J Epidemiol* **1992**; **136(11)**, pp. 1400-1413.
22. Reeves GK, Cox DR, Darby SC, Whitley E. Some aspects of measurement error in explanatory variables for continuous and binary regression models, *Stat Med* **1998**; **17(19)**, pp. 2157-2177.
23. Keogh RH, Strawbridge AD, White IR. Effects of classical exposure measurement error on the shape of exposure-disease associations, *Epidemiol Method* **2012**; **1(1)**, pp. 13-32.
24. Di Castelnuovo A, Costanzo S, Bagnardi V, et al. Alcohol dosing and total mortality in men and women, *Arch Intern Med* **2006**; **166**, pp. 2437-2445.
25. Knott CS, Coombs N, Stamatakis E, Biddulph JP. All cause mortality and the case for age specific alcohol consumption guidelines: pooled analyses of up to 10 population based cohorts, *Bmj* **2015**; **350**, pp. h384.

26. Stockwell T, Zhao J, Panwar S, et al. Do “moderate” drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality, *J Stud Alcohol Drugs* **2016**; **77(2)**, pp. 185-198.
27. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. CRC Press, 2003.
28. Peter Congdon. *Bayesian Statistical Modelling*. 2nd ed. John Wiley & Sons, Ltd: West Sussex, England, 2006.
29. Muff S, Riebler A, Held L, et al. Bayesian analysis of measurement error models using integrated nested Laplace approximations, *J R Stat Soc Ser C (Applied Stat)* **2015**; **64(2)**, pp. 231-252.
30. Taylor B, Irving HM, Baliunas D, et al. Alcohol and hypertension: Gender differences in dose-response relationships determined through systematic review and meta-analysis, *Addiction* **2009**; **104(12)**, pp. 1981-1990.
31. Irving HM, Samokhvalov A V., Rehm J. Alcohol as a risk factor for pancreatitis. A systematic review and meta-analysis, *J Pancreas* **2009**; **10(4)**, pp. 387-392.
32. Costanzo S, Di Castelnuovo A, Donati MB, et al. Alcohol consumption and mortality in patients with cardiovascular disease. A meta-analysis, *J Am Coll Cardiol* **2010**; **55(13)**, pp. 1339-1347.
33. Armstrong B. Measurement error in the generalised linear model, *Commun Stat - Simul Comput* **1985**; **14(3)**, pp. 529-544.
34. Rosner B, Spiegelman D, Willett WC, et al. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error, *Am J Epidemiol* **1990**; **132(4)**, pp. 734-745.
35. Richardson S, Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models, *Am J Epidemiol* **1993**; **138(6)**, pp. 430-442.
36. Richardson S, Gilks WR. Conditional independence models for epidemiological studies with covariate measurement error, *Stat Med* **1993**; **12(18)**, pp. 1703-1722.
37. Bartlett JW, Keogh RH. Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration, *Stat Methods Med Res* **2018**; **27(6)**, pp. 1695-1708.
38. Hossain S, Gustafson P. Bayesian adjustment for covariate measurement errors: A flexible

- parametric approach, *Stat Med* **2009**; **28(28)**, pp. 1580-1600.
39. Muller P, Roeder K. A Bayesian semiparametric model for case-control studies with errors in variables, *Biometrika* **1997**; **84(3)**, pp. 523-537.
  40. Richardson S, Leblond L, Jaussent I, Green PJ. Mixture models in measurement error problems, with reference to epidemiological studies, *J R Stat Soc Ser A (Statistics Soc)* **2002**; **165(3)**, pp. 549-566.
  41. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *J R Stat Soc Ser B Stat Methodol* **2009**; **71(2)**, pp. 319-392.
  42. Ferkingstad E, Rue H. Improving the INLA approach for approximate Bayesian inference for latent Gaussian models, **2015**; **9(2)**, pp. 2706-2731.
  43. Rue H, Riebler A, Sørbye SH, et al. Bayesian computing with INLA: A review, *Annu Rev Stat Its Appl* **2017**; **4**, pp. 395-421.
  44. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Vol 81. 1st ed. John Wiley & Sons: New York, NY, 1987.
  45. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values, *Stat Med* **2010**; **29(28)**, pp. 2920-2931.
  46. Bartlett JW, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model, *Stat Methods Med Res* **2015**; **24(4)**, pp. 462-487.
  47. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction, *Int J Epidemiol* **2006**; **35(4)**, pp. 1074-1081.
  48. Freedman LS, Midthune D, Carroll RJ, Kipnis V. A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression, *Stat Med* **2008**; **27(25)**, pp. 5195-5216.
  49. Strawbridge AD. Modelling non-linear exposure–disease relationships in a large individual participant meta-analysis allowing for the effects of exposure measurement error, PhD thesis, Robinson College, University of Cambridge, 2011.
  50. Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies, *Am J Epidemiol* **1999**; **150(6)**, pp. 642-651.

51. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity, *Am J Epidemiol* **2007**; **166(7)**, pp. 832-840.
52. Bhadra A, Wei R, Lilly E, et al. *Measurement Error Models with Zero Inflation and Hard Zeros, with Applications to Never-Consumers in Nutrition*. Working paper (unpublished) 2017.
53. Berry SM, Carroll RJ, Ruppert D. Bayesian smoothing and regression splines for measurement error problems, *J Am Stat Assoc* **2002**; **97(457)**, pp. 160-169.
54. Gossel C, Küchenhoff H. Bayesian analysis of logistic regression with an unknown change point and covariate measurement error, *Stat Med* **2001**; **20(3)**, pp. 3109-3121.
55. Bové Sabanés D, Held L. Bayesian fractional polynomials, *Stat Comput* **2011**; **21(3)**, pp. 309-324.
56. Zeger SL, Thomas D, Dominici F, et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences, *Environ Health Perspect* **2000**; **108(5)**, pp. 419-426.
57. Dominici F, Zeger SL, Samet JM. A measurement error model for time-series studies of air pollution and mortality, *Biostatistics* **2000**; **1(2)**, pp. 157-175.
58. Mwalili SM, Lesaffre E, Declerck D. A Bayesian ordinal logistic regression model to correct for interobserver measurement error in a geographical oral health study, *J R Stat Soc Ser C (Applied Stat)* **2005**; **54(2)**, pp. 77-93.
59. Zhang S, Midthune D, Guenther PM, et al. A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment, *Ann Appl Stat* **2011**; **5(2B)**, pp. 1456-1487.
60. Kuo L, Mallick B. Variable selection for regression models, *Sankhyā Indian J Stat Ser B* **1998**; **60(1)**, pp. 65-81.
61. Brown P, Vannucci M, Fearn T. Multivariate Bayesian variable selection and prediction, *J R Stat Soc Ser B (Statistical Methodol)* **1998**; **60(3)**, pp. 627-641.
62. Plummer M. Just Another Gibbs Sampler (JAGS), version 4.3.0, 2017.
63. Pickering TG, Hall JE, Appel LJ, et al. Recommendations for blood pressure measurement in humans and experimental animals: Part 1: Blood pressure measurement in humans: A statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Cou, *Circulation* **2005**; **111(5)**, pp. 697-716.

64. Toozé JA, Midthune D, Dodd KW, et al. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution, *J Am Diet Assoc* **2006**; **106**(10), pp. 1575-1587.
65. Kipnis V, Freedman LS, Carroll RJ, Midthune D. A bivariate measurement error model for semicontinuous and continuous variables: Application to nutritional epidemiology, *Biometrics* **2016**; **72**(1), pp. 106-115.
66. Kannel WB, Neaton JD, Wentworth D, et al. Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for the MRFIT, *Am Heart J* **1986**; **112**(4), pp. 825-836.
67. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease, *Lancet* **1990**; **335**(8693), pp. 765-774.
68. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer., *Br J Cancer* **1999**; **80** Suppl 1(1993), pp. 95-103.
69. Riboli E, Hunt K, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection, *Public Health Nutr* **2002**; **5**(6b), pp. 1113.
70. Bingham S, Welch A, McTaggart A, et al. Nutritional methods in the European Prospective Investigation of Cancer in Norfolk, *Public Health Nutr* **2001**; **4**(3), pp. 847.
71. Welch A, McTaggart A, Mulligan A, et al. 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 Nutr* **2001**; **4**(6), pp. 1253-1265.
72. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement error bias in nutritional epidemiology, *Am J Clin Nutr* **1997**; **65**(4), pp. 1179s-1186s.
73. Hughes MD. Regression dilution in the proportional hazards model., *Biometrics* **1993**; **49**(4), pp. 1056-1066.
74. Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument, *Stat Med* **2001**; **20**(1), pp. 139-160.
75. DiCiccio TJ, Efron B. Bootstrap confidence intervals, *Stat Sci* **1996**; **11**(3), pp. 189-228.

76. Efron B, Tibshirani R. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, *Stat Sci* **1986**; **1(1)**, pp. 54-75.
77. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, *Stat Med* **2000**; **19(9)**, pp. 1141-1164.
78. Hardin JW, Schmiediche H, Carroll RJ. The regression calibration method for fitting generalized linear models with additive measurement error, *Stata J* **2003**; **3(4)**, pp. 373-385.
79. Metropolis N, Rosenbluth AW, Rosenbluth MN, et al. Equation of state calculations by fast computing machines, *J Chem Phys* **1953**; **21(6)**, pp. 1087-1092.
80. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications, *Source: Biometrika Biometrika* **1970**; **57(1)**, pp. 97-109.
81. Martins TG, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: New features, *Comput Stat Data Anal* **2013**; **67**, pp. 68-83.
82. Brooks S, Gelman A, Jones GL, Meng X-L, eds. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.
83. Duane S, Kennedy AD, Pendleton BJ, Roweth D. Hybrid Monte Carlo, *Phys Lett B* **1987**; **195(2)**, pp. 216-222.
84. Stephens D, Dellaportas P. Bayesian analysis of generalised linear models with covariate measurement error, *Bayesian Stat* **1992**; **4**, pp. 813-820.
85. Kruschke JK. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press, 2010.
86. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans Pattern Anal Mach Intell* **1984**; **(6)**, pp. 721-741.
87. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. 3rd ed. Chapman & Hall/CRC, 2013.
88. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations, *J Comput Graph Stat* **1998**; **7(4)**, pp. 434-455.
89. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *Bayesian Stat* **1992**; **4**, pp. 641-649.
90. Chen MH, Shao QM. Monte Carlo estimation of Bayesian credible and HPD intervals, *J Comput Graph Stat* **1999**; **8(1)**, pp. 69-92.

91. Efron B. Frequentist accuracy of Bayesian estimates, *J R Stat Soc Ser B (Statistical Methodol)* **2015**; **77(Part 3)**, pp. 617–646.
92. Carroll BRJ, Maca JD, Ruppert D. Nonparametric regression in the presence of measurement error, *Biometrika* **1999**; **86(3)**, pp. 541-554.
93. Spiegelhalter DJ, Thomas A, Best N, Lunn D. *OpenBUGS User Manual*. 2014.
94. Thomas A, Best N, Way R. WinBUGS User Manual, 2003, pp. 1-60.
95. Plummer M. R package rjags: Bayesian graphical models using MCMC, version 4-6, 2015.
96. Thomas A, Lunn D, Spiegelhalter D, et al. R package BRugs: Interface to the “OpenBUGS” MCMC software, version 0.8-6, 2014.
97. Plummer M, Best N, Cowles K, et al. R package coda: Output analysis and diagnostics for MCMC, version 0.19-1, 2015.
98. Carpenter JR, Kenward MG. *Multiple Imputation and Its Application*. 1st ed. John Wiley & Sons, Ltd, 2013.
99. Royston P, White I. Multiple imputation by chained equations (MICE): implementation in Stata, *J Stat Softw* **2011**; **45(4)**, pp. 1-20.
100. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained in R, *J Stat Softw* **2011**; **45(3)**.
101. Bartlett JW, Morris TP. Multiple imputation of covariates by substantive model compatible fully conditional specification, *Stata J* **2015**; **15(2)**, pp. 437-456.
102. Wang N, Robins JM. Large-sample theory for parametric multiple imputation procedures, *Biometrika* **1998**; **85(4)**, pp. 935-948.
103. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred, *J Clin Epidemiol* **2006**; **59(10)**, pp. 1092-1101.
104. Liu J, Gelman A, Hill J, et al. On the stationary distribution of iterative imputations, *Biometrika* **2013**; **101(1)**, pp. 155-173.
105. Buuren S Van. Multiple imputation of discrete and continuous data by fully conditional specification, *Stat Methods Med Res* **2007**; **16**, pp. 219-242.
106. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice, *Stat Med* **2011**; **30(4)**, pp. 377-399.



107. Bartlett JW. R package smcfcs: Multiple imputation of covariates by substantive model compatible fully conditional specification, version 1.3.0, 2015.
108. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models, *J Am Stat Assoc* **1994**; **89(428)**, pp. 1314-1328.
109. White IR. simsum: Analyses of simulation studies including Monte Carlo error, *Stata J* **2010**; **10(3)**, pp. 369-385.
110. Kuha J. Estimation by data augmentation in regression models with continuous and discrete covariates measured with error, *Stat Med* **1997**; **16**, pp. 189-201.
111. Von Hippel PT. How to impute interactions, squares, and other transformed variables, *Sociol Methodol* **2009**; **39(1)**, pp. 265-291.
112. Vink G, van Buuren S. Multiple imputation of squared terms, *Sociol Methods Res* **2013**; **42(4)**, pp. 598-607.
113. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods, *BMC Med Res Methodol* **2012**; **12(1)**, pp. 46.
114. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press, 2003.
115. Akaike H. Information theory and an extension of the maximum likelihood principle, *Int Symp Inf theory* **1973**; **(1973)**, pp. 267-281.
116. Dziak JJ, Coffman DL, Lanza ST, Runze L. Sensitivity and specificity of information criteria, *Methodol Cent Dep Stat Penn State* **2012**; **16(30)**, pp. 140.
117. Leeb H, Pötscher BM. Model Selection In: Andersen T, Davis RA, Kreiß JP, Mikosch T, eds. *Handbook of Financial Time Series*. Springer: Berlin,; 2009.
118. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion), *J R Stat Soc Ser B (Statistical Methodol)* **2002**; **64(4)**, pp. 583-639.
119. Spiegelhalter DJ, Best NG, Carlin BP, Linde A Van Der. The deviance information criterion: 12 years on (with discussion), *J R Stat Soc Ser B* **2014**; **64**, pp. 485-493.
120. O'Hara RB, Sillanpää MJ. A review of Bayesian variable selection methods: What, how and which, *Bayesian Anal* **2009**; **4(1)**, pp. 85-118.
121. Hoeting JA, Ibrahim JG. Bayesian predictive simultaneous variable and transformation

- selection in a linear model, *J Comput Stat Data Anal* **1998**; **28(1)**, pp. 87-103.
122. Hoeting JA, Raftery A, Madigan D. Bayesian variable and transformation selection in linear regression, *J Comput Graph Stat* **2002**; **11(3)**, pp. 485-507.
  123. Gottardo R, Raftery A. Bayesian robust transformation and variable selection: a unified approach, *Can J Stat* **2009**; **37(3)**, pp. 361-380.
  124. George EI, McCulloch RE. Variable selection via Gibbs sampling, *J Am Stat Assoc* **1993**; **88(423)**, pp. 881-889.
  125. Jeffreys H. *Theory of Probability*. Vol 2. 1961.
  126. Lee MD, Wagenmakers E-J. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press, 2014.
  127. Hoff PD. *A First Course in Bayesian Statistical Methods*. Springer Science & Business Media, 2009.
  128. Frühwirth-Schnatter S. Markov chain monte carlo estimation of classical and dynamic switching and mixture models, *J Am Stat Assoc* **2001**; **96(453)**, pp. 194-209.
  129. Plummer M. Cuts in Bayesian graphical models, *Stat Comput* **2014**; **25(1)**, pp. 37-43.
  130. Fraser GE, Stram DO. Regression calibration in studies with correlated variables measured with error, *Am J Epidemiol* **2001**; **154(9)**, pp. 836-844.
  131. Freedman LS, Tasevska N, Kipnis V, et al. Gains in statistical power from using a dietary biomarker in combination with self-reported intake to strengthen the analysis of a diet-disease association: An example from CAREDS, *Am J Epidemiol* **2010**; **172(7)**, pp. 836-842.
  132. Tidwell PW, Box GEP. Transformation of the independent variables, *Technometrics* **1962**; **4(4)**, pp. 531-550.
  133. Ambler G, Benner A. R package mfp: Multivariable fractional polynomials, version 1.5.2, 2015.
  134. Ambler G, Royston P. Fractional polynomial selection procedures: investigation of type I error rate, *J Stat Comput Simul* **2001**; **69(1999)**, pp. 89-108.
  135. Lunn D, Best N, Spiegelhalter D, et al. Combining MCMC with “sequential” PKPD modelling, *J Pharmacokinet Pharmacodyn* **2009**; **36(1)**, pp. 19-38.
  136. Zigler CM, Watts K, Yeh RW, et al. Model feedback in Bayesian propensity score

- estimation, *Biometrics* **2013**; **69(1)**, pp. 263-273.
137. McCandless LC, Douglas IJ, Evans SJ, Smeeth L. Cutting feedback in Bayesian regression adjustment for the propensity score, *Int J Biostat* **2010**; **6(2)**.
  138. Fillmore KM, Stockwell T, Chikritzhs T, et al. Moderate alcohol use and reduced mortality risk: Systematic error in prospective studies and new hypotheses, *Ann Epidemiol* **2007**; **17(5S)**, pp. s16-S23.
  139. Ferrari P, Licaj I, Muller DC, et al. Lifetime alcohol use and overall and cause-specific mortality in the European Prospective Investigation into Cancer and nutrition (EPIC) study, *BMJ Open* **2014**; **4(7)**, pp. e005245.
  140. Thun MJ, Peto R, Lopez AD, et al. Alcohol consumption and mortality among middle-aged and elderly U.S. adults, *N Engl J Med* **1997**; **337(24)**, pp. 1705-1714.
  141. Khaw K-T, Wareham N, Bingham S, et al. Combined impact of health behaviours and mortality in men and women: The EPIC-Norfolk prospective population study, *PLoS Med* **2008**; **5(1)**, pp. e12.
  142. Box GEP, Cox DR. An analysis of transformations, *J R Stat Soc Ser B* 1964, pp. 211-252.
  143. Alvarez I, Niemi J, Simpson M. Bayesian inference for a covariance matrix, *Ann Stat* **1992**; **20(4)**, pp. 1669-1696.
  144. Raftery a. E, Lewis SM. The number of iterations, convergence diagnostics and generic Metropolis algorithms, *Pract Markov Chain Monte Carlo* **1995**; **7(98)**, pp. 763-773.
  145. R Core Team (R Foundation for Statistical Computing). R: A language and environment for statistical computing, 2016.
  146. Bergmann MM, Rehm J, Klipstein-Grobusch K, et al. The association of pattern of lifetime alcohol use and cause of death in the European Prospective Investigation into Cancer and Nutrition (EPIC) study, *Int J Epidemiol* **2013**; **42(6)**, pp. 1772-1790.
  147. Jayasekara H, English DR, Room R, MacInnis RJ. Alcohol consumption over time and risk of death: A systematic review and meta-analysis, *Am J Epidemiol* **2014**; **179(9)**, pp. 1049-1059.
  148. Morris TP, White IR, Carpenter JR, et al. Combining fractional polynomial model building with multiple imputation, *Stat Med* **2015**; **34(25)**, pp. 3298-3317.
  149. Szpiro AA. Measurement error in two-stage analyses, with application to air pollution epidemiology, *Environmetrics* **2013**; **24(8)**, pp. 501-517.

150. Lesaffre E, Lawson B. A. *Bayesian Biostatistics*. John Wiley & Sons, 2012.
151. McCandless LC. Bayesian propensity score analysis for observational data, *Stat Med* **2009**; **28**(1), pp. 94-112.
152. Muff S, Ott M, Braun J, Held L. Bayesian two-component measurement error modelling for survival analysis using INLA—A case study on cardiovascular disease mortality in Switzerland, *Comput Stat Data Anal* **2017**; **113**, pp. 177-193.
153. Hollander N. Estimating the functional form of the effect of a continuous covariate on survival time, 2002.
154. Martino S, Akerkar R, Rue H. Approximate Bayesian inference for survival models, *Scand J Stat* **2011**; **38**(3), pp. 514-528.
155. Govindarajulu US, Malloy EJ, Ganguli B, et al. The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study, *Int J Biostat* **2009**; **5**(1).
156. Zhang L, Beal SL, Sheiner LB. Simultaneous vs. sequential analysis for population PK/PD data II: Robustness of methods, *J Pharmacokinet Pharmacodyn* **2003**; **30**(6), pp. 405-416.
157. Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods, *J R Stat Soc Ser B (Statistical Methodol)* **2011**; **73**(2), pp. 123-214.
158. Gelman A, Lee D, Guo J. Stan: A probabilistic programming language for Bayesian inference and optimization, *J Educ Behav Stat* **2015**; **40**(5), pp. 530-543.

# Appendices

## A. CODE FOR MEASUREMENT ERROR CORRECTION METHODS IN GENERALIZED LINEAR MODELS WHEN THE EXPOSURE IS UNTRANSFORMED

---

The following code can be used to apply RC as presented in Section 2.2, Bayesian analysis using MCMC as presented in Section 2.4, Bayesian analysis using INLA as presented in Section 2.5, and the SMC-FCS method of MI as presented in Section 2.6.2. The code for each method may be applied with either a validation or a replicate study and either linear or logistic regression. This code was used for the simulation studies in Chapter 3.

```
#####  
# Logistic regression validation study (30%) #  
# log(OR)=0.35, #  
# measurement error equal to variance of X #  
#####  
set.seed(1989746)  
n=2000  
x_sd=1  
m_error_sd=1  
target.beta=0.35  
num_reps=1  
prop_observed=0.3  
  
X <- rnorm(n, 10, x_sd)  
logit_Y <- -5.7 + target.beta*X  
pr <- exp(logit_Y)/(1 + exp(logit_Y))  
Y <- rbinom(n, 1, pr)  
w1 <- X + rnorm(n, 0, m_error_sd)  
set <- sample.int(n, size=floor(n*prop_observed), replace=F)  
observed <- c(1:n)  
observed[T] <- observed %in% set  
xobs <- X; xobs[observed==F] <- NA  
  
d <- data.frame(Y, w1, xobs)  
rm(X, logit_Y, pr, Y, w1, set, observed, xobs)  
  
#####  
# Linear regression replicate study (100%) #  
# effect size=1.4 #  
# measurement error equal to variance of X #  
#####  
set.seed(1989746)  
n=2000  
x_sd=1  
m_error_sd=1  
y_sd=1  
target.beta=1.4  
num_reps=2  
  
X <- rnorm(n, 10, x_sd)  
Y <- -(y_sd/2 * target.beta) + target.beta*X + rnorm(n, 0, y_sd)  
w1 <- X + rnorm(n, 0, m_error_sd)  
w2 <- X + rnorm(n, 0, m_error_sd)  
  
d <- data.frame(Y,w1,w2)  
rm(X, Y, w1, w2)  
  
#####  
# Regression calibration #  
#####  
  
naive.function <- function(d, target.beta, mod.family, Z=c(NA), pred='w1') {  
  if (is.na(Z[1])) {  
    my.formula = as.formula(paste('Y ~', pred))
```

```

} else {
  my.formula = as.formula(paste('Y ~', pred, '+', paste0(Z, collapse=' + ' ) ))
}
naive.mod <- glm(my.formula, family=mod.family, data=d)
return(c(beta_naive=naive.mod$coeff[2], beta_naive_var=summary(naive.mod)$coeff[2,2]^2))
}

rc.w <- function(d, m1, m2, Z = c(NA)) {
  d$m1 <- d[,m1]
  d$m2 <- d[,m2]

  if (is.na(Z[1])) {
    pred_w <- lm(m2 ~ m1, na.action=na.exclude, data=d)
  } else {
    my.formula <- as.formula(paste("m2", '~', "m1", " + ", paste(Z, collapse=' + ' ) ))
    pred_w <- lm(my.formula, data=d, na.action=na.exclude)
  }
  return(list(fitted=fitted(pred_w), resid = resid(pred_w),
            rc_coef=pred_w$coeff[2],
            rc_coef_var = (summary(pred_w)$coefficients[2,2]^2))
)

rc.delta <- function(beta_naive, rc_coef, beta_naive_var, rc_coef_var) {
  se <- sqrt((beta_naive_var/rc_coef^2) + ((beta_naive/rc_coef^2)^2)*rc_coef_var)
}

rc.function <- function(d, mod.family='binomial', m1='w1', m2='w2') {
  #for validation, replace m2='w2' with m2='Xobs'

  d$m1 <- d[,m1]
  d$m2 <- d[,m2]
  rc.x <- rc.w(d, m1, m2)
  naive <- naive.function(d, target.beta, mod.family)
  rc.beta <- naive[1]/rc.x[[3]]
  rc.se.delta <- rc.delta(naive[1], rc.x[[3]],naive[2],rc.x[[4]])
  cis <- sort(rc.beta+qnorm(c(0.025, 0.975))*rc.se.delta)

  return(c(rc.beta=rc.beta,
          rdr=rc.x[[3]],
          rc.se.delta=rc.se.delta,
          lci=cis[1],uci=cis[2]))
}

#rc.function(d, m2='Xobs')

#####
# Multiple imputation #
# Adapted from available R package 'smcfcs' #
# to extend for use with replicate studies #
#####
smcfcs <- function(originaldata, smtype, smformula, method, predictorMatrix = NULL,
                  m = 10, numit = 10, rjlimit=10000, ftl=25, noisy = FALSE,
                  errorProneMatrix = NULL,
                  priors='gustafson', title) {

  require(rje)
  require(MASS)

  outcomeCol <- which(colnames(originaldata) == as.formula(smformula)[[2]])

  #checking incoming data is correcting defined
  stopifnot(is.data.frame(originaldata))
  if (ncol(originaldata) != length(method)) {
    stop("Method argument must have the same length as the number of
         columns in the data frame.")
  } else {
    if (sum(method[outcomeCol] != c("", "")) > 0)
      stop("The elements of the method argument corresponding to the
           outcome variables should be empty.")
  }

  n <- dim(originaldata)[1]
  r <- 1 * (is.na(originaldata) == 0)

  smcovnames <- attr(terms(as.formula(smformula)), "term.labels")
  smcovcols <- (1:ncol(originaldata))[colnames(originaldata) %in% smcovnames]
  partialvars <- which((method == "norm") | (method == "latnorm"))
  fullobsvars <- which((colSums(r) == n) & (colnames(originaldata) %in% smcovnames))
  passivevars <- which((method != "") & (method != "norm") & (method != "latnorm"))

  if (noisy == T) {
    print(paste("Outcome variable(s):",
              paste(colnames(originaldata)[outcomeCol], collapse = ", ")))
    print(paste("Passive variables:",
              paste(colnames(originaldata)[passivevars], collapse = ", ")))
    print(paste("Partially obs. variables:",

```

```

        paste(colnames(originaldata)[partialVars], collapse = ","))
print(paste("Fully obs. substantive model variables:",
           paste(colnames(originaldata)[fullObsVars], collapse = ",")))
}

imputations <- list()
for (imp in 1:m) {
  imputations[[imp]] <- originaldata
}

#LOOP OF EACH IMPUTATION
for (imp in 1:m) {
  print(paste("Imputation ", imp, timestamp()))
  for (var in 1:length(partialVars)) {
    targetCol <- partialVars[var]
    if (method[targetCol] == "latnorm") {
      #Replicate study
      errorProneCols <- which(errorProneMatrix[targetCol, ] == 1)
      imputations[[imp]][, targetCol] <- imputations[[imp]][, errorProneCols[1]]
    } else {
      #Validation study
      imputations[[imp]][r[, targetCol] == 0, targetCol] <-
        sample(imputations[[imp]][r[, targetCol] == 1, targetCol],
              size = sum(r[, targetCol] == 0), replace = TRUE)
    }
  }
}

#LOOP OF EACH CYCLE OR ITERATION
for (cyclenum in 1:numit) {
  if (noisy == T) {
    print(paste("Impuation", imp, "and Iteration", cyclenum))
  }
  #update the passive variables (ie. X^2) based on X^(t-1)
  imputations[[imp]] <- updatePassiveVars(imputations[[imp]],
                                          method, passiveVars)

  #For each variable to be imputed - modified code only
  # implemented for 1 partially observed variable
  for (var in 1:length(partialVars)) {
    targetCol <- partialVars[var]
    if (is.null(predictorMatrix)) {
      predictorCols <- c(partialVars[!partialVars %in% targetCol], fullObsVars)
    } else {
      predictorCols <- which(predictorMatrix[targetCol, ] == 1)
      predictorCols <- predictorCols[!predictorCols %in% outcomeCol]
    }

    if ((noisy == T) & (imp == 1) & (cyclenum == 1)) {
      if (method[targetCol] == "latnorm") {
        print(paste("Imputing: ", colnames(imputations[[imp]][targetCol],
                                           " using ", paste(colnames(imputations[[imp]][c(predictorCols,
                                           which(errorProneMatrix[targetCol, ] ==
                                           1))], collapse = ","), " plus outcome",
                                           collapse = ",")))
      } else {
        print(paste("Imputing: ", colnames(imputations[[imp]][targetCol],
                                           " using ", paste(colnames(imputations[[imp]][predictorCols],
                                           collapse = ","), " plus outcome", collapse = ",")))
      }
    }

    if (length(predictorCols) > 0) {
      xmodformula <- as.formula(paste(colnames(imputations[[imp]][targetCol],
                                             "~", paste(colnames(imputations[[imp]][predictorCols],
                                             collapse = "+"), sep = ""))
    } else {
      xmodformula <- as.formula(paste(colnames(imputations[[imp]][targetCol],
                                             "~1", sep = ""))
    }

    if (method[targetCol] == "norm") {
      xmod <- lm(xmodformula, data = imputations[[imp]])
      sample.alpha <- xmod$coef
      sample.sigmaxsq <- summary(xmod)$sigma^2
      sigmaxsq <- (sample.sigmaxsq * xmod$df)/rchisq(1, xmod$df)
      covariance <- (sigmaxsq/sample.sigmaxsq) * vcov(xmod)
      alpha = sample.alpha +
        MASS::mvrnorm(1, mu = rep(0, ncol(covariance)), sigma = covariance)
      xfitted <- model.matrix(xmod) %>% alpha
    } else if (method[targetCol] == "latnorm") {
      xmod <- lm(xmodformula, data = imputations[[imp]])
      sample.alpha <- xmod$coef
      sample.sigmaxsq <- summary(xmod)$sigma^2

      if (priors=='flat') {

```

```

prior.sigmaxsq <- xmod$df/rchisq(1, xmod$df)
sigmaxsq <- sample.sigmaxsq * prior.sigmaxsq
covariance <- (sigmaxsq/sample.sigmaxsq) * vcov(xmod)
alpha <- sample.alpha +
  MASS::mvrnorm(1, mu = rep(0, ncol(covariance)), Sigma = covariance)
}
if (priors=='gustafson') {
  k1 = 2; k2 = 2
  prior.est.sigmaxsq = 1
  prior.est.x = 0
  sigmaxsq <- MCMCpack::rinvgamma(1, (xmod$df+k2)/2,
    (xmod$df*sample.sigmaxsq + k2*prior.est.sigmaxsq)/2)
  alpha <- rep(rnorm(1, (xmod$df/(xmod$df + k1))*sample.alpha +
    (k1/(xmod$df + k1))*prior.est.x,
    sqrt(sigmaxsq/(xmod$df + k1))), n)
}
xfitted <- model.matrix(xmod) * alpha
}

if (method[targetCol] == "latnorm") {
#Replicate study
errorProneCols <- which(errorProneMatrix[targetCol, ] == 1)
wmean <- rowMeans(imputations[[imp]][, errorProneCols], na.rm = TRUE)
n_i <- rowSums(!is.na(imputations[[imp]][, errorProneCols]))
sum_ni <- sum(n_i)

if (cyclenum == 1) {
  xmat <- matrix(wmean, nrow = nrow(imputations[[imp]]),
    ncol = length(errorProneCols))
  uvec <- c(as.matrix(imputations[[imp]][, errorProneCols] - xmat))
  sample.sigmaxsq <- sum(uvec^2, na.rm = TRUE)/(sum_ni - n)
} else {
  xmat <- matrix(imputations[[imp]][, targetCol],
    nrow = nrow(imputations[[imp]]), ncol = length(errorProneCols))
  uvec <- c(as.matrix(imputations[[imp]][, errorProneCols] - xmat))
  sample.sigmaxsq <- sum(uvec^2, na.rm = TRUE)/sum_ni
}

if (priors=='flat') {
  prior.sigmaxsq <- sum_ni/rchisq(1, sum_ni)
  sigmaxsq <- sample.sigmaxsq * prior.sigmaxsq
}
if (priors=='gustafson') {
  k3=2
  prior.est.sigmaxsq = 1
  sigmaxsq <- MCMCpack::rinvgamma(1, (sum_ni+k3)/2, (sum_ni*sample.sigmaxsq +
    k3*prior.est.sigmaxsq)/2)
}

lambda <- sigmaxsq/(sigmaxsq + sigmaxsq/n_i)
xfitted <- xfitted + lambda * (wmean - xfitted)
sigmaxwsq <- sigmaxsq * (1 - lambda)
} else {
#Validation study
sigmaxwsq <- sigmaxsq
}

#DRAW THE PARAMETERS FOR THE SUBSTANTIVE MODEL BASED ON THE DATA WITH X^(t-1)
if (smtype == "logistic") {
  ymod <- glm(as.formula(smformula), family = "binomial", imputations[[imp]])
  sample.beta <- ymod$coef
  sample.varcov <- vcov(ymod)

  outcomeModBeta = sample.beta +
    MASS::mvrnorm(1, mu = rep(0, ncol(sample.varcov)), Sigma =
sample.varcov)

} else if (smtype == "lm") {
  ymod <- lm(as.formula(smformula), imputations[[imp]])
  sample.beta <- ymod$coef
  sample.sigmaxsq <- summary(ymod)$sigma^2
  sample.varcov <- vcov(ymod)

  if (priors=='flat') {
    outcomeModResVar <- (sample.sigmaxsq * ymod$df)/rchisq(1,ymod$df)
    covariance_y <- (outcomeModResVar/sample.sigmaxsq) * vcov(ymod)
    outcomeModBeta = sample.beta +
      MASS::mvrnorm(1, mu = rep(0, ncol(covariance_y)), Sigma = covariance_y)
  }

  if (priors=='gustafson') {
    k4=2
    prior.est.sigmaxsq=1
    outcomeModResVar <- MCMCpack::rinvgamma(1, (ymod$df + k4)/2,
      (ymod$df*sample.sigmaxsq + k4*prior.est.sigmaxsq)/2)
    covariance_y <- (outcomeModResVar/sample.sigmaxsq) * vcov(ymod)
  }
}

```



```

        outcomeModBeta = sample.beta +
          MASS::mvrnorm(1, mu = rep(0, ncol(covariance_y)), Sigma = covariance_y)
      }
    }

#A MATRIX TO HOLD THE SUBSTANTIVE MODEL PARAMETERS FIT AT EACH CYCLE
if ((imp == 1) & (cyclenum == 1) & (var == 1)) {
  if (method[targetCol] == "latnorm") {
    #Replicate study
    smCoefIter <- array(0, dim = c(m, length(outcomeModBeta) +
      length(alpha) + 4, numit))
  } else {
    #Validation study
    smCoefIter <- array(0, dim = c(m, length(outcomeModBeta) +
      length(alpha) + 1, numit))
  }
}
if (method[targetCol] == "latnorm") {
  #Replicate study
  smCoefIter[imp, , cyclenum] <- c(outcomeModBeta, alpha, sigmaxsq,
    sigmausq, sigmaxwsq[1], lambda[1])
} else {
  #Validation study
  smCoefIter[imp, , cyclenum] <- c(outcomeModBeta, alpha, sigmaxwsq[1])
}

#A VECTOR SPECIFYING WITH ROWS NEED IMPUTING
imputationNeeded <- (1:n)[r[, targetCol] == 0]

#REJECTION SAMPLING
firstTryLimit <- ftl
j <- 1
while ((length(imputationNeeded) > 0) & (j <= firstTryLimit)) {
  #generate new potential x values
  imputations[[imp]][imputationNeeded, targetCol] <- rnorm(length(imputationNeeded),
    xfitted[imputationNeeded],
    sigmaxwsq^0.5)
  imputations[[imp]] <- updatePassiveVars(imputations[[imp]], method, passiveVars)

  #draw from a uniform distribution & apply relevant rejection rule
  uDraw <- runif(length(imputationNeeded))

  if (smtype == "logistic") {
    outmodxb <- model.matrix(as.formula(smformula),
      imputations[[imp]]) %% outcomeModBeta
    prob = expit(outmodxb[imputationNeeded])
    prob = prob * imputations[[imp]][imputationNeeded,
      outcomeCol] + (1 - prob) *
      (1 - imputations[[imp]][imputationNeeded, outcomeCol])
    reject = 1 * (uDraw > prob)
  } else if (smtype == "lm") {
    outmodxb <- model.matrix(as.formula(smformula),
      imputations[[imp]]) %% outcomeModBeta
    deviation <- imputations[[imp]][imputationNeeded,
      outcomeCol] - outmodxb[imputationNeeded]
    reject = 1 * (log(uDraw) > -(deviation^2)/(2 * array(outcomeModResVar,
      dim = c(length(imputationNeeded), 1))))
  }
  imputationNeeded <- imputationNeeded[reject == 1]
  j <- j + 1
}

for (i in imputationNeeded) {
  tries=firstTryLimit
  found = 0
  while (found==0 & tries<1000000) {
    tries=tries + rjlimit

    #More efficient search for difficult missing data
    tempData <- imputations[[imp]][i, ]
    tempData <- tempData[rep(1, rjlimit), ]
    if (method[targetCol] == "norm") {
      tempData[, targetCol] <- rnorm(rjlimit, xfitted[i], sigmaxwsq^0.5)
    } else if (method[targetCol] == "latnorm") {
      tempData[, targetCol] <- rnorm(rjlimit, xfitted[i], sigmaxwsq^0.5)
    }
    tempData <- updatePassiveVars(tempData, method, passiveVars)
    uDraw <- runif(rjlimit)

    if (smtype == "lm") {
      outmodxb <- model.matrix(as.formula(smformula), tempData) %% outcomeModBeta
      deviation <- tempData[, outcomeCol] - outmodxb
      reject = 1 * (log(uDraw) > -(deviation^2)/(2 * array(outcomeModResVar,
        dim = c(rjlimit, 1))))
    } else if (smtype == "logistic") {

```



```

cis <- sort(mime.beta + qt(c(0.025, 0.975), mime.beta.df)*mime.var.se)
return(list(c(mime.beta=mime.beta, mime.var.se=mime.var.se,
             lci=cis[1], uci=cis[2]), betas, betavars, smCoefIter))
}
#smcfc.function(d, target.beta, 100, 100, rep.vars=c('w1'), title='temp')[[1]]

#####
# INLA #
# Adapted from Muff et al 2015 #
# extended for use with validation studies #
#####

#Install INLA
source("http://www.math.ntnu.no/inla/givemeINLA.R")

#Generate priors to be used for both INLA and MCMC
gen.priors <- function(d, mod.family='binomial') {
  prior.beta0 <- c(0, 0.0001)
  if (mod.family=='binomial') prior.beta.x <- c(0, 1/(exp(1.39)))
  if (mod.family=='gaussian') prior.beta.x <- c(0, 0.0001)
  prior.alpha0 <- c(0, 0.001)

  k.x = 2 #effective sample size of tau.x prior
  k.u = 2 #effective sample size of tau.u prior
  est.tau.x = 1 #guess at value of precision of x (1/sigma.x)
  est.tau.u = 1 #guess at value of precision of u (1/sigma.u)

  #conditionally conjugate priors
  #based on Appendix of Gustafson 2004
  prior.tau.x <- c(k.x/2, 2*est.tau.x/k.x)
  prior.tau.u <- c(k.u/2, 2*est.tau.u/k.u)

  if (mod.family=='gaussian') prior.tau.y <- c(1, 1)

  return.priors <- c(prior.beta0, prior.beta.x, prior.alpha0, prior.tau.x, prior.tau.u)
  if (mod.family=='gaussian') return.priors <- c(return.priors, prior.tau.y)

  return(return.priors)
}

inla.function <- function(d, mod.family='binomial', rep.vars=c('w1','w2'),
                        correct=NA, prior=c(NA), hyperpar=F, threads=8) {
  require(INLA)

  dcenter <- d[,c('Y', rep.vars)]
  for (r in rep.vars) {
    dcenter[r] <- d[r] - mean(unlist(d[,rep.vars]), na.rm=T)
  }
  if (!is.na(Z[1])) dcenter$Z <- d[,Z]

  # validation study
  if ('Xobs' %in% names(d)) {
    dcenter$X <- d$Xobs
    dcenter$X <- dcenter$X - mean(dcenter$X, na.rm=T)
  } else dcenter$X <- NA

  n <- dim(dcenter)[1]
  n1 <- sum(!is.na(dcenter$X)) #number with validation data

  if (n1 > 0) {
    d1 <- dcenter[!is.na(dcenter$X),]
    d2 <- dcenter[is.na(dcenter$X),]
    dcenter <- data.frame(rbind(d1,d2))
  }

  if (is.na(prior[1])) prior <- gen.priors(d, mod.family)
  prior.beta0=prior[1:2]; prior.beta.x=prior[3:4]; prior.alpha0=prior[5:6];
  prior.tau.x=prior[7:8]; prior.tau.u=prior[9:10]
  if (mod.family=='gaussian') prior.tau.y=prior[15:16]

  tau.u <- 1 #initial values for the precision of x
  tau.x <- 1 #initial values for the precision of u
  if (mod.family=='gaussian') { tau.y <- 1 }

  # the response matrix Ymat and the data vectors are filled according to the
  # naming of the above joint model equation:
  # Y = coefbeta0*beta.0 + coefbetax*beta.x + idx.x + coefalpha0*alpha.0
  num_col <- 3 + ifelse(n1>0, 1, 0)
  num_row <- n*dim(dcenter)[2] + n1
  Ymat <- matrix(NA, nrow=num_row, ncol=num_col)
  Ymat[1:n, 1] <- dcenter$Y
  Ymat[n+(1:n), 2] <- rep(0, n) #latent x
  for (r in 1:length(rep.vars)) {
    Ymat[((1+r)*n+(1:n)), 3] <- dcenter[,rep.vars[r]] #w1 & w2
  }
}

```

```

if (n1 > 0) Ymat[(2+length(rep.vars))*n+(1:n1), 4] <- d1$x
Ymat <- Ymat

beta.0 <- c(rep(1, n), rep(rep(NA, n), 1 + length(rep.vars)), rep(NA, n1))
beta.x <- c(1:n, rep(rep(NA, n), 1 + length(rep.vars)), rep(NA, n1))
idx.x <- c(rep(NA, n), rep(1:n, 1 + length(rep.vars)), rep(NA, n1))
if (n1 > 0) idx.xobs <- c(rep(rep(NA, n), 2 + length(rep.vars)), 1:n1)

alpha.0 <- c(rep(NA, n), rep(1, n), rep(rep(NA, n), length(rep.vars)), rep(NA, n1))
weight.x <- c(rep(NA, n), rep(-1, n), rep(rep(1, n), length(rep.vars)), rep(NA, n1))
Ntrials <- c(rep(1, n), rep(rep(NA, n), 1 + length(rep.vars)), rep(NA, n1))

if (n1==0) {
  data.joint <- data.frame(Ymat=Ymat, beta.0=beta.0, beta.x=beta.x, idx.x=idx.x,
                          weight.x=weight.x, alpha0=alpha.0, Ntrials=Ntrials)
  formula <- Ymat ~ f(beta.x, copy = "idx.x",
                    hyper = list(beta = list(param = prior.beta.x, fixed = FALSE))) +
    f(idx.x, weight.x, model = "iid", values = 1:n,
      hyper = list(prec = list(initial = -15, fixed = TRUE))) +
    beta.0 - 1 + alpha.0
} else {
  data.joint <- data.frame(Ymat=Ymat, beta.0=beta.0, beta.x=beta.x, idx.x=idx.x,
                          weight.x=weight.x, alpha0=alpha.0, idx.xobs=idx.xobs,
                          Ntrials=Ntrials)
  formula <- Ymat ~ f(beta.x, copy = "idx.x",
                    hyper = list(beta = list(param = prior.beta.x, fixed = FALSE))) +
    f(idx.x, weight.x, model = "iid", values = 1:n,
      hyper = list(prec = list(initial = -15, fixed = TRUE))) +
    beta.0 - 1 + alpha.0 + f(idx.xobs, copy="idx.x",
                          hyper = list(beta = list(initial = 1, fixed = TRUE)))
}

control.family.list <- list(
  list(hyper = list()),
  list(hyper = list(
    prec = list(initial = log(tau.x),
                param = prior.tau.x,
                fixed = FALSE))),
  list(hyper = list(
    prec = list(initial=log(tau.u),
                param = prior.tau.u,
                fixed = FALSE))))
if (mod.family=='gaussian') {
  control.family.list[[1]] <- list(hyper = list(
    prec = list(initial = log(tau.y),
                param = prior.tau.y,
                fixed = FALSE)))
}
if (n1 > 0) control.family.list[[length(control.family.list)+1]] <- list(hyper = list())

control.fixed.list <- list(
  mean = list(beta.0=prior.beta0[1],
              alpha.0=prior.alpha0[1]),
  prec = list(beta.0=prior.beta0[2],
              alpha.0=prior.alpha0[2]))

#Explanation of options:
# family: There are three different likelihoods here, namely the binomial
#          likelihood of the regression model and two Gaussian likelihoods,
#          one for the exposure and one for the error model. They correspond to
#          the different columns in the response matrix Y.
# control.family: Specification of the hyperparameters for the three likelihoods,
#                 in the same order as given in family. The binomial likelihood does
#                 not contain any hyperparameters, thus the respective list
#                 is empty. In the second and third likelihoods the hyperparameters
#                 tau.x and tau.u need to be specified, respectively.
# control.fixed: Prior specification for the fixed effects

if (is.na(correct)) {
  r <- inla(formula,
            family = c(mod.family, rep("gaussian", 2 + ifelse(n1>0, 1, 0))),
            Ntrials = Ntrials, data = data.joint,
            control.family = control.family.list,
            control.fixed = control.fixed.list,
            num.threads = threads)
} else {
  r <- inla(formula,
            family = c(mod.family, rep("gaussian", 2 + ifelse(n1>0, 1, 0))),
            Ntrials = Ntrials, data = data.joint,
            control.family = control.family.list,
            control.fixed = control.fixed.list,
            control.inla = list(
              correct = TRUE,
              correct.factor = correct))
}

```

```

}
repeats=0
while (as.numeric(r$logfile[which(r$logfile=='Eigenvalues of the Hessian')+1])<0 &
  repeats < 50) {
  r <- inla.rerun(r)
  repeats = repeats + 1
  print(paste0("Negative Hessian value, repeat #", repeats))
}

if (hyperpar==T) r_more <- inla.hyperpar(r) else r_more=r

#The last call (inla.hyperpar) is not required. It is used to improve
# the estimates of the posterior marginals for the hyperparameters
# using a finer grid in the numerical integration.

beta.x <- r_more$summary.hyperpar['Beta for beta.x',1]

return(list(c(inla.beta=r_more$summary.hyperpar['Beta for beta.x',1],
  inla.se=r_more$summary.hyperpar['Beta for beta.x',2],
  lci=r_more$summary.hyperpar['Beta for beta.x',3],
  uci=r_more$summary.hyperpar['Beta for beta.x',5]), r_more))
}
inla.function(d, rep.vars=c('w1'))

#####
# MCMC #
# Using JAGS, 'rjags' and 'R2jags' wrapper package #
#####

mcmc.function <- function(d, mod.family='binomial', num.chains=3,
  n.adapt=20000, n.burn=25000, n.samples=20000, n.thin=20,
  num_reps=2, prior=c(NA)) {

  require(R2jags)
  require(coda)

  if (num_reps==1) {
    #Validation study
    d1 <- d[!is.na(d$xobs),]
    d2 <- d[is.na(d$xobs),]
    d <- data.frame(rbind(d1, d2))
    rm(d1, d2)
    n <- dim(d)[1]; n1=sum(!is.na(d$xobs))
  } else {
    #Replicate study
    n <- dim(d)[1]; n1=n
  }

  if (is.na(prior[1])) prior <- gen.priors(d, mod.family, Z)
  prior.beta0=prior[1:2]; prior.beta.x=prior[3:4]; prior.alpha0=prior[5:6];
  prior.tau.x=prior[7:8]; prior.tau.u=prior[9:10]
  if (mod.family=='gaussian') prior.tau.y=prior[11:12]

  inits.list <- vector("list", num.chains)
  for (i in 1:num.chains) {
    inits.list[[i]]$RNG.name <- "base:wichmann-hill"
    inits.list[[i]]$RNG.seed <- i
  }

  var.list <- list("Y"=d$Y, "w1"=d$w1,
    "Nobservations"=n,
    "betamean"=c(prior.beta0[1],prior.beta.x[1]),
    "betaprec"=diag(c(prior.beta0[2],prior.beta.x[2])),
    "alphamean"=prior.alpha0[1],
    "alphaprec"=prior.alpha0[2],
    "taux_a"=prior.tau.x[1], "taux_b"=prior.tau.x[2],
    "tauu_a"=prior.tau.u[1], "tauu_b"=prior.tau.u[2])

  if (num_reps==1) var.list <- c(var.list, list("X"=d$xobs), list("Nvalidated"=n1))
  if (num_reps==2) var.list <- c(var.list, list("w2"=d$w2), list("Nreplications"=n1))

  param.list <- c("alpha", "tau", "taux", "beta")
  if (mod.family=='gaussian') {
    var.list <- c(var.list, list("tauy_a"=prior.tau.y[1], "tauy_b"=prior.tau.y[2]))
    param.list <- c(param.list, "tauy")
  }

  mcmc.fit <- jags(data = var.list, inits=inits.list, param=param.list, n.thin=20,
    n.chains=num.chains, n.iter=n.burn, n.burnin=0,
    model=ifelse(mod.family=='binomial',
      ifelse(num_reps==1, "bugmods/valmodel_logistic.bug",
        "bugmods/repmodel_logistic.bug"),
      ifelse(num_reps==1, "bugmods/valmodel_linear.bug",

```

```

                                "bugmods/repmodel_linear.bug"))
mcmc.fit <- update(mcmc.fit, n.iter=n.samples, n.thin=n.thin)

rhat <- mcmc.fit$BUGSoutput$summary[, 'Rhat']
my.mcmc <- as.mcmc(mcmc.fit)
beta.x <- mcmc.fit$BUGSoutput$median$beta[2]
beta.x.ts.se <- summary(my.mcmc)[[1]]['beta[2]',4] #SE corrected for autocorrelation
rdr <- (1/mcmc.fit$BUGSoutput$median$taux)/((1/mcmc.fit$BUGSoutput$median$taux)+
                                             (1/mcmc.fit$BUGSoutput$median$tauu))

return(c(mcmc.beta=beta.x,
         mcmc.se=beta.x.ts.se,
         rhat=rhat[3],
         rdr=rdr,
         lci=interval[1], uci=interval[2]))
}

# file "valmodel_logistic.bug"
model {
  for (i in 1:Nobservations) {
    Y[i] ~ dbern(mu[i])
    logit(mu[i]) <- beta[1] + beta[2]*x[i]
    w1[i] ~ dnorm(x[i], tauu)
  }

  for (i in 1:Nvalidated) {
    x[i] <- X[i]
  }

  for (i in (Nvalidated+1):Nobservations) {
    x[i] ~ dnorm(mux[i], tauu)
    mux[i] <- alpha[1]
  }

  #Priors
  tauu ~ dgamma(tauu_a, tauu_b)
  alpha ~ dnorm(alphamean, alphaprec)
  taux ~ dgamma(taux_a, taux_b)
  beta ~ dnorm(betamean, betaprec)
}

# file "valmodel_linear.bug"
model {
  for (i in 1:Nobservations) {
    Y[i] ~ dnorm(mu[i], tauy)
    mu[i] <- beta[1] + beta[2]*x[i]
    w1[i] ~ dnorm(x[i], tauu)
  }

  for (i in 1:Nvalidated) {
    x[i] <- X[i]
  }

  for (i in (Nvalidated+1):Nobservations) {
    x[i] ~ dnorm(mux[i], tauu)
    mux[i] <- alpha[1]
  }

  #Priors
  tauu ~ dgamma(tauu_a, tauu_b)
  alpha ~ dnorm(alphamean, alphaprec)
  taux ~ dgamma(taux_a, taux_b)
  beta ~ dnorm(betamean, betaprec)
  tauy ~ dgamma(tauy_a, tauy_b)
}

# file "repmodel_logistic.bug"
model {
  for (i in 1:Nobservations) {
    Y[i] ~ dbern(mu[i])
    logit(mu[i]) <- beta[1] + beta[2]*x[i]

    X[i] ~ dnorm(mux[i], tauu)
    mux[i] <- alpha[1]
    w1[i] ~ dnorm(X[i], tauu)
  }

  for (i in 1:Nreplications) {
    w2[i] ~ dnorm(X[i], tauu)
  }

  #Priors
  tauu ~ dgamma(tauu_a, tauu_b)
  alpha ~ dnorm(alphamean, alphaprec)
  taux ~ dgamma(taux_a, taux_b)
  beta ~ dnorm(betamean, betaprec)
}

```

```

}
# file "repmoel_linear.bug"
model {
  for (i in 1:Nobservations) {
    Y[i] ~ dnorm(mu[i], tauy)
    mu[i] <- beta[1] + beta[2]*x[i]

    x[i] ~ dnorm(mux[i], taux)
    mux[i] <- alpha[1]
    w1[i] ~ dnorm(x[i], tauu)
  }

  for (i in 1:Nreplications) {
    w2[i] ~ dnorm(x[i], tauu)
  }

  #Priors
  tauu ~ dgamma(tauu_a, tauu_b)
  alpha ~ dnorm(alphamean, alphaprec)
  taux ~ dgamma(taux_a, taux_b)
  beta ~ dnorm(betamean, betaprec)
  tauy ~ dgamma(tauy_a, tauy_b)
}

```

## B. CODE FOR MEASUREMENT ERROR CORRECTION METHODS WITH A QUADRATIC TRANSFORMATION OF THE ERROR-PRONE EXPOSURE

The following code can be used to apply RC as presented in Section 4.4.1, Bayesian analysis using MCMC as presented in Section 4.4.2, Bayesian regression calibration using MCMC (MCMC-RC) as presented in Section 4.4.4, Bayesian regression calibration INLA as presented in Section 4.4.4, and selection of a quadratic model using Bayesian transformation selection (BTS) while correcting for measurement error as presented in Section 5.4.2. The code for each method may be applied with either a validation or a replicate study and either linear or logistic regression. This code was used for the simulation studies in Chapters 4 and 5.

```
#Chapter 4 example code

#####
# Logistic regression, validation study (30%)      #
# J-shaped association,                          #
# measurement error variance 1/4 the variance of X #
#####
set.seed(1989746)
n=2000
x_sd=1
m_error_sd=0.5
num_reps=1
prop_observed=0.3

X <- rnorm(n, 10, x_sd)
logit_Y <- -1.8 + 0.2*(X-9)^2
pr <- exp(logit_Y)/(1 + exp(logit_Y))
Y <- rbinom(n, 1, pr)
w1 <- X + rnorm(n, 0, m_error_sd)
set <- sample.int(n, size=floor(n*prop_observed), replace=F)
observed <- c(1:n)
observed[T] <- observed %in% set
xobs <- X; xobs[observed==F] <- NA

d <- data.frame(Y, w1, xobs)
rm(X, logit_Y, pr, w1, set, observed, xobs)

#####
# Linear regression, replicate study (100%)      #
# weak quadratic association,                   #
# measurement error equal to variance of X      #
#####
set.seed(1989746)
n=2000
x_sd=1
m_error_sd=1
num_reps=2
prop_observed=1

X <- rnorm(n, 10, x_sd)
Y <- 0.3*X^2 + rnorm(n, 0, 6)
w1 <- X + rnorm(n, 0, m_error_sd)
w2 <- X + rnorm(n, 0, m_error_sd)

d <- data.frame(Y,w1,w2)
rm(X, w1, w2)

#####
# Regression calibration #
#####
#rc.w function unchanged from previous chapter

rc.quad.boot <- function(data, indices, m1, m2, mod.family, Z=c(NA), Xvalues) {
  d <- data[indices,]
```



```

d$m1 <- d[,m1]
d$m2 <- d[,m2]

rc.x <- rc.w(d, m1, m2, Z)
mod2 <- glm(d$Y~fitted + I(fitted^2), family=mod.family, data=rc.x)
betaXsq <- mod2$coef[3]
quad.coefs <- c(mod2$coef[1] - betaXsq*rc.x$rc_xw_var, mod2$coef[-1])
quad.Xtvalues <- matrix(unlist(S.pow(Xvalues, c(1,2))), nrow=2, byrow=T)
my.Yval <- quad.coefs %%% rbind(rep(1, length(Xvalues)), quad.Xtvalues)

return(c(quad.coefs[2:3],my.Yval))
}

rc.quad.function <- function(d, Y, mod.family='binomial', m1='w1', m2='w2',
                             Xvalues=seq(7,13,by=0.1)) {
  rc.x <- rc.w(d, 'w1', ifelse(prop_observed<1, 'Xobs', 'w2'), Z=c(NA))
  rc.mod1 <- glm(Y~rc.x$fitted, family=mod.family)
  rc.mod2 <- glm(Y~rc.x$fitted + I(rc.x$fitted^2), family=mod.family)

  my.strata <- rep(1,length(Y)); if (prop_observed<1) my.strata[is.na(d$Xobs)] <- 2
  rc.boots <- boot(data=data.frame(cbind(d,Y)), statistic=rc.quad.boot, R=1000,
                  strata=my.strata, m1='w1', m2=ifelse(num_reps==1, 'Xobs', 'w2'),
                  mod.family=mod.family, Xvalues=Xvalues)

  rc.x <- rc.w(d, 'w1', ifelse(prop_observed<1, 'Xobs', 'w2'), Z=c(NA))
  ll <- lmtest::lrtest(rc.mod1,rc.mod2)

  lin.AIC <- AIC(rc.mod1)
  quad.AIC <- AIC(rc.mod2)
  AIC.diff = ifelse(quad.AIC<lin.AIC, 1, 0)

  bx1.boot.cis <- boot.ci(rc.boots, index=1, conf=0.95, type='basic')
  bx2.boot.cis <- boot.ci(rc.boots, index=2, conf=0.95, type='basic')

  quad.Xtvalues <- matrix(unlist(S.pow(Xvalues, c(1,2))), nrow=2, byrow=T)
  betaXsq <- rc.mod2$coef[3]
  quad.coefs <- c(rc.mod2$coef[1] - betaXsq*rc.x$rc_xw_var, rc.mod2$coef[-1])

  quad.Yval <- quad.coefs %%% rbind(rep(1, length(Xvalues)), quad.Xtvalues)
  lin.Yval <- rc.mod1$coef %%% rbind(rep(1, length(Xvalues)), Xvalues)

  return(list(lratio=ll$Pr[2],
             AIC.diff=AIC.diff,
             bx1.cis=bx1.boot.cis,
             bx2.cis=bx2.boot.cis,
             quad.coef=quad.coefs, lin.coef=rc.mod1$coef,
             quad.Yval=quad.Yval,
             lin.Yval=lin.Yval))
}

#rc.quad.function(d, Y, m2='Xobs')

#####
#MCMC quadratic #
#####
mcmc.quad <- function(d, Y, mod.family='gaussian', m1='w1', m2='w2',
                     Xvalues=seq(7,13,by=0.1),
                     n.adapt, n.burn, n.samples, n.thin, n.chains) {

  require(rjags)

  n <- dim(d)[1]
  d$m1 <- d[,m1]
  d$m2 <- d[,m2]

  n1 <- sum(!is.na(d$m2)) #number with validation data

  if (mod.family=='gaussian') {sdofY = sd(Y)} else {sdofY=1}
  if (m2=='Xobs') {
    sdofx <- sd(d$m2, na.rm=T)
    sdofxsq <- sd(d$m2^2, na.rm=T)
    mux <- mean(d$m2/sdofx, na.rm=T)
    muxsq <- mean(d$m2^2/sdofxsq, na.rm=T)

    #sort observed values of X to be first
    d$Y <- Y
    d1 <- d[!is.na(d$m2),]
    d2 <- d[is.na(d$m2),]
    d <- data.frame(rbind(d1,d2))
    Y <- d$Y
    d <- d[-length(d)]
    rm(d1,d2)
  } else {
    sdofx <- sqrt(cov(d$m1, d$m2))
    sdofxsq <- sqrt(cov(d$m1^2, d$m2^2))
    mux <- mean(c(d$m1, d$m2)/sdofx, na.rm=T)
    muxsq <- mean(c(d$m1, d$m2)^2/sdofxsq, na.rm=T)
  }
}

```

```

}
var.list <- list("w1"=(d$m1/sdofx)-mux, "Nobservations"=n,
               "betamean"=c(0, 0, 0),
               "betaprec"=diag(c(0.0001, 0.0001, 0.0001)),
               "alphamean"=c(0),
               "alphaprec"=c(0.0001),
               "taux_a"=1, "taux_b"=1,
               "tauu_a"=1, "tauu_b"=1,
               'sdofxsq'=sdofxsq, 'sdofx'=sdofx,
               'myx'=mux, 'myxsq'=muxsq)

if (mod.family=='gaussian') {
  var.list <- c(var.list, list("Y"=Y/sdofY, "tauy_a"=1, "tauy_b"=1, "sdofy"=sdofY))
} else {
  var.list <- c(var.list, list("Y"=Y))
  var.list$betaprec = diag(c(0.000001, 0.00001, 0.00001))
}

if (m2=='Xobs') {
  var.list <- c(var.list, list("Xobs"=(d$m2/sdofx)-mux, "nvalidated"=n1))
} else if (m2=='w2') {
  var.list <- c(var.list, list("w2"=(d$m2/sdofx)-mux, "Nreplications"=n1))
}

param.list <- c("betax", "deviance", "tau", "taux")
load.module("dic", quiet = TRUE)
inits.list <- vector("list", n.chains)
for (i in 1:n.chains) {
  inits.list[[i]]$RNG.name <- "base::wichmann-Hill"
  inits.list[[i]]$RNG.seed <- i
}

modfile <- ifelse(mod.family=='gaussian',
                 ifelse(m2=='Xobs', "bugmods/linquad_validation.bug",
                        "bugmods/linquad_replicate.bug"),
                 ifelse(m2=='Xobs', "bugmods/logquad_validation.bug",
                        "bugmods/logquad_replicate.bug"))

sim.fit <- jags.model(modfile, data = var.list, inits=inits.list,
                    n.chains = n.chains, n.adapt=n.adapt)
burn.samples <- coda.samples(sim.fit, param.list, n.burn, thin = n.thin)
samples <- coda.samples(sim.fit, param.list, n.samples, n.thin)

means <- summary(samples)[[1]][,1]
quad.Yval <- means['betax[1]'] + means['betax[2]']*Xvalues +
means['betax[3]']*Xvalues^2

dic.s <- dic.samples(sim.fit, n.iter = 1000, type = 'pD')
DIC <- (sum(dic.s$deviance) + sum(dic.s$penalty))

coefficients=c(means['betax[1]'],
               means['betax[2]'],
               means['betax[3]'])

bx2.cis <- HPDinterval(long.mcmc,0.95)['betax[3]',]
bx1.cis <- HPDinterval(long.mcmc,0.95)['betax[2]',]

out.list <- list(coef=coefficients,
                bx2.cis=bx2.cis, bx1.cis=bx1.cis,
                quad.Yval=quad.Yval,
                DIC=DIC,
                sim.fit=sim.fit)

return(out.list)
}
#mcmc.mod2 <- mcmc.quad(d, Y, mod.family="binomial", m1='w1', m2='Xobs',
#                      n.adapt=70000, n.burn=10000, n.samples=20000, n.thin=1,
#                      n.chains=3)

#file 'linquad_validation.bug'
model {
  for (i in 1:Nobservations) {
    Y[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta[1] + beta[2]*X[i] + beta[3]*Xt[i]

    Xt[i] <- (pow((X[i] + myx)*sdofx, 2)/sdofxsq) - myxsq
    w1[i] ~ dnorm(X[i], tauu)
  }
  for (i in 1:Nvalidated) {
    X[i] <- Xobs[i]
  }
  for (i in (Nvalidated+1):Nobservations) {
    X[i] ~ dnorm(mux[i], tau)
    mux[i] <- alpha[1]
  }
}

```

```

#Priors
tauu ~ dgamma(tauu_a, tauu_b)
alpha ~ dnorm(alphamean, alphaprec)
taux ~ dgamma(taux_a, taux_b)
beta ~ dnorm(betamean, betaprec)
tauy ~ dgamma(tauy_a, tauy_b)

#Unstandardize variables
betax[2] <- beta[2]*sdofy/sdofx
betax[3] <- beta[3]*sdofy/sdofxsq
betax[1] <- beta[1]*sdofy - (betax[2]*(myx*sdofx) + betax[3]*(myxsq*sdofxsq))
}

#file 'logquad_replicate.bug'
model {
  for (i in 1:Nobservations) {
    Y[i] ~ dbern(mu[i])
    logit(mu[i]) <- beta[1] + beta[2]*X[i] + beta[3]*xt[i]

    Xt[i] <- (pow(X[i]*sdofx + myx, 2)/sdofxsq) - myxsq
    X[i] ~ dnorm(mux[i], taux)
    mux[i] <- alpha

    w1[i] ~ dnorm(X[i], tauu)
  }
  for (i in 1:Nreplications) {
    w2[i] ~ dnorm(X[i], tauu)
  }

#Priors
tauu ~ dgamma(tauu_a, tauu_b)
alpha ~ dnorm(alphamean, alphaprec)
taux ~ dgamma(taux_a, taux_b)
beta ~ dnorm(betamean, betaprec)

#Unstandardize variables
betax[2] <- beta[2]/sdofx
betax[3] <- beta[3]/sdofxsq
betax[1] <- beta[1] - (betax[2]*(myx*sdofx) + betax[3]*(myxsq*sdofxsq))
}

#####
#MCMC-RC #
#####
#MCMC implementation of the classical error model (no substantive model)
mcmc.rc.function <- function(d, validation=T,
                             n.adapt, n.burn, n.samples, n.thin, n.chains) {

  require(rjags)
  n <- dim(d)[1]
  d <- cbind(d, 'orig.no'=1:n)

  if (validation != T) {
    dsort <- rbind(d[!is.na(d$w2),], d[is.na(d$w2),])
    cc <- as.vector(cbind(dsort$w1, dsort$w2))
    cc <- cc[!is.na(cc) & cc!=0]
    sdofw = sd(cc, na.rm=T)
    meanofw = mean(cc, na.rm=T)
    w1 = (dsort$w1-meanofw)/sdofw; w2 = (dsort$w2-meanofw)/sdofw
    var.list <- list("w1"=w1, "w2"=w2, "Nobservations"=n,
                   "Nreplications"=sum(!is.na(dsort$w2)),
                   "alphamean"=c(0),
                   "alphaprec"=c(0.0001),
                   "taux_a"=1, "taux_b"=1,
                   "tauu_a"=1, "tauu_b"=1)
  } else {
    n1 <- sum(!is.na(d$xobs))
    dsort <- rbind(d[!is.na(d$xobs),], d[is.na(d$xobs),])
    sdofx = sd(dsort$xobs, na.rm=T)
    meanofx = mean(dsort$xobs, na.rm=T)
    w1 = (dsort$w1-meanofx)/sdofx; Xobs = (dsort$xobs-meanofx)/sdofx
    var.list <- list("w1"=w1, "Xobs"=Xobs, "Nobservations"=n, "Nvalidated"=n1,
                   "alphamean"=c(0),
                   "alphaprec"=c(0.0001),
                   "taux_a"=1, "taux_b"=1,
                   "tauu_a"=1, "tauu_b"=1)
  }

  num.chains=n.chains
  inits.list <- vector("list", num.chains)
  for (j in 1:num.chains) {
    inits.list[[j]] <- list(.RNG.name="base::wichmann-Hill", .RNG.seed=j)
  }

  param.list <- c("alpha", "tauu", "taux", "X")
}

```

```

modfile <- ifelse(validation != T, "bugmods/classical.bug",
"bugmods/classical_val.bug")

sim.fit <- jags.model(modfile, data = var.list, inits=inits.list,
n.chains = n.chains, n.adapt=n.adapt)
burn.samples <- coda.samples(sim.fit, param.list, n.burn, n.thin)
samples <- coda.samples(sim.fit, param.list, n.samples, n.thin)

#Isolate x samples and put back in original order
X.samples <- lapply(samples, function(x) x[,1:n][,order(dsort$orig.no)])

if (n.chains > 1) {
  for (i in 2:n.chains) {
    X.samples[[1]] <- rbind(X.samples[[1]], X.samples[[i]])
  }
}
X.samples <- as.matrix(X.samples[[1]])
if (validation != T) X.samples = X.samples*sdfw + meanofw else {
  X.samples = X.samples*sdfX + meanofX}

return(X.samples)
}

X.samples <- mcmc.rc.function(d, validation = T,
n.adapt = 30000, n.burn = 10000, n.samples=6000,
n.thin=1, n.chains=3)

#Transform samples and find expectations
Xsq.samples = X.samples^2
mean.X <- apply(X.samples, 2, mean)
mean.Xsq <- apply(Xsq.samples, 2, mean)

#Fit linear and quadratic models (regression calibration)
mod1 <- glm(Y~mean.X, family=mod.family)
mod2 <- glm(Y~mean.X+mean.Xsq, family=mod.family)

ll <- lmtest::lrtest(mod1,mod2)
cis <- confint.default(mod2)
AIC.diff = ifelse(AIC(mod2)<AIC(mod1), 1, 0)

quad.Xtvalues <- matrix(unlist(S.pow(Xvalues, c(1,2))), nrow=2, byrow=T)
quad.Yval <- predict(mod2, newdata=data.frame(mean.X=Xvalues,
mean.Xsq=quad.Xtvalues[2,]),
se.fit=F, response='link')

#file 'classical.bug'
model {
  for (i in 1:Nobservations) {
    X[i] ~ dnorm(mux[i], taux)
    mux[i] <- alpha[1]
    w1[i] ~ dnorm(X[i], tauu)
  }
  for (i in 1:Nreplications) {
    w2[i] ~ dnorm(X[i], tauu)
  }
  tauu ~ dgamma(tauu_a, tauu_b)
  taux ~ dgamma(taux_a, taux_b)
  alpha ~ dnorm(alphamean,alphaprec)
}

#file 'classical_val.bug'
model {
  for (i in 1:Nobservations) {
    w1[i] ~ dnorm(X[i], tauu)
  }
  for (i in 1:Nvalidated) {
    X[i] <- Xobs[i]
  }
  for (i in (Nvalidated+1):Nobservations) {
    X[i] ~ dnorm(mux[i], taux)
    mux[i] <- alpha[1]
  }
  tauu ~ dgamma(tauu_a, tauu_b)
  taux ~ dgamma(taux_a, taux_b)
  alpha ~ dnorm(alphamean,alphaprec)
}

#####
#INLA-RC #
#####
#MCMC implementation of the classical error model (no substantive model)

inla.rc.function <- function(d, rep.vars=c('w1','w2'), validation=F,
n.samples) {

```

```

require(INLA)

n <- dim(d)[1]
if (validation==T) n1 <- sum(!is.na(d$Xobs)) else n1=0 #number with validation data

dcenter <- cbind(d[,c('X', rep.vars)], 'orig.no'=1:n)
if ('Xobs' %in% names(d)) {
  shift = mean(d$Xobs, na.rm=T)
  scale = sd(d$Xobs, na.rm=T)
  dcenter$X <- (d$Xobs - shift)/scale
} else {
  shift = mean(unlist(d[,rep.vars]), na.rm=T)
  scale = sqrt(cov(d[,rep.vars[1]], d[,rep.vars[2]]))
  dcenter$X <- NA
}
for (r in rep.vars) {
  dcenter[r] <- (d[r] - shift)/scale
}

if (n1 > 0) {
  d1 <- dcenter[!is.na(dcenter$X),]
  d2 <- dcenter[is.na(dcenter$X),]
  dcenter <- data.frame(rbind(d1,d2))
}

tau.u <- 1 #initial values for the precision of x
tau.x <- 1 #initial values for the precision of u

num_col <- 2 + ifelse(n1 > 0, 1, 0)
num_row <- n*(1 + length(rep.vars)) + ifelse(n1 > 0, n1, 0)
Xmat <- matrix(NA, nrow=num_row, ncol=num_col)
Xmat[1:n, 1] <- rep(0, n) #latent x
for (r in 1:length(rep.vars)) {
  Xmat[((r)*n+(1:n)), 2] <- dcenter[,rep.vars[r]] #w1 & w2
}
if (n1 > 0) {Xmat[(1+length(rep.vars))*n+1:n1, 3] <- dcenter$X[1:n1]}
Xmat <- Xmat

#data vectors
idx.x <- c(rep(1:n, 1 + length(rep.vars))); if (validation==T) idx.x <- c(idx.x, 1:n1)
alpha.0 <- c(rep(1, n), rep(rep(NA, n), length(rep.vars))); if (validation==T) alpha.0
<- c(alpha.0, rep(NA, n1))
weight.x <- c(rep(-1, n), rep(rep(1, n), length(rep.vars))); if (validation==T)
weight.x <- c(weight.x, rep(1,n1))

data.joint <- data.frame(Xmat=Xmat, idx.x=idx.x,
                        weight.x=weight.x, alpha0=alpha.0)
formula <- Xmat ~ f(idx.x, weight.x, model = "iid", values = 1:n,
                    hyper = list(prec = list(initial = -15, fixed = TRUE))) - 1 +
alpha.0

control.family.list <- list(
  list(hyper = list(
    prec = list(initial = tau.x,
                param = c(1,1),
                fixed = FALSE))),
  list(hyper = list(
    prec = list(initial=tau.u,
                param = c(1,1),
                fixed = FALSE)))
)
if (n1 > 0) control.family.list[[length(control.family.list)+1]] <- list(hyper =
list())

control.fixed.list <- list(
  mean = list(alpha.0=0),
  prec = list(alpha.0=0.0001))

r <- inla(formula,
          family = c(rep("gaussian", 2+ifelse(n1 > 0, 1, 0))),
          data = data.joint,
          control.family = control.family.list,
          control.fixed = control.fixed.list,
          num.threads = 8,
          control.compute = list(config=TRUE))
summary(r)

post.a0 <- r$summary.fixed[,'mean']+shift
post.taux <- r$summary.hyperpar[1,'mean']
post.tauu <- r$summary.hyperpar[2,'mean']
if (n1 > 0) post.tauxobs <- r$summary.hyperpar[3,'mean'] #should be very high
indicating a very low variance

post.vars <- c(post.a0, post.taux, post.tauu)
names(post.vars) <- c('a0', 'taux', 'tauu')
if (n1 > 0) {post.vars <- c(post.vars, post.tauxobs); names(post.vars)[4] <-
'tauxobs'}

```

```

print(post.vars)

samples = inla.posterior.sample(n = n.samples, result = r)
X.samples <- sapply(samples, FUN= function(x) x$latent[(n+1):(2*n)])
#NOTE: in samples[[s]]$latent, the 1:n are the -X values in the equation 0 = -X + a0
# the (n+1):(2n) are the X in the w1 = X + U, also
4001:6001 for replicate w2 = X + U
# the (2n+1):(2n+1) are the Xobs = X values for
validation
# for validation (2n+1):(3n+1) are the a0 values as a
part of 0 = -X + a0
# for replicate (3n+1):(4n) are the a0 values
X.samples <- t(X.samples[order(dcenter$orig.no),]*scale+shift)

return(X.samples)
}

#Transform samples and find expectations
Xsq.samples = X.samples^2
mean.X <- apply(X.samples, 2, mean)
mean.Xsq <- apply(Xsq.samples, 2, mean)

#Fit linear and quadratic models (regression calibration)
mod1 <- glm(Y~mean.X, family=mod.family)
mod2 <- glm(Y~mean.X+mean.Xsq, family=mod.family)

ll <- lmtest::lrtest(mod1,mod2)
cis <- confint.default(mod2)
AIC.diff = ifelse(AIC(mod2)<AIC(mod1), 1, 0)

quad.Xtvalues <- matrix(unlist(S.pow(Xvalues, c(1,2))), nrow=2, byrow=T)
quad.Yval <- predict(mod2, newdata=data.frame(mean.X=Xvalues,
mean.Xsq=quad.Xtvalues[2,]),
se.fit=F, response='link')

#####
#Implementation of Kuo & Mallick method #
#of Bayesian variable selection #
#for selecting the quadratic model vs #
#the linear model #
#####

mcmc.quad.km <- function(d, Y, mod.family='gaussian',
m1='w1', m2='Xobs', xvalues=seq(7,13,by=0.1),
pi=0.5,
n.adapt, n.burn, n.samples, n.thin, n.chains) {

require(rjags)

n <- dim(d)[1]
d$m1 <- d[,m1]
d$m2 <- d[,m2]

n1 <- sum(!is.na(d$m2)) #number with validation data

if (mod.family=='gaussian') {sdofY = sd(Y)} else {sdofY=1}
if (m2=='Xobs') {
sdofx <- sd(d$m2, na.rm=T)
sdofxsq <- sd(d$m2^2, na.rm=T)
mux <- mean(d$m2/sdofx, na.rm=T)
muxsq <- mean(d$m2^2/sdofxsq, na.rm=T)

d$Y <- Y
d1 <- d[!is.na(d$m2),]
d2 <- d[is.na(d$m2),]
d <- data.frame(rbind(d1,d2))
Y <- d$Y
d <- d[-length(d)]
rm(d1,d2)
} else {
sdofx <- sqrt(cov(d$m1, d$m2))
sdofxsq <- sqrt(cov(d$m1^2, d$m2^2))
mux <- mean(c(d$m1, d$m2)/sdofx, na.rm=T)
muxsq <- mean(c(d$m1, d$m2)^2/sdofxsq, na.rm=T)
}

var.list <- list("w1"=(d$m1/sdofx)-mux, "Nobservations"=n,
"betamean"=c(0, 0, 0),
"betaprec"=diag(c(0.0001, 0.0001, 0.01)),
"alphamean"=c(0),
"alphaprec"=c(0.0001),
"taux_a"=1, "taux_b"=1,
"tauu_a"=1, "tauu_b"=1,
'sdofxsq'=sdofxsq, 'sdofx'=sdofx,
'myx'=mux, 'myxsq'=muxsq,

```

```

      "pi"=pi)

if (mod.family=='gaussian') {
  var.list <- c(var.list, list("Y"=Y/sdofY, "tauy_a"=1, "tauy_b"=1, "sdofY"=sdofY))
} else {
  var.list <- c(var.list, list("Y"=Y))
  var.list$betaprec = diag(c(0.000001, 0.00001, 0.01))
}

if (m2=='Xobs') {
  var.list <- c(var.list, list("Xobs"=(d$m2/sdofx)-mux, "nvalidated"=n1))
} else if (m2=='W2') {
  var.list <- c(var.list, list("W2"=(d$m2/sdofx)-mux, "Nreplications"=n1))
}

param.list <- c("alpha", "betax", "deviance", "indsq")
load.module("dic", quiet = TRUE)

inits.list <- vector("list", n.chains)
for (i in 1:n.chains) {
  inits.list[[i]] <- list('indsq'=ifelse((i %% 2)==0, 1, 0),
                        .RNG.name="base::Wichmann-Hill",
                        .RNG.seed=i)
}
modfile <- ifelse(mod.family=="gaussian",
                 ifelse(m2=='Xobs', "bugmods/linquad-opt_validation.bug",
                        "bugmods/linquad-opt_replicate.bug"),
                 ifelse(m2=='Xobs', "bugmods/logquad-opt_validation.bug",
                        "bugmods/logquad-opt_replicate.bug"))

sim.fit <- jags.model(modfile, data = var.list, inits=inits.list,
                    n.chains = n.chains, n.adapt=n.adapt)
burn.samples <- coda.samples(sim.fit, param.list, n.burn, thin = n.thin)
samples <- coda.samples(sim.fit, param.list, n.samples, n.thin)

samples <- bvs.mod[[1]]
sims.mat1 <- samples[[1]][samples[[1]][,'indsq']==1,]
sims.mat0 <- samples[[1]][samples[[1]][,'indsq']==0,]
for (cc in 2:n.chains) {
  sims.mat1 <- rbind(sims.mat1, samples[[cc]][samples[[cc]][,'indsq']==1,])
  sims.mat0 <- rbind(sims.mat0, samples[[cc]][samples[[cc]][,'indsq']==0,])
}
if (nrow(sims.mat1) == 0) stop('Error: None of the samples selected indsq==1')

sims.mat1 <- mcmc(sims.mat1)
sims.mat0 <- mcmc(sims.mat0)

means1 <- summary(sims.mat1)[[1]][,1]
quad.Yval <- means1['betax[1]'] + means1['betax[2]']*xvalues +
means1['betax[3]']*xvalues^2
means0 <- summary(sims.mat0)[[1]][,1]
lin.Yval <- means0['betax[1]'] + means0['betax[1]']*xvalues

quad.coef=c(means1['betax[1]'],
            means1['betax[2]'],
            means1['betax[3]'])
lin.coef=c(means0['betax[1]'],
           means0['betax[2]'])

quad.cov = cov(sims.mat1[,grep('betax', colnames(sims.mat1))])
lin.cov = cov(sims.mat0[,grep('betax', colnames(sims.mat0))])

bx2.cis <- HPDinterval(sims.mat1[, 'betax[3]', ], 0.95)
bx1.cis <- HPDinterval(sims.mat1[, 'betax[3]', ], 0.95)

out.list <- list(indsq=summary(samples)[[1]][,'indsq',1],
                quad.coef=quad.coef, lin.coef=lin.coef,
                bx2.cis=bx2.cis, bx1.cis=bx1.cis,
                quad.cov=quad.cov, lin.cov=lin.cov,
                quad.Yval=quad.Yval, lin.Yval=lin.Yval)
return(out.list)
}

```

## C. CODE FOR MEASUREMENT ERROR CORRECTION IN THE CONTEXT OF THE FRACTIONAL POLYNOMIAL METHOD

The following code can be used to apply RC as presented in Section 6.4.1, Bayesian regression calibration using MCMC (MCMC-RC) and INLA (INLA-RC) as presented in Section 6.4.2, and selection of the best fractional polynomial model using Bayesian transformation selection (BTS) while correcting for measurement error as presented in Section 6.4.3. The code for each method may be applied with either a validation or a replicate study and either linear or logistic regression. This code was used for the simulation studies in Chapters 6.

```
#####  
# Logistic regression, validation study (30%) #  
# Weak quadratic association, #  
# measurement error variance 1/4 the variance of X #  
#####  
set.seed(1989746)  
n=2000  
x_sd=1  
m_error_sd=0.5  
num_reps=1  
prop_observed=0.3  
  
X <- rnorm(n, 10, x_sd)  
logit_Y <- -7.8 + 0.06*X^2  
pr <- exp(logit_Y)/(1 + exp(logit_Y))  
Y <- rbinom(n, 1, pr)  
w1 <- X + rnorm(n, 0, m_error_sd)  
set <- sample.int(n, size=floor(n*prop_observed), replace=F)  
observed <- c(1:n)  
observed[T] <- observed %in% set  
Xobs <- X; Xobs[observed==F] <- NA  
  
d <- data.frame(Y, w1, Xobs)  
rm(X, logit_Y, pr, w1, set, observed, Xobs)  
  
#####  
# Linear regression, replicate study (100%) #  
# Asymptotic association, #  
# measurement error equal to variance of X #  
#####  
set.seed(1989746)  
n=2000  
x_sd=1  
m_error_sd=1  
num_reps=2  
prop_observed=1  
  
X <- rnorm(n, 10, x_sd)  
Y <- (4-X)^-1  
w1 <- X + rnorm(n, 0, m_error_sd)  
w2 <- X + rnorm(n, 0, m_error_sd)  
  
d <- data.frame(Y,w1,w2)  
rm(X, w1, w2)  
  
#Values over which predicted outcome values are estimated  
Xvalues <- seq(7, 13, by=0.1)  
#Yvalues is the application of the underlying model to the Xvalues  
# these values (Xvalues & Yvalues) are used in the functions  
  
#####  
# Helper functions for fractional polynomials #  
#####  
require(mfp)  
require(rjags)  
require(MASS)
```



```

require(boot)
S <- c(-2, -1, -0.5, 0, 0.5, 1, 2, 3)
# Returns either h(x, p) (Equation 1.5) or h'(x, p1, p2) (Equation 1.6)
S.pow <- function(x, p) {
  if (length(p)==1) {
    if (p != 0) return(x^p) else return(log(x))
  } else if (length(p)==2) {
    x1 <- S.pow(x, p[1])
    if (p[1]!=p[2]) {
      x2 <- S.pow(x, p[2])
    } else {
      x2 <- x1*log(x)
    }
  }
  return(list(x1,x2))
} else {
  stop("p must be length 1 or 2")
}
}

# Creates of a matrix of combinations of 'p' for FP2 models
p.mat <- t(cbind(combn(S,2),rbind(S,S)))
p.mat <- p.mat[order(p.mat[,1], p.mat[,2]),]

#####
# Regression calibration #
#####
rc.fp.boot <- function(data, indices, p, m1, m2, mod.family, xvalues) {
  # Bootstraps just the chosen model
  # Returns model coefficients and predicted outcome values over exposure range "xvalues)
  # data needs to include Y
  # p is the chosen powers

  d <- data[indices,]

  w1 <- d[,m1]
  w2 <- d[,m2]

  Y <- d$Y
  n = length(w1)

  rcmod <- lm(log(w2)~log(w1))
  sigma_lnxsq <- cov(log(w1), log(w2), use="na.or.complete")
  mu_lnx <- mean(c(log(w1), log(w2)), na.rm=T)
  lambda <- rcmod$coefficients[2]

  kp <- function(power) {exp((1-lambda)*((sigma_lnxsq*power^2)/2 + power*mu_lnx))}
  Exp.Xp <- function(w, power) kp(power)*w^(lambda*power)
  Exp.logX <- function(w) log(w^(lambda)) + (1-lambda)*mu_lnx
  Exp.XplogX <- function(w, power) kp(power)*
    ((w^(lambda*power))*log(w^lambda) +
    (1-lambda)*(mu_lnx + power*sigma_lnxsq)*(w^(lambda*power)))
  Exp.logXlogX <- function(w) (1-lambda)*(sigma_lnxsq + (1-lambda)*mu_lnx^2) +
    log(w^lambda)*(log(w^lambda) + 2*(1-lambda)*mu_lnx)

  if (length(p)==1) {
    if (p != 0) ExpXt <- Exp.Xp(w1, p) else ExpXt <- Exp.logX(w1)
    mod <- glm(Y ~ ExpXt, family=mod.family)
  } else if (length(p)==2) {
    if (p[1] != 0) ExpXt1 <- Exp.Xp(w1, p[1]) else ExpXt1 <- Exp.logX(w1)
    if (p[2]==p[1]) {
      if (p[1] != 0) ExpXt2 <- Exp.XplogX(w1, p[2]) else ExpXt2 <- Exp.logXlogX(w1)
    } else {
      if (p[2] != 0) ExpXt2 <- Exp.Xp(w1, p[2]) else ExpXt2 <- Exp.logX(w1)
    }
    mod <- glm(Y ~ ExpXt1 + ExpXt2, family=mod.family)
  }

  Xtvalues <- matrix(unlist(S.pow(Xvalues, p)), nrow=length(p), byrow=T)
  my.Yval <- mod$coef %%% rbind(rep(1, length(Xvalues)), Xtvalues)

  return(c(mod$coef, my.Yval))
}

rc.fp <- function(d, Y, mod.family='binomial', m1="w1", m2="xobs", prop_observed=0.3,
  Xvalues, Yvalues) {
  w1 <- d[,m1]
  w2 <- d[,m2]
  n = length(w1)

  if (m2=="xobs") {
    sigma_lnxsq <- var(log(d[,m2]), na.rm=T)
    sigmausq <- var(log(w1)) - sigma_lnxsq
  } else {
    sigma_lnxsq <- cov(log(w1), log(w2), use="na.or.complete")
  }
}

```

```

    sigmausq <- var(log(w1) - log(w2), na.rm = T)/2
  }
  lambda <- lm(log(w2)~log(w1))$coef[2]
  mu_lnx <- mean(c(log(w1), log(w2)), na.rm=T)

  print(paste0("Est RC lambda: ", lambda))

  kp <- function(power) {exp((1-lambda)*((sigma_lnxsq*power^2)/2 + power*mu_lnx))}
  Exp.Xp <- function(w, power) kp(power)*w^(lambda*power)
  Exp.logX <- function(w) log(w^lambda) + (1-lambda)*mu_lnx
  Exp.XplogX <- function(w, power) kp(power)*
    ((w^(lambda*power))*log(w^lambda) +
    (1-lambda)*(mu_lnx + power*sigma_lnxsq)*(w^(lambda*power)))
  Exp.logXlogX <- function(w) (1-lambda)*(sigma_lnxsq + (1-lambda)*mu_lnx^2) +
    log(w^lambda)*(log(w^lambda) + 2*(1-lambda)*mu_lnx)

  nullmod <- glm(Y ~ 1, family=mod.family)
  null.dev <- nullmod$dev

  fp1.devs <- c()
  for (p in 1:length(S)) {
    if (S[p] != 0) ExpXt <- Exp.Xp(w1, S[p]) else ExpXt <- Exp.logX(w1)
    mod <- glm(Y ~ ExpXt, family=mod.family)
    fp1.devs[p] <- mod$deviance
    if (S[p]==1) {
      lin.dev <- mod$dev
      lin.mod <- mod
    }
  }

  fp1.p <- S[fp1.devs==min(fp1.devs)]
  if (aic.select==T) fp1.p <- S[fp1.aics==min(fp1.aics)]
  if (fp1.p!=0) {
    fp1.Xt = Exp.Xp(w1, fp1.p)
  } else {
    fp1.Xt = Exp.logX(w1)
  }
  fp1.mod <- glm(Y~fp1.Xt, family=mod.family, data=d)
  fp1.dev <- fp1.mod$deviance

  # Note pvalue comparison is chisq using 2 df to FP1 and 3 to linear and 4 to null
  fp2.devs <- c()
  for (pp in 1:dim(p.mat)[1]) {
    if (p.mat[pp,1] != 0) ExpXt1 <- Exp.Xp(w1, p.mat[pp,1]) else {
      ExpXt1 <- Exp.logX(w1)}
    if (p.mat[pp,2]==p.mat[pp,1]) {
      if (p.mat[pp,1] != 0) ExpXt2 <- Exp.XplogX(w1, p.mat[pp,2]) else {
        ExpXt2 <- Exp.logXlogX(w1)}
    } else {
      if (p.mat[pp,2] != 0) ExpXt2 <- Exp.Xp(w1, p.mat[pp,2]) else {
        ExpXt2 <- Exp.logX(w1) }
    }
    mod <- glm(Y ~ ExpXt1 + ExpXt2, family=mod.family)
    fp2.devs[pp] <- mod$deviance
  }

  fp2.p <- p.mat[fp2.devs==min(fp2.devs),]
  if (fp2.p[1]!=0) Xt1 = Exp.Xp(w1, fp2.p[1]) else Xt1 <- Exp.logX(w1)
  if (fp2.p[2]==fp2.p[1]) {
    if (fp2.p[1]!=0) Xt2 = Exp.XplogX(w1, fp2.p[2]) else Xt2 <- Exp.logXlogX(w1)
  } else {
    if (fp2.p[2]!=0) Xt2 = Exp.Xp(w1, fp2.p[2]) else Xt2 <- Exp.logX(w1)
  }
  fp2.mod <- glm(Y~Xt1+Xt2, family=mod.family, data=d)
  fp2.dev <- fp2.mod$dev

  dispersion <- if (mod.family=='gaussian') {
    if (lin.mod$df.residual > 0) sum(lin.mod$residuals^2)/lin.mod$df.residual
  } else 1

  p.null1 <- pchisq((null.dev - fp1.dev)/dispersion, df=2, lower.tail=F)
  p.null2 <- pchisq((null.dev - fp2.dev)/dispersion, df=4, lower.tail=F)
  p.lin1 <- pchisq((lin.dev - fp1.dev)/dispersion, df=1, lower.tail=F)
  p.lin2 <- pchisq((lin.dev - fp2.dev)/dispersion, df=3, lower.tail=F)
  p.fp <- pchisq((fp1.dev - fp2.dev)/dispersion, df=2, lower.tail=F)

  my.strata <- rep(1,length(Y))
  if (prop_observed<1) my.strata[is.na(d$Xobs)] <- 2
  lin.boots <- boot(data=data.frame(cbind(d,Y)), statistic=rc.fp.boot, R=1000,
    strata=my.strata, p=c(1), m1='w1',
    m2=ifelse(prop_observed<1, 'Xobs', 'w2'),
    mod.family=mod.family, xvalues=xvalues)
  fp1.boots <- boot(data=data.frame(cbind(d,Y)), statistic=rc.fp.boot, R=1000,
    strata=my.strata, p=fp1.p, m1='w1',
    m2=ifelse(prop_observed<1, 'Xobs', 'w2'),

```

```

        mod.family=mod.family, Xvalues=Xvalues)
fp2.boots <- boot(data=data.frame(cbind(d,Y)), statistic=rc.fp.boot, R=1000,
                 strata=my.strata, p=fp2.p, m1='w1',
                 m2=ifelse(prop_observed<1, 'Xobs', 'w2'),
                 mod.family=mod.family, Xvalues=Xvalues)

lin.cis <- rbind(boot.ci(lin.boots,index=1,conf=0.95,type=c('basic'))$basic[4:5],
                boot.ci(lin.boots,index=2,conf=0.95,type=c('basic'))$basic[4:5])
fp1.cis <- rbind(boot.ci(fp1.boots,index=1,conf=0.95,type=c('basic'))$basic[4:5],
                boot.ci(fp1.boots,index=2,conf=0.95,type=c('basic'))$basic[4:5])
fp2.cis <- rbind(boot.ci(fp2.boots,index=1,conf=0.95,type=c('basic'))$basic[4:5],
                boot.ci(fp2.boots,index=2,conf=0.95,type=c('basic'))$basic[4:5],
                boot.ci(fp2.boots,index=3,conf=0.95,type=c('basic'))$basic[4:5])

lin.vcov <- cov(lin.boots$t[,1:2])
fp1.vcov <- cov(fp1.boots$t[,1:2])
fp2.vcov <- cov(fp2.boots$t[,1:3])

lin.coef <- lin.mod$coef; fp1.coef <- fp1.mod$coef; fp2.coef <- fp2.mod$coef

lin.Yval <- lin.coef %%% rbind(rep(1, length(Xvalues)), Xvalues)
fp1.Yval <- fp1.coef %%% rbind(rep(1, length(Xvalues)), S.pow(Xvalues, fp1.p))
fp2.Yval <- fp2.coef %%% rbind(rep(1, length(Xvalues)),
                               matrix(unlist(S.pow(Xvalues, fp2.p)), nrow=2, byrow=T))

lin.Yval.cis <- fp1.Yval.cis <- fp2.Yval.cis <- matrix(nrow=2, ncol=length(Xvalues))
for (cc in 1:length(Xvalues)) {
  lin.Yval.cis[1:2,cc] <- boot.ci(lin.boots,index=2+cc,
                                conf=0.95, type='basic')$basic[4:5]
  fp1.Yval.cis[1:2,cc] <- boot.ci(fp1.boots,index=2+cc,
                                conf=0.95, type='basic')$basic[4:5]
  fp2.Yval.cis[1:2,cc] <- boot.ci(fp2.boots,index=3+cc,
                                conf=0.95, type='basic')$basic[4:5]
}

lin.mise <- sum((Yvalues - lin.Yval)^2)/length(Yvalues)
fp1.mise <- sum((Yvalues - fp1.Yval)^2)/length(Yvalues)
fp2.mise <- sum((Yvalues - fp2.Yval)^2)/length(Yvalues)

return(list(p.null1=p.null1,p.null2=p.null2,p.lin1=p.lin1,p.lin2=p.lin2,p.fp=p.fp,
           fp1.p=fp1.p, fp2.p=fp2.p,
           lin.coef=lin.mod$coef, fp1.coef=fp1.mod$coef, fp2.coef=fp2.mod$coef,
           lin.cis=lin.cis, fp1.cis=fp1.cis, fp2.cis=fp2.cis,
           lin.dev=lin.dev, fp1.dev=fp1.dev, fp2.dev=fp2.dev,
           lin.mise=lin.mise, fp1.mise=fp1.mise, fp2.mise=fp2.mise,
           lin.vcov=lin.vcov, fp1.vcov=fp1.vcov, fp2.vcov=fp2.vcov,
           lin.Yval=lin.Yval, fp1.Yval=fp1.Yval, fp2.Yval=fp2.Yval,
           lin.Yval.cis=lin.Yval.cis, fp1.Yval.cis=fp1.Yval.cis,
           fp2.Yval.cis=fp2.Yval.cis))
}
#rc.fp(d, Y, mod.family='gaussian', m1="w1", m2="Xobs", prop_observed=0.3,
#      Xvalues, Yvalues)

#####
#Fully Bayesian analysis #
#####

# BTS "FP1" method
mcmc.fp1 <- function(d, Y, mod.family='gaussian', m1='w1', m2='Xobs', pi=1/32,
                    n.adapt, n.burn, chains=c(1:5), uniform=F) {
  n.chains=length(chains)
  require(rjags)

  n <- dim(d)[1]
  n1 <- sum(!is.na(d$m2)) #number with validation data
  d$m1 <- d[,m1]
  d$m2 <- d[,m2]

  if (mod.family=='gaussian') {sdoFY = sd(Y)} else {sdoFY=1}
  if (m2=='Xobs') {
    d$Y <- Y
    d1 <- d[!is.na(d$m2),]
    d2 <- d[is.na(d$m2),]
    d <- data.frame(rbind(d1,d2))
    Y <- d$Y
    d <- d[-length(d)]
    rm(d1,d2)

    sdofx <- sd(d$m2, na.rm=T)
    mux <- mean(d$m2/sdofx, na.rm=T)
  } else {
    sdofx <- sqrt(cov(d$m1, d$m2))
    mux <- mean(c(d$m1, d$m2)/sdofx, na.rm=T)
  }

  xt <- xts <- matrix(ncol=length(S), nrow=n)

```

```

for (j in 1:length(S)) {
  xt[,j] <- S.pow(d$m2, S[j])
}
sdtransforms <- apply(xt, 2, sd, na.rm=T)
for (j in 1:(2*length(S))) {
  xts[,j] <- xt[,j]/sdtransforms[j]
}
mutransforms <- apply(xts, 2, mean, na.rm=T)

var.list <- list("w1"=(d$m1/sdofx)-mux, "Nobservations"=n,
               "S"=S, 'Ntransform'=length(S), 'pi'=pi,
               "sdtransforms"=sdtransforms, "mutransforms"=mutransforms,
               "betamean"=0, "betaprec"=0.33, "beta0mean"=0,
               "beta0prec"=0.01, "alphamean"=c(0), "alphaprec"=c(0.01),
               "tauu_a"=1, "tauu_b"=1, "tauu_a"=1, "tauu_b"=1,
               'myx'=mux, 'sdofx'=sdofx)

if (mod.family=='gaussian') {
  var.list <- c(var.list, list("Y"=Y/sdofY, "tauu_a"=1, "tauu_b"=1, "sdofy"=sdofY)) #,
'muy'=muy))
} else {var.list <- c(var.list, list("Y"=Y))}

if (m2=='Xobs') {
  var.list <- c(var.list, list("Xobs"=(d$m2/sdofx)-mux, "nvalidated"=n1))
} else if (m2=='w2') {
  var.list <- c(var.list, list("w2"=(d$m2/sdofx)-mux, "Nreplications"=n1))
}

param.list <- c("beta.0", "betaT", "beta0", "deviance", 'ind', 'tauu', 'taux', 'alpha')
load.module("dic", quiet = TRUE)
inits.list <- vector("list", n.chains)
for (i in chains) {
  inits.list[[i]] <- list(.RNG.name="base::wichmann-Hill",
                        .RNG.seed=i)
}

modfile <- ifelse(mod.family=='gaussian',
                 ifelse(m2=='Xobs', "bugmods/linvselect_fp1_val_ssvs.bug",
                        "bugmods/linvselect_fp1_rep_ssvs.bug"),
                 ifelse(m2=='Xobs', "bugmods/logvselect_fp1_val_ssvs.bug",
                        "bugmods/logvselect_fp1_rep_ssvs.bug"))

sim.fit <- jags.model(modfile, data = var.list, inits=inits.list,
                    n.chains = n.chains, n.adapt=n.adapt)
samples <- coda.samples(sim.fit, param.list, n.burn)

g <- sapply(geweke.diag(samples, 0.1, 0.5), function(x) 2*pnorm(-abs(x[[1]])))
subsamples <- list()
for (i in 1:n.chains) {
  subsamples[[i]] <- samples[[i]][,c(grep('beta.0', colnames(samples[[1]])),
                                     grep('betaT', colnames(samples[[1]])))]
}
subsamples <- mcmc.list(subsamples)
rhat <- gelman.diag(subsamples)[[2]]
merror_v_sd <- summary(samples)[[1]][c('alpha'),2]/summary(samples)[[1]][,4]
eff_sample_size <- effectiveSize(samples)
table <- cbind(mean=summary(samples)[[1]][,1], lci=summary(samples)[[2]][, '2.5%'],
uci=summary(samples)[[2]][, '97.5%'],
               geweke=g, merror_v_sd, eff_sample_size, rhat)

# Returns the burn samples collected, the jags fit object and information
# about the burn samples
return(list(samples, sim.fit=sim.fit,
           table=table, g=g, rhat=rhat))
}

# BTS "FP1" method
mcmc.fp2 <- function(d, Y, mod.family='gaussian', m1='w1', m2='Xobs', pi=1/32,
                    n.adapt, n.burn, chains=c(1:5)) {
  n.chains=length(chains)
  require(rjags)

  n <- dim(d)[1]
  n1 <- sum(!is.na(d$m2)) #number with validation data
  d$m1 <- d[,m1]
  d$m2 <- d[,m2]

  if (mod.family=='gaussian') {sdofY = sd(Y)} else {sdofY=1}
  if (m2=='Xobs') {
    d$Y <- Y
    d1 <- d[!is.na(d$m2),]
    d2 <- d[is.na(d$m2),]
    d <- data.frame(rbind(d1,d2))
    Y <- d$Y
    d <- d[-length(d)]
    rm(d1,d2)
  }
}

```

```

    sdofx <- sd(d$m2, na.rm=T)
    mux <- mean(d$m2/sdofx, na.rm=T)
  } else {
    sdofx <- sqrt(cov(d$m1, d$m2))
    mux <- mean(c(d$m1, d$m2)/sdofx, na.rm=T)
  }

xt <- xts <- matrix(ncol=2*length(S), nrow=n)
for (j in 1:length(S)) {
  xt[,j] <- S.pow(d$m2, S[j])
}
for (j in (length(S)+1):(2*length(S))) {
  xt[,j] <- S.pow(d$m2, S[j-length(S)])*log(d$m2)
}
sdtransforms <- apply(xt, 2, sd, na.rm=T)
for (j in 1:(2*length(S))) {
  xts[,j] <- xt[,j]/sdtransforms[j]
}
mutransforms <- apply(xts, 2, mean, na.rm=T)

var.list <- list("w1"=(d$m1/sdofx)-mux, "Nobservations"=n,
               "S"=S, 'Ntransform'=length(S), 'pi'=pi,
               "sdtransforms"=sdtransforms, "mutransforms"=mutransforms,
               "betamean"=0, "betaprec"=0.33, "beta0mean"=0,
               "beta0prec"=0.01, "alphamean"=c(0), "alphaprec"=c(0.01),
               "taux_a"=1, "taux_b"=1, "tauu_a"=1, "tauu_b"=1,
               'myx'=mux, 'sdofx'=sdofx)

if (mod.family=='gaussian') {
  var.list <- c(var.list, list("Y"=Y/sdofY, "tauy_a"=1, "tauy_b"=1, "sdofy"=sdofY)) #,
'muy'=muy))
} else {
  var.list <- c(var.list, list("Y"=Y))
}

if (m2=='xobs') {
  var.list <- c(var.list, list("Xobs"=(d$m2/sdofx)-mux, "Nvalidated"=n1))
} else if (m2=='w2') {
  var.list <- c(var.list, list("w2"=(d$m2/sdofx)-mux, "Nreplications"=n1))
}

param.list <- c("beta.0", "betax", "beta0", "betaT", "deviance", 'ind', 'alpha')
load.module("dic", quiet = TRUE)
inits.list <- vector("list", n.chains)
for (i in chains) {
  inits.list[[i]] <- list(.RNG.name="base:Wichmann-Hill",
                        .RNG.seed=i)
}

modfile <- ifelse(mod.family=='gaussian',
                 ifelse(m2=='xobs', "bugmods/linvareselect_fp2_val_ssvs.bug",
                        "bugmods/linvareselect_fp1_rep_ssvs.bug"),
                 ifelse(m2=='xobs', "bugmods/logvareselect_fp1_val_ssvs.bug",
                        "bugmods/logvareselect_fp1_rep_ssvs.bug"))

sim.fit <- jags.model(modfile, data = var.list, inits=inits.list,
                    n.chains = n.chains, n.adapt=n.adapt)
samples <- coda.samples(sim.fit, param.list, n.burn)

g <- sapply(geweke.diag(samples, 0.1, 0.5), function(x) 2*pnorm(-abs(x[[1]])))
subsamples <- list()
for (i in 1:n.chains) {
  subsamples[[i]] <- samples[[i]][,c(grep('beta.0', colnames(samples[[1]])),
                                   grep('betaT', colnames(samples[[1]])))]
}
subsamples <- mcmc.list(subsamples)
rhat <- gelman.diag(subsamples)[[2]]
merror_v_sd <- summary(samples)[[1]][c('alpha'),2]/summary(samples)[[1]][,4]
eff_sample_size <- effectiveSize(samples)
table <- cbind(mean=summary(samples)[[1]][,1], lci=summary(samples)[[2]][, '2.5%'],
uci=summary(samples)[[2]][, '97.5%'],
               geweke=g, merror_v_sd, eff_sample_size, rhat)

# Returns the burn samples collected, the jags fit object and information
# about the burn samples
return(list(samples, sim.fit=sim.fit,
           table=table, g=g, rhat=rhat))
}

# Draws final samples for inference using the saved states of the jags objects
more.mcmc <- function(bts.mod, n.samples, n.thin=1, more.samples=F, recompile=F) {
  n.chains <- length(bts.mod[[1]])

  if (recompile==T) {
    endstate <- bts.mod$sim.fit$state(internal = T)

```

```

data <- bts.mod$sim.fit$data()
themodel <- bts.mod$sim.fit$model()

modelname <- paste0("model", round(runif(1,1,99),0), ".bug")
fileConn<-file(modelname)
writeLines(themodel, fileConn)
close(fileConn)

sim.fit <- jags.model(modelname, data = data[-grep("x$", names(data))],
                     inits = endstate, n.chains = n.chains, n.adapt=0)
} else sim.fit <- bts.mod$sim.fit

param.list <- c("beta0", "betaX", 'ind', "deviance", "betaT", "tauy")
samples <- coda.samples(sim.fit, param.list, n.samples, n.thin=n.thin)

merror_v_sd <- summary(samples)[[1]][,2]/summary(samples)[[1]][,4]
eff_sample_size <- effectiveSize(samples)
table <- cbind(mean=summary(samples)[[1]][,1], lci=summary(samples)[[2]][, '2.5%'],
uci=summary(samples)[[2]][, '97.5%'],
               geweke=g, merror_v_sd, eff_sample_size)

if (more.samples==T) {
  samples1 <- bts.mod[[1]]
  for (cc in 1:sim.fit$nchain()) {
    samples[[cc]] <- rbind(samples1[[cc]], samples[[cc]])
  }
}

# Returns the samples for inferenct, the jags fit object and information
# about the samples. Function can be run repeatedly if determining
# whether the stationary distribution has been reached
return(list(samples, sim.fit=sim.fit,
           table=table, g=bts.mod$g, rhat=bts.mod$rhat))
}

mcmc.fp1.results <- function(bts.mod, Xvalues=Xvalues, Yvalues=Yvalues) {
  samples <- bvs.mod[[1]]
  n.chains <- length(samples)
  sim.fit <- bvs.mod$sim.fit

  if (n.chains > 1) {
    for (cc in 2:n.chains) {
      samples[[1]] <- rbind(samples[[1]], samples[[cc]])
    }
    samples <- mcmc.list(mcmc(samples[[1]]))
  }
  n.samples <- dim(samples[[1]])[1]

  # Create a matrix of frequencies of indicator values eq 1
  ind.mat <- samples[[1]][,grep('ind', colnames(samples[[1]]))]
  ind.out <- summary(ind.mat)[[1]][,1]
  dt <- data.table::as.data.table(ind.mat)
  ind.freq <- dt[, .N/n.samples, by = names(dt)]
  ind.freq <- ind.freq[order(ind.freq$V1, decreasing = T),]
  ind.freq[1:5,]

  num.models <- dim(ind.freq)[1]
  linear.keepers=NA;linear.samples=NA; num.lin=0
  for (mm in 1:num.models) {
    model.keepers <- c(1:8)[as.logical(unlist(ind.freq[mm,1:8]))]
    model.keep.samples <- apply(ind.mat, 1, function(x)
      all(as.logical(x)==as.logical(unlist(ind.freq[mm,1:8]))))
    if (mm==1) {
      keepers=model.keepers
      keep.samples=model.keep.samples
    }
    if (all(model.keepers == c(6))) {
      num.lin = unlist(ind.freq[mm,9])*n.samples
      linear.keepers=model.keepers
      linear.samples=model.keep.samples
    }
  }

  beta.mat.all <- samples[[1]][,grep('betaX', colnames(samples[[1]])]
  beta0 <- samples[[1]][keep.samples,'beta0']
  beta.mat <- mcmc(beta.mat.all[keep.samples,])
  fp1.p <- S[keepers]
  fp1.coef <- c(mean(beta0), summary(beta.mat)[[1]][keepers,1])
  fp1.cis <- HPDinterval(beta.mat)[keepers,]
  fp1.vcov <- cov(cbind(beta0, beta.mat))[c(1,keepers+1),c(1,keepers+1)]

  Xval.mat <- matrix(ncol=length(Xvalues), nrow=length(fp1.p))
  for (pp in 1:length(fp1.p)) Xval.mat[pp,] <- S.pow(Xvalues, fp1.p[pp])

  fp1.Yval <- fp1.coef %%% rbind(rep(1, length(Xvalues)), Xval.mat)
  fp1.mise <- sum((Yvalues - fp1.Yval)^2)/length(Yvalues)

```

```

# Yvalue cis for just the model selected
pred.mean.dist <- as.mcmc(beta0 + beta.mat[,keepers] %%% Xval.mat)
fp1.Yval.cis <- matrix(ncol=length(Xvalues), nrow=2)
for (xx in 1:length(Xvalues)) {
  fp1.Yval.cis[,xx] <- HPDinterval(pred.mean.dist[,xx])
}

# Linear model samples
lin.Yval <- lin.mise <- lin.Yval.cis <- NA
if (!is.na(linear.keepers)) {
  lin.beta.mat <- mcmc(beta.mat.all[linear.samples,])
  lin.coef <- c(mean(samples[[1]][linear.samples, 'beta0']),
               summary(lin.beta.mat)[[1]][linear.keepers,1])
  lin.Yval <- lin.coef %%% rbind(rep(1, length(Xvalues)), Xvalues)
  lin.mise <- sum((Yvalues - lin.Yval)^2)/length(Yvalues)
  lin.pred.mean.dist <- as.mcmc(samples[[1]][linear.samples, 'beta0'] +
                               lin.beta.mat[,linear.keepers] %%%
                               matrix(Xvalues, ncol=length(Xvalues), nrow=1))
  lin.Yval.cis <- matrix(ncol=length(Xvalues), nrow=2)
  for (xx in 1:length(Xvalues)) {
    lin.Yval.cis[,xx] <- HPDinterval(lin.pred.mean.dist[,xx])
  }
}

out <- list(ind=ind.out, ind.freq=ind.freq, fp1.p=fp1.p, fp1.coef=fp1.coef,
           fp1.cis=fp1.cis, fp1.mise=fp1.mise, fp1.vcov=fp1.vcov,
           fp1.Yval=fp1.Yval, fp1.Yval.cis=fp1.Yval.cis,
           num.lin=num.lin, lin.Yval=lin.Yval, lin.pred.err=lin.pred.err,
           lin.Yval.cis=lin.Yval.cis)
return(out)
}

mcmc.fp2.results <- function(bts.mod, Xvalues=Xvalues, Yvalues=Yvalues) {
  samples <- bts.mod[[1]]
  n.chains <- length(samples)
  sim.fit <- bts.mod$sim.fit

  if (n.chains > 1) {
    for (cc in 2:n.chains) {
      samples[[1]] <- rbind(samples[[1]], samples[[cc]])
    }
    samples <- mcmc.list(mcmc(samples[[1]]))
  }
  n.samples <- dim(samples[[1]])[1]

  # Create a matrix of frequencies of indicator values eq 1
  ind.mat <- samples[[1]][,grep('ind', colnames(samples[[1]])]]
  ind.out <- summary(ind.mat)[[1]][,1]
  dt <- data.table::as.data.table(ind.mat)
  ind.freq <- dt[, .N/n.samples, by = names(dt)]
  ind.freq <- ind.freq[order(ind.freq$V1, decreasing = T),]
  ind.freq[1:5,]

  num.models <- dim(ind.freq)[1]
  linear.keepers=NA; linear.samples=NA; num.lin=0

  for (mm in 1:num.models) {
    model.keepers <- c(1:16)[as.logical(unlist(ind.freq[mm,1:16]))]
    model.keep.samples <- apply(ind.mat, 1, function(x)
                              all(as.logical(x)==as.logical(unlist(ind.freq[mm,1:16]))))
    if (mm==1) {
      keepers=model.keepers
      keep.samples=model.keep.samples
    }
    if (all(model.keepers == c(6))) {
      num.lin = unlist(ind.freq[mm,17])*n.samples
      linear.keepers=model.keepers
      linear.samples=model.keep.samples
    }
  }

  beta.mat.all <- samples[[1]][,grep('betax', colnames(samples[[1]]))]
  beta0 <- samples[[1]][keep.samples,'beta0']
  beta.mat <- mcmc(beta.mat.all[keep.samples,])
  keepers1 <- keepers[keepers<=8]
  keepers2 <- keepers[keepers>8]
  fp2.p1 <- S[keepers1]
  fp2.p2 <- S[keepers2-8]
  fp2.coef <- c(mean(beta0), summary(beta.mat)[[1]][keepers,1])
  fp2.cis <- HPDinterval(beta.mat)[keepers,]
  for (kk in keepers) {
    if (kk==keepers[1]) {
      fp2.cisET <- quantile(beta.mat[,kk], c(0.0275, 0.975)) } else {
      fp2.cisET <- rbind(fp2.cisET, quantile(beta.mat[,kk], c(0.0275, 0.975)))
    }
  }
}

```

```

}
fp2.vcov <- cov(cbind(beta0, beta.mat))[c(1,keepers+1),c(1,keepers+1)]

xval.mat <- matrix(ncol=length(xvalues), nrow=length(keepers))
place = 0
if (length(fp2.p1) > 0) for (pp in 1:length(fp2.p1)) {
  xval.mat[pp,] <- S.pow(xvalues, fp2.p1[pp])
  place=place+1
}
if (length(fp2.p2) > 0) for (pp in 1:length(fp2.p2)) {
  xval.mat[pp+place,] <- S.pow(xvalues, fp2.p2[pp])*log(xvalues)
}

fp2.Yval <- fp2.coef %%% rbind(rep(1, length(xvalues)), xval.mat)
fp2.mise <- sum((Yvalues - fp2.Yval)^2)/length(Yvalues)
pred.mean.dist <- as.mcmc(beta0 + beta.mat[,keepers] %%% xval.mat)
fp2.Yval.cis <- matrix(ncol=length(xvalues), nrow=2)
for (xx in 1:length(xvalues)) {
  fp2.Yval.cis[,xx] <- HPDinterval(pred.mean.dist[,xx])
}

#linear output
lin.Yval <- lin.pred.err <- lin.Yval.cis <- NA
if (!is.na(linear.keepers)) {
  lin.beta.mat <- mcmc(beta.mat.all[linear.samples,])
  lin.coef <- c(mean(samples[[1]][linear.samples, 'beta0']),
               summary(lin.beta.mat)[[1]][linear.keepers,1])
  lin.Yval <- lin.coef %%% rbind(rep(1, length(xvalues)), xvalues)
  lin.mise <- sum((Yvalues - lin.Yval)^2)/length(Yvalues)
  lin.pred.mean.dist <- as.mcmc(samples[[1]][linear.samples, 'beta0'] +
                                lin.beta.mat[,linear.keepers] %%%
                                matrix(xvalues, ncol=length(xvalues), nrow=1))

  lin.Yval.cis <- matrix(ncol=length(xvalues), nrow=2)
  for (xx in 1:length(xvalues)) {
    lin.Yval.cis[,xx] <- HPDinterval(lin.pred.mean.dist[,xx])
  }
}
out <- list(ind=ind.out,
            ind.freq=ind.freq,
            fp2.p1=fp2.p1,
            fp2.p2=fp2.p2,
            fp2.coef=fp2.coef,
            fp2.cis=fp2.cis,
            fp2.mise=fp2.mise,
            fp2.vcov=fp2.vcov,
            fp2.Yval=fp2.Yval,
            fp2.Yval.cis=fp2.Yval.cis,
            num.lin=num.lin,
            lin.Yval=lin.Yval,
            lin.pred.err=lin.pred.err,
            lin.Yval.cis=lin.Yval.cis)

return(out)
}

# bts.mod <- mcmc.fp1(d, Y, mod.family=mod.family, m1='w1', m2='xobs', pi=bts.pi,
#                    n.adapt=30000, n.burn=10000, chains=c(1:5))
# bts.mod2 <- more.mcmc.fp1(bts.mod, n.samples=10000, n.thin=1)
# mcmc.fp1.results(bts.mod2, xvalues, Yvalues)

#file 'linvselect_fp1_rep_ssvs.bug'
model {
  for (i in 1:Nobservations) {
    X[i] ~ dnorm(mux[i], taux)
    mux[i] <- alpha[1]
    w1[i] ~ dnorm(X[i], tauu)

    #All potential transformations of X[i]
    for (k in 1:Ntransform) {
      xt1[i,k] <- (ifelse(S[k]==0, log((X[i]+myx)*sdofx),
                          pow((X[i]+myx)*sdofx + myx, S[k]))/
                  sdtransforms[k]-mutransforms[k])
    }

    #Linear outcome model
    Y[i] ~ dnorm(mu[i], tauy)
    mu[i] <- beta.0 + inprod(betaT, xt1[i,])
  }
  #Replicate measures
  for (i in 1:Nreplications) w2[i] ~ dnorm(X[i], tauu)

  ### Priors ###
  for (k in 1:Ntransform) {
    indA[k] ~ dcat(PInd[])
    ind[k] <- indA[k]-1
    betaT[k] ~ dnorm(betamean, TauM[indA[k]])
  }
}

```



```

TauM[1] <- 3600 #precision of spike
TauM[2] <- betaprec #precision of slab
PInd[2] <- pi
PInd[1] <- 1 - PInd[2]

beta.0 ~ dnorm(beta0mean, beta0prec)
tauu ~ dgamma(tauu_a, tauu_b)
taux ~ dgamma(taux_a, taux_b)
tauy ~ dgamma(tauy_a, tauy_b)
alpha ~ dnorm(alphamean, alphaprec)

# Re-scaling of beta coefficients back to original x scale
for (k in 1:Ntransform) {
  betax[k] <- ind[k]*betaT[k]*sdofy/sdtransforms[k]
}
beta0 <- (beta.0 - inprod(betaT, mutransforms))*sdofy
}

#file 'logvselect_fp1_val_ssvs.bug'
model {
  for (i in 1:Nobservations) {
    #All potential transformations of X[i]
    for (k in 1:Ntransform) {
      xt1[i,k] <- (ifelse(S[k]==0, log(X[i]*sdofx + myx),
                        pow(X[i]*sdofx + myx, S[k])/
                        sdtransforms[k])-mutransforms[k]
    )
  }

  w1[i] ~ dnorm(X[i], tauu)

  #Logistic outcome model
  Y[i] ~ dbern(mu[i])
  logit(mu[i]) <- beta.0 + inprod(betaT, xt1[i,])
}
#Exposure model
for (i in 1:Nvalidated) X[i] <- Xobs[i]
for (i in (Nvalidated+1):Nobservations) {
  X[i] ~ dnorm(mux[i], taux)
  mux[i] <- alpha[1]
}
### Priors ###
for (k in 1:Ntransform) {
  indA[k] ~ dcat(PInd[])
  ind[k] <- indA[k]-1
  betaT[k] ~ dnorm(betamean, TauM[indA[k]])
}
TauM[1] <- 3600 #precision of spike
TauM[2] <- betaprec #precision of slab
PInd[2] <- pi
PInd[1] <- 1 - PInd[2]

beta.0 ~ dnorm(beta0mean, beta0prec)
tauu ~ dgamma(tauu_a, tauu_b)
taux ~ dgamma(taux_a, taux_b)
alpha ~ dnorm(alphamean, alphaprec)

#re-scaling of beta coefficients back to original x scale
for (k in 1:Ntransform) {
  betax[k] <- ind[k]*betaT[k]/sdtransforms[k]
}
beta0 <- (beta.0 - inprod(betaT, mutransforms))
}

#file 'linvselect_fp2_val_ssvs.bug'
model {
  for (i in 1:Nobservations) {
    #All potential transformations of X[i]
    for (k in 1:Ntransform) {
      xt1[i,k] <- (ifelse(S[k]==0, log((X[i]+myx)*sdofx),
                        pow((X[i]+myx)*sdofx,
                        S[k])/sdtransforms[k])-mutransforms[k]
    )
  }
  for (k in (Ntransform+1):(2*Ntransform)) {
    xt1[i,k] <- (ifelse(S[k-Ntransform]==0,
                      log((X[i]+myx)*sdofx)*log((X[i]+myx)*sdofx),
                      pow((X[i]+myx)*sdofx,
                      S[k-Ntransform])*log(X[i]*sdofx + myx))/
                      sdtransforms[k])-mutransforms[k]
  }
}

  w1[i] ~ dnorm(X[i], tauu)

  #Linear outcome model
  Y[i] ~ dnorm(mu[i], tauy)
  mu[i] <- beta.0 + inprod(betaT, xt1[i,])
}

```

```

#Exposure model
for (i in 1:Nvalidated) X[i] <- Xobs[i]
for (i in (Nvalidated+1):Nobservations) {
  X[i] ~ dnorm(mux[i], tau)
  mux[i] <- alpha[1]
}
### Priors ###
for (k in 1:(2*Ntransform)) {
  indI[k] ~ dbern(pi)
}
for (k in 1:Ntransform) {
  ind[k] <- indI[k]
  ind[k+Ntransform] <- indI[k]*indI[k+Ntransform]
}
for (k in 1:(2*Ntransform)) {
  indA[k] <- ind[k] + 1
  betaT[k] ~ dnorm(betamean, TauM[indA[k]])
}

TauM[1] <- 3600 #precision of spike
TauM[2] <- betaprec #precision of slab

beta.0 ~ dnorm(beta0mean, beta0prec)
tauu ~ dgamma(tauu_a, tauu_b)
tau ~ dgamma(tau_a, tau_b)
tauy ~ dgamma(tauy_a, tauy_b)
alpha ~ dnorm(alphamean, alphaprec)

#re-scaling of beta coefficients back to original x scale
for (k in 1:(2*Ntransform)) {
  betaX[k] <- ind[k]*betaT[k]*sdofy/sdtransforms[k]
}
beta0 <- (beta.0 - inprod(betaT, mutransforms))*sdofy
}

#####
#Bayesian RC #
#####
# MCMC implementation of the classical error model (no substantive model)
# The mcmc.rc.function and bugs models are unchanged from Chapter 4/Appendix C
# The inla.rc.function is unchanged from Chapter 4/Appendix C

# Generate samples via MCMC
X.samples <- mcmc.rc.function(d, validation = T,
                             n.adapt = 30000, n.burn = 10000, n.samples=6000,
                             n.thin=1, n.chains=3)
X.samples <- mcmc.rc[[1]]

# Generate samples via INLA
inla.rc <- inla.rc.function(d, rep.vars=c('w1','w2'), validation=F,
                            xvalues, yvalues, n.samples=3000)
X.samples <- inla.rc

# Transform samples, find expecations, apply the fractional polynomial method
fp1.devs <- c()
for (pp in 1:length(S)) {
  ExpXt <- apply(S.pow(X.samples, S[pp]), 2, mean)
  mod <- glm(Y ~ ExpXt, family=mod.family)
  fp1.devs[pp] <- mod$deviance
  if (S[pp]==1) {
    lin.dev <- mod$dev
    lin.mod <- mod
  }
}
fp1.p <- S[fp1.devs==min(fp1.devs)]
fp1.Xt = apply(S.pow(X.samples, fp1.p), 2, mean)
fp1.mod <- glm(Y~fp1.Xt, family=mod.family, data=d)
fp1.dev <- fp1.mod$deviance

fp2.devs <- c()
for (jj in 1:dim(p.mat)[1]) {
  ExpXt <- lapply(S.pow(X.samples, p.mat[jj,]), FUN=function(x) apply(x, 2, mean))
  mod <- glm(Y ~ ExpXt[[1]] + ExpXt[[2]], family=mod.family)
  fp2.devs[jj] <- mod$deviance
}

fp2.p <- p.mat[fp2.devs==min(fp2.devs),]
fp2.Xt = lapply(S.pow(X.samples, fp2.p), FUN=function(x) apply(x, 2, mean))
Xt1 <- fp2.Xt[[1]]; Xt2 <- fp2.Xt[[2]]
fp2.mod <- glm(Y~Xt1+Xt2, family=mod.family, data=d)
fp2.dev <- fp2.mod$dev

nullmod <- glm(Y ~ 1, family=mod.family)
null.dev <- nullmod$dev

```

```

dispersion <- if (mod.family=='gaussian') {
  if (lin.mod$df.residual > 0) sum(lin.mod$residuals^2)/lin.mod$df.residual
} else 1

p.null1 <- pchisq((null.dev - fp1.dev)/dispersion, df=2, lower.tail=F)
p.null2 <- pchisq((null.dev - fp2.dev)/dispersion, df=4, lower.tail=F)
p.lin1 <- pchisq((lin.dev - fp1.dev)/dispersion, df=1, lower.tail=F)
p.lin2 <- pchisq((lin.dev - fp2.dev)/dispersion, df=3, lower.tail=F)
p.fp <- pchisq((fp1.dev - fp2.dev)/dispersion, df=2, lower.tail=F)

lin.cis <- confint.default(lin.mod)
fp1.cis <- confint.default(fp1.mod)
fp2.cis <- confint.default(fp2.mod)

lin.vcov <- vcov(lin.mod)
fp1.vcov <- vcov(fp1.mod)
fp2.vcov <- vcov(fp2.mod)

lin.AIC <- AIC(lin.mod); fp1.AIC <- AIC(fp1.mod); fp2.AIC <- AIC(fp2.mod)

lin.Yval <- predict(lin.mod, newdata=data.frame(ExpXt=xvalues), se.fit=T,
  response='link')
lin.mise <- sum((Yvalues - lin.Yval[[1]])^2)/length(Yvalues)

fp1.Yval <- predict(fp1.mod, newdata=data.frame(fp1.Xt=S.pow(Xvalues, fp1.p)),
  se.fit=T, response='link')
fp1.mise <- sum((Yvalues - fp1.Yval[[1]])^2)/length(Yvalues)

xtvalues <- S.pow(Xvalues, fp2.p)
fp2.Yval <- predict(fp2.mod, newdata=data.frame(Xt1=xtvalues[[1]],
  Xt2=xtvalues[[2]]),
  se.fit=T, response='link')
fp2.mise <- sum((Yvalues - fp2.Yval[[1]])^2)/length(Yvalues)

lin.Yval.cis <- rbind(lin.Yval[[1]] - 1.96*lin.Yval[[2]],
  lin.Yval[[1]] + 1.96*lin.Yval[[2]])
fp1.Yval.cis <- rbind(fp1.Yval[[1]] - 1.96*fp1.Yval[[2]],
  fp1.Yval[[1]] + 1.96*fp1.Yval[[2]])
fp2.Yval.cis <- rbind(fp2.Yval[[1]] - 1.96*fp2.Yval[[2]],
  fp2.Yval[[1]] + 1.96*fp2.Yval[[2]])

```

## D. CODE OF THE BAYESIAN IMPLEMENTATION OF THE EPISODIC CONSUMERS MODEL IN JAGS AND INLA SOFTWARE

The following code can be used to fit the EC model as presented in Section 7.3 using either MCMC via JAGS or INLA via the INLA software. Both software are accessed through R packages 'rjags' and 'INLA', respectively. The resulting samples can be used within the MCMC-RC and INLA-RC methods using code in Appendix B or Appendix C.

```
#####
# Data preparation #
#####

# Functions for finding the best BoxCox transformation, applying it, and for
# centering and scaling
bestboxcox <- function(a) {
  bc <- boxcox(lm(a[a>0]~1), plotit=F, lambda=seq(0.00, 0.5, length=51))
  bc$x[bc$y==max(bc$y)]
}
BoxCoxAARP <- function(x, lambda){
  y = x; ispositive = 1.0 * (y > 0); a0 <- min(y[y > 0])/2; y[y <= 0] <- a0
  if (lambda == 0) uu = log(y)
  if (abs(lambda) > 0) uu = (y^lambda - 1)/lambda
  uu = uu * ispositive
  list(x.bc = uu)
}
t.cov <- function(c) {
  a <- c
  lambda = bestboxcox(a)
  b <- unlist(BoxCoxAARP(c, lambda))
  print(lambda)
  (b - mean(b, na.rm=T))/sd(b, na.rm=T)
}
# Apply the best Box Cox transformation, centering and scaling to continuous covariates
t.covariates <- data.frame(apply(d[,c("meanFFQ_NUT002", "baselinebmi", "age")], 2,
  t.cov))

# Create a binary value for consumption/not consumption
r1 <- r.alc; r2 <- r.alc
r1 <- ifelse(r.alc>0, 1, 0)

# Apply the best BoxCox transformation to the positive alcohol 7DD measurements
a = c(d$D1_NUT002, d$D2_NUT002, d$DF_NUT002); lambda.r <- bestboxcox(a)
r2 <- BoxCoxAARP(r.alc, lambda.r)$x.bc
r2[r1==0] <- 0

# Scale positive alcohol values
sdoflambda = sd(as.vector(r2)[as.vector(r2)!=0 & !is.na(as.vector(r2))])
s2 = r2/sdoflambda

#####
## Fitting the EC model in INLA ##
#####
n==sum(!is.na(s2[,1]))
n2 <- sum(!is.na(s2[,2]))
n3 <- sum(!is.na(s2[,3]))

rep.vars=c("D1_NUT002", "D2_NUT002", "DF_NUT002")
dcenter <- cbind(s2, ec.covar1, 'orig.no'=1:n, Y15, r1=r1, 'X'=NA)
d1 <- dcenter[order(dcenter$DF_NUT002, dcenter$D2_NUT002, na.last = T),]

tau.e <- 1 #initial value for the precision of the measurement error

# the response matrix
# Two outcome models - R2 ~ gamma.2 + gamma.z.2*Z + u2 + eps (1st component)
# - logit(R1) ~ gamma.1 + gamma.z.1*Z + u1 (2nd component)
num_col <- 2
num_row <- 2*n + 2*n2 + 2*n3
Xmat <- matrix(NA, nrow=num_row, ncol=num_col)
```

```

Xmat[1:n, 1] <- d1[,rep.vars[1]]
Xmat[(n+1):(n+n2), 1] <- d1[,rep.vars[2]][1:n2]
Xmat[(n+n2+1):(n+n2+n3), 1] <- d1[,rep.vars[3]][1:n3]
nn <- n+n2+n3
Xmat[(nn+1):(nn+n), 2] <- d1[, 'r1.D1_NUT002']
Xmat[(nn+n+1):(nn+n+n2), 2] <- d1[, 'r1.D2_NUT002']
Xmat[(nn+n+n2+1):(nn+n+n2+n3), 2] <- d1[, 'r1.DF_NUT002']
Xmat <- Xmat

#data vectors
idx.x <- rep(c(1:n, 1:n2, 1:n3), 2)
us <- c(c(1:n, 1:n2, 1:n3), c((n+1):(n+n), (n+1):(n+n2), (n+1):(n+n3)))
z.mat <- rbind(d1[,4:16], d1[1:n2,4:16], d1[1:n3,4:16])
z.NA <- matrix(rep(NA, 13*dim(z.mat)[1]), ncol=13)
colnames(z.NA) <- colnames(z.mat)
gamma.z.1 <- rbind(z.NA, z.mat)
gamma.z.2 <- rbind(z.mat, z.NA)
sex.1 <- gamma.z.1$sex
sex.2 <- gamma.z.2$sex
age.1 <- gamma.z.1$age
age.2 <- gamma.z.2$age

ffq.1 <- gamma.z.1$meanFFQ_NUT002
bmi.1 <- gamma.z.1$baselinebmi
cig2.1 <- gamma.z.1$cigstat2
cig3.1 <- gamma.z.1$cigstat3
ed1.1 <- gamma.z.1$edlevel1
ed2.1 <- gamma.z.1$edlevel2
ed3.1 <- gamma.z.1$edlevel3
sc2.1 <- gamma.z.1$sc2
sc3.1 <- gamma.z.1$sc3
sc4.1 <- gamma.z.1$sc4
sc5.1 <- gamma.z.1$sc5
ffq.2 <- gamma.z.2$meanFFQ_NUT002
bmi.2 <- gamma.z.2$baselinebmi
cig2.2 <- gamma.z.2$cigstat2
cig3.2 <- gamma.z.2$cigstat3
ed1.2 <- gamma.z.2$edlevel1
ed2.2 <- gamma.z.2$edlevel2
ed3.2 <- gamma.z.2$edlevel3
sc2.2 <- gamma.z.2$sc2
sc3.2 <- gamma.z.2$sc3
sc4.2 <- gamma.z.2$sc4
sc5.2 <- gamma.z.2$sc5

gamma.1 <- c(rep(NA, nn), rep(1, nn))
gamma.2 <- c(rep(1, nn), rep(NA, nn))
Ntrials <- c(rep(NA, nn), rep(1, nn))

data.joint <- data.frame(Xmat=Xmat, idx.x=idx.x, us, sex.1, sex.2, age.1, age.2,
                        ffq.1, bmi.1, cig2.1, cig3.1, ed1.1, ed2.1, ed3.1,
                        sc2.1, sc3.1, sc4.1, sc5.1,
                        ffq.2, bmi.2, cig2.2, cig3.2, ed1.2, ed2.2, ed3.2,
                        sc2.2, sc3.2, sc4.2, sc5.2,
                        gamma.1, gamma.2, Ntrials)

#us is included as a 2-dimensional random effects model
formula <- Xmat ~ f(us, model="iid2d", n=n*2, hyper = list(theta =
  list(prior = "wishart2d", param =c(1, 1, 0.5)))) - 1 +
  sex.1 + sex.2 + age.1 + age.2 + ffq.1 + bmi.1 + cig2.1 + cig3.1 +
  ed1.1 + ed2.1 + ed3.1 + sc2.1 + sc3.1 + sc4.1 + sc5.1 + ffq.2 +
  bmi.2 + cig2.2 + cig3.2 + ed1.2 + ed2.2 + ed3.2 + sc2.2 + sc3.2 +
  sc4.2 + sc5.2 + gamma.1 + gamma.2

# Defines hyperparameters of the two outcome models
# Only 1st component has any hyperparameters - the error term eps~N(0, tau.e^2),
# tau.e~Gamma(1,1)
# Uses Gamma prior instead of uniform b/c INLA does not have a uniform prior command
control.family.list <- list(list(hyper = list(prec = list(initial = tau.e,
  param = c(1,1), fixed = FALSE))), list(hyper = list()))

control.fixed.list <- list(mean = list(sex.1=0, sex.2=0, age.1=0, age.2=0,
  ffq.1=0, bmi.1=0, cig2.1=0, cig3.1=0, ed1.1=0, ed2.1=0, ed3.1=0,
  sc2.1=0, sc3.1=0, sc4.1=0, sc5.1=0, ffq.2=0, bmi.2=0, cig2.2=0,
  cig3.2=0, ed1.2=0, ed2.2=0, ed3.2=0, sc2.2=0, sc3.2=0, sc4.2=0,
  sc5.2=0, gamma.1=0, gamma.2=0),
  prec = list(sex.1=0.01, sex.2=0.01, age.1=0.01, age.2=0.01,
  ffq.1=0.01, bmi.1=0.01, cig2.1=0.01, cig3.1=0.01, ed1.1=0.01,
  ed2.1=0.01, ed3.1=0.01, sc2.1=0.01, sc3.1=0.01, sc4.1=0.01, sc5.1=0.01,
  ffq.2=0.01, bmi.2=0.01, cig2.2=0.01, cig3.2=0.01, ed1.2=0.01,
  ed2.2=0.01, ed3.2=0.01, sc2.2=0.01, sc3.2=0.01, sc4.2=0.01, sc5.2=0.01,
  gamma.1=0.01, gamma.2=0.01))

r <- inla(formula, Ntrials=Ntrials,
  family = c("gaussian", "binomial"),
  data = data.joint,

```

```

        control.family = control.family.list,
        control.fixed = control.fixed.list,
        num.threads = 8,
        control.compute = list(config=TRUE))

# Check that the Hessian value is not negative, if TRUE, then repeat
as.numeric(r$logfile[which(r$logfile=='Eigenvalues of the Hessian')+1])<0
#r <- inla.rerun(r)

means <- r$summary.fixed[, 'mean']
sds <- r$summary.fixed[, 'sd']
names(means) <- rownames(r$summary.fixed)

# Draw samples from estimated posterior distribution
n.samples=10000
i.samples = inla.posterior.sample(n = n.samples, result = r, seed = 1266)

# Translate back into the expected exposure on the boxcox/scaled scale ("that")
# Translate that back to the original alcohol scale
x <- temp <- matrix(nrow=n.samples, ncol=n)
for (ss in 1:n.samples) {
  u <- i.samples[[ss]]$latent[grep('us', rownames(i.samples[[1]]$latent))]
  u1 <- u[(n+1):(2*n)]
  u2 <- u[1:n]
  that <- rje::expit(means['gamma.1'] + means['sex.1']*sex.1[(nn+1):(nn+n)] +
    means['age.1']*age.1[(nn+1):(nn+n)] + means['ffq.1']*ffq.1[(nn+1):(nn+n)] +
    means['bmi.1']*bmi.1[(nn+1):(nn+n)] + means['cig2.1']*cig2.1[(nn+1):(nn+n)] +
    means['cig3.1']*cig3.1[(nn+1):(nn+n)] + means['ed1.1']*ed1.1[(nn+1):(nn+n)] +
    means['ed2.1']*ed2.1[(nn+1):(nn+n)] + means['ed3.1']*ed3.1[(nn+1):(nn+n)] +
    means['sc2.1']*sc2.1[(nn+1):(nn+n)] + means['sc3.1']*sc3.1[(nn+1):(nn+n)] +
    means['sc4.1']*sc4.1[(nn+1):(nn+n)] + means['sc5.1']*sc5.1[(nn+1):(nn+n)] +
    u1)*
    (means['gamma.2'] + means['sex.2']*sex.2[1:n] + means['age.2']*age.2[1:n] +
    means['ffq.2']*ffq.2[1:n] + means['bmi.2']*bmi.2[1:n] +
    means['cig2.2']*cig2.2[1:n] + means['cig3.2']*cig3.2[1:n] +
    means['ed1.2']*ed1.2[1:n] + means['ed2.2']*ed2.2[1:n] + means['ed3.2']*ed3.2[1:n]
  +
    means['sc2.2']*sc2.2[1:n] + means['sc3.2']*sc3.2[1:n] +
    means['sc4.2']*sc4.2[1:n] + means['sc5.2']*sc5.2[1:n] + u2)
  x[ss,] <- (sdoflambd*lambda.r*that + 1)^(1/lambda.r)
}
X.samples <- x[,order(d1$orig.no)] #put back in original order!
exp.that <- apply(x, 2, mean)

#####
## Fitting the EC model using MCMC in JAGS ##
#####
require(rjags)
require(coda)

n <- dim(r1)[1]
num.chains=3

inits.list <- vector("list", num.chains)
for (i in 1:num.chains) {
  inits.list[[i]]$RNG.name <- "base::wichmann-Hill"
  inits.list[[i]]$RNG.seed <- i
}

ec.covar1 <- cbind(1, t.covariates, as.numeric(d$sex==2),
  as.numeric(d$cigstat==2), as.numeric(d$cigstat==3),
  as.numeric(d$edlevel09==1), as.numeric(d$edlevel09==2),
  as.numeric(d$edlevel09==3), as.numeric(d$sc==2),
  as.numeric(d$sc==3), as.numeric(d$sc==4), as.numeric(d$sc==5))
ec.covar2 <- ec.covar1
nGammas=length(ec.covar1)

var.list <- list("r1"=r1, "r2"=s2, "Nobservations"=n,
  "Nrepeats"=rep(J_recalls,n),
  "exp.invsig.u"=solve(matrix(c(1, 0.5, 0.5, 1), byrow=T, ncol=2)),
  "sdoflambd"=sdoflambd, "lambda.r"=lambda.r,
  "zero.u"=c(0,0), "ec.covar1"=ec.covar1, "ec.covar2"=ec.covar2,
  "gammamean"=rep(0,nGammas), "gammaprec"=diag(rep(0.01,nGammas)))

param.list <- c("gamma.1", "gamma.2", "var.u1", "var.u2", "rho.out", "sigma.e", "x")

sim.fit <- jags.model("bugmods/ecmodel-covariates.bug", data = var.list,
  inits=inits.list, n.chains = num.chains, n.adapt=100000)
samples <- coda.samples(sim.fit, variable.names = param.list, n.iter = 100000,
  thin=200)

# Find the posterior expectation for every X[i]
model.sims.array <- function(samples) {
  parameter.names <- colnames(samples[[1]])
  num.chains <- length(samples)
  sims.array <- array(NA, c(dim(samples[[1]])[1], num.chains, length(parameter.names)))
}

```

```

dimnames(sims.array)[[3]] <- parameter.names
for (i in 1:num.chains){
  sims.array[,i,] <- samples[[i]]
}
return(sims.array)
}
sims.array <- model.sims.array(samples)
exp.that <- apply(sims.array[,grep('x', dimnames(sims.array)[[3]])],2,mean)

#file 'ecmodel-covariates.bug'
model {
  for (i in 1:Nobservations) {
    #Person-specific random effects
    u[i, 1:2] ~ dnorm(zero.u, invSigma.u)

    for (j in 1:Nrepeats[i]) {
      #Probability of consumption
      r1[i,j] ~ dbern(p1[i,j])
      logit(p1[i,j]) <- gamma.1 %%% ec.covar1[i,] + u[i,1]

      #Amount consumed
      r2[i,j] ~ dnorm(mur[i,j], tau.e)
      mur[i,j] <- gamma.2 %%% ec.covar2[i,] + u[i,2]
    }

    #Predicted alcohol intake on boxcox/scaled scale
    that[i] <- p1[i,1]*(gamma.2 %%% ec.covar2[i,] + u[i,2])

    #Predicted alcohol intake on original scale
    x[i] <- pow(sdoflambda*lambda.r*that[i] + 1,(1/lambda.r))
  }

  ### Priors ###
  u.df <- 5
  invSigma.u ~ dwish((u.df-2-1)*exp.invSig.u, u.df)
  Sig.out <- inverse(invSigma.u)
  sigma.e ~ dunif(0,3)
  tau.e <- pow(sigma.e,-2)
  gamma.2 ~ dnorm(gammamean, gammaprec)
  gamma.1 ~ dnorm(gammamean, gammaprec)

  ### Estimates ###
  var.u1 <- Sig.out[1,1]
  var.u2 <- Sig.out[2,2]
  rho.out <- Sig.out[1,2]/(sqrt(var.u1)*sqrt(var.u2))
}

```

## **E. MANUSCRIPT PREPARED FOR SUBMISSION TO STATISTICS IN MEDICINE**

---

### **Comparison of methods for measurement error correction: Regression calibration, multiple imputation and Bayesian methods**

Christen M. Gray<sup>1</sup>, Karla DiazOrdaz<sup>1</sup>, Jonathan W. Bartlett<sup>2</sup>, and Ruth H. Keogh<sup>1</sup>

<sup>1</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

<sup>2</sup>Statistical Innovation Group, AstraZeneca, Cambridge, UK

#### **ABSTRACT**

In studies of exposure-outcome relationships, measurement error in continuous exposure variables leads to biased estimates of associations and loss of power.

Attributes of the error distribution can be estimated with a replicate study, i.e. obtaining repeated measures of the error-prone variable, or a validation study wherein the true exposure is determined in a subsample.

The most commonly used method for measurement error correction is regression calibration (RC) which requires an approximation for non-linear models. Bayesian analysis using Markov Chain Monte Carlo (MCMC), multiple imputation (MI), and integrated nested Laplace approximations (INLA), all based on Bayesian principles, have been suggested as potential alternatives as they are more easily adaptable to complex settings. We compare the methods using simulation studies, alongside RC, for both replicate studies and validation studies. Linear and logistic regression models are considered, with varying effects sizes and degrees of measurement error.

MCMC gives unbiased estimates and efficiency gains over RC in all scenarios. Using validation data, MI can be used with standard software by treating the ‘true’ exposure as missing. We extend MI for use with a replicate study. MI gives unbiased estimates when a validation study is available; however, for a replicate study, MI is less efficient and often more computationally intensive than MCMC. INLA gives approximately unbiased estimates for low to moderate measurement error with large efficiency gains over RC.

We recommend increased use of Bayesian methods for measurement error correction and provide the means for implementation via example code using R and JAGS.

*Keywords:* Measurement error, Bayesian analysis, INLA, regression calibration



## F. MANUSCRIPT IN REVIEW FOR THE BIOMETRICAL JOURNAL

---

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

Christen M. Gray<sup>\*1</sup>, Raymond J. Carroll<sup>2</sup>, Marleen A.H. Lentjes<sup>3</sup>, and Ruth H. Keogh<sup>1</sup>

<sup>1</sup> Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT

<sup>2</sup> Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843

<sup>3</sup> Department of Public Health & Primary Care, University of Cambridge, Strangeways Research Laboratories, Worts Causeway, Cambridge CB1 8RN

Received zzz, revised zzz, accepted zzz

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, i.e. 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.

*Key words:* Bayesian modelling; Episodic consumers model; Fractional polynomials; Measurement error; Regression calibration.

---

\*Corresponding author: e-mail: cmgray@alum.mit.edu, Phone: +44-7455-830-503

## G. ADDITIONAL PLOTS OF CHAPTER 3 SIMULATIONS

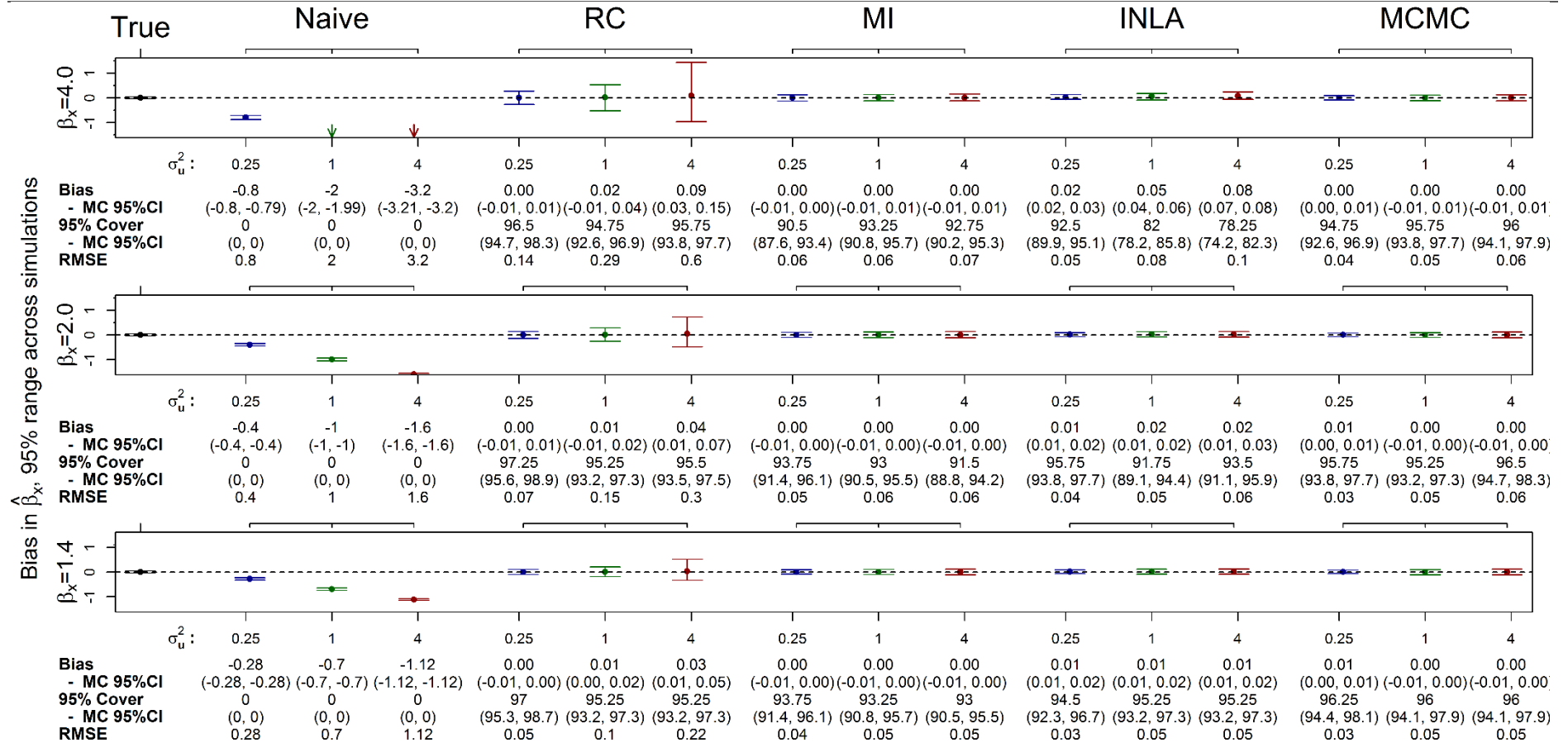


Figure G.1. Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure  $X$  is observed for 10% ( $P = 0.1$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo sampling from a Bayesian model (MCMC). The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

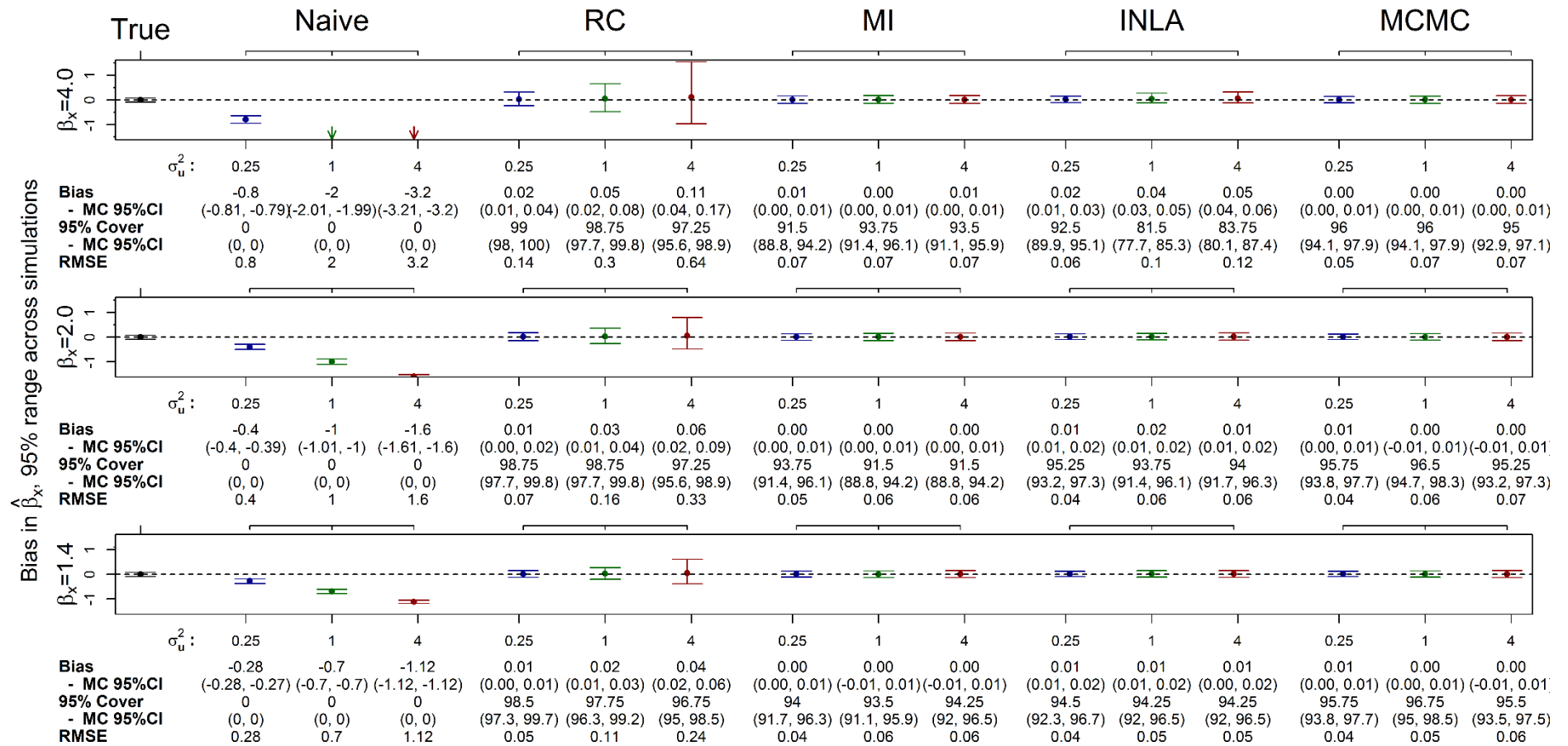


Figure G.2. Simulation study results in the linear regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure  $X$  is observed for 30% ( $P = 0.3$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo sampling from a Bayesian model (MCMC). The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (left, middle, and right) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

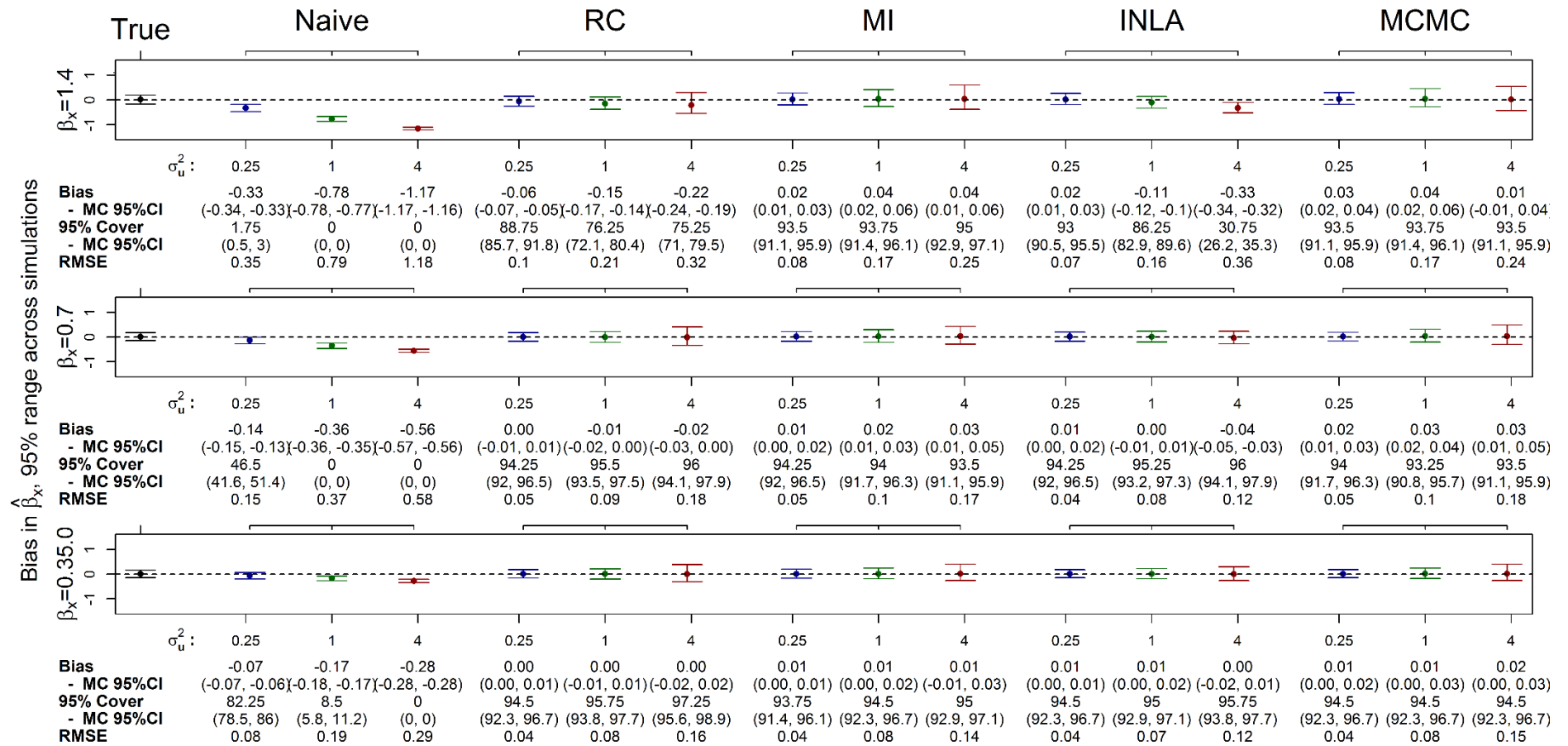


Figure G.3. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a validation study in which the true exposure  $X$  is observed for 10% ( $P = 0.1$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo sampling from a Bayesian model (MCMC). The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

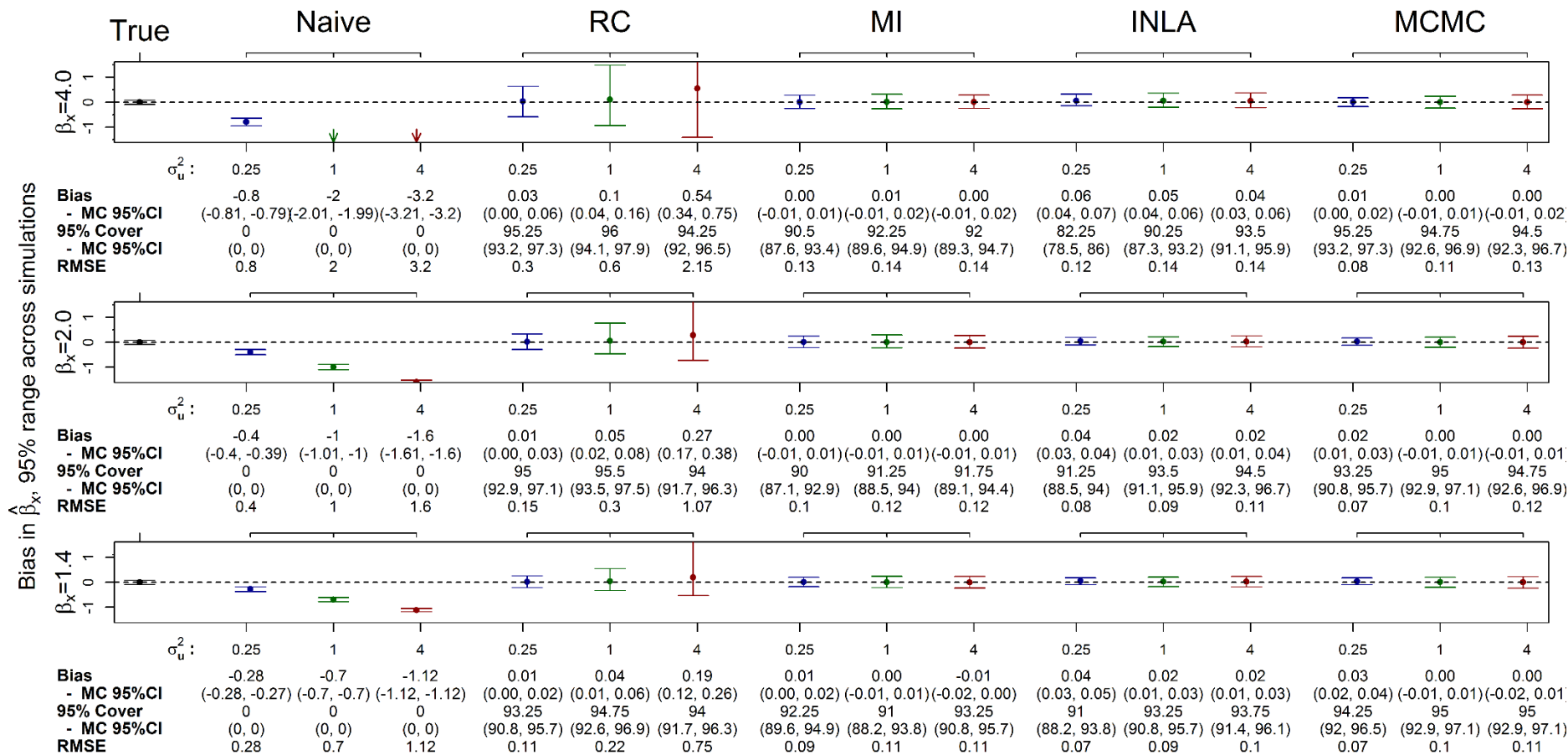


Figure G.4. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 500 and a validation study in which the true exposure  $X$  is observed for 30% ( $P = 0.3$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo sampling from a Bayesian model (MCMC). The association between the true predictor and the outcome,  $\beta_X$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_X$  and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_X$ ,  $\hat{\beta}_X$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

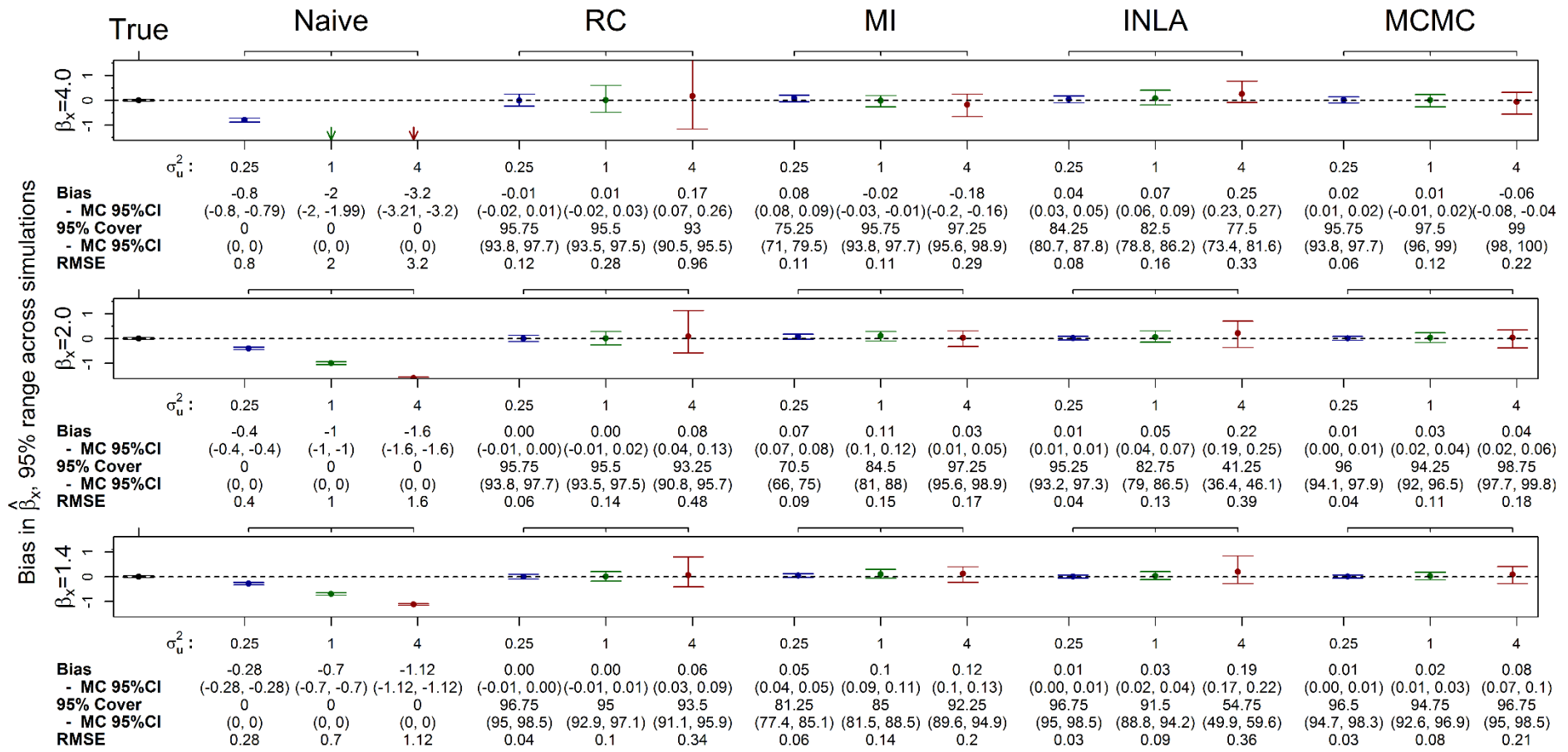


Figure G.5. Simulation study results in the linear regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for 30% ( $P = 0.3$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo sampling from a Bayesian model (MCMC). The association between the true predictor and the outcome,  $\beta_x$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_x$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_x$ ,  $\hat{\beta}_x$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

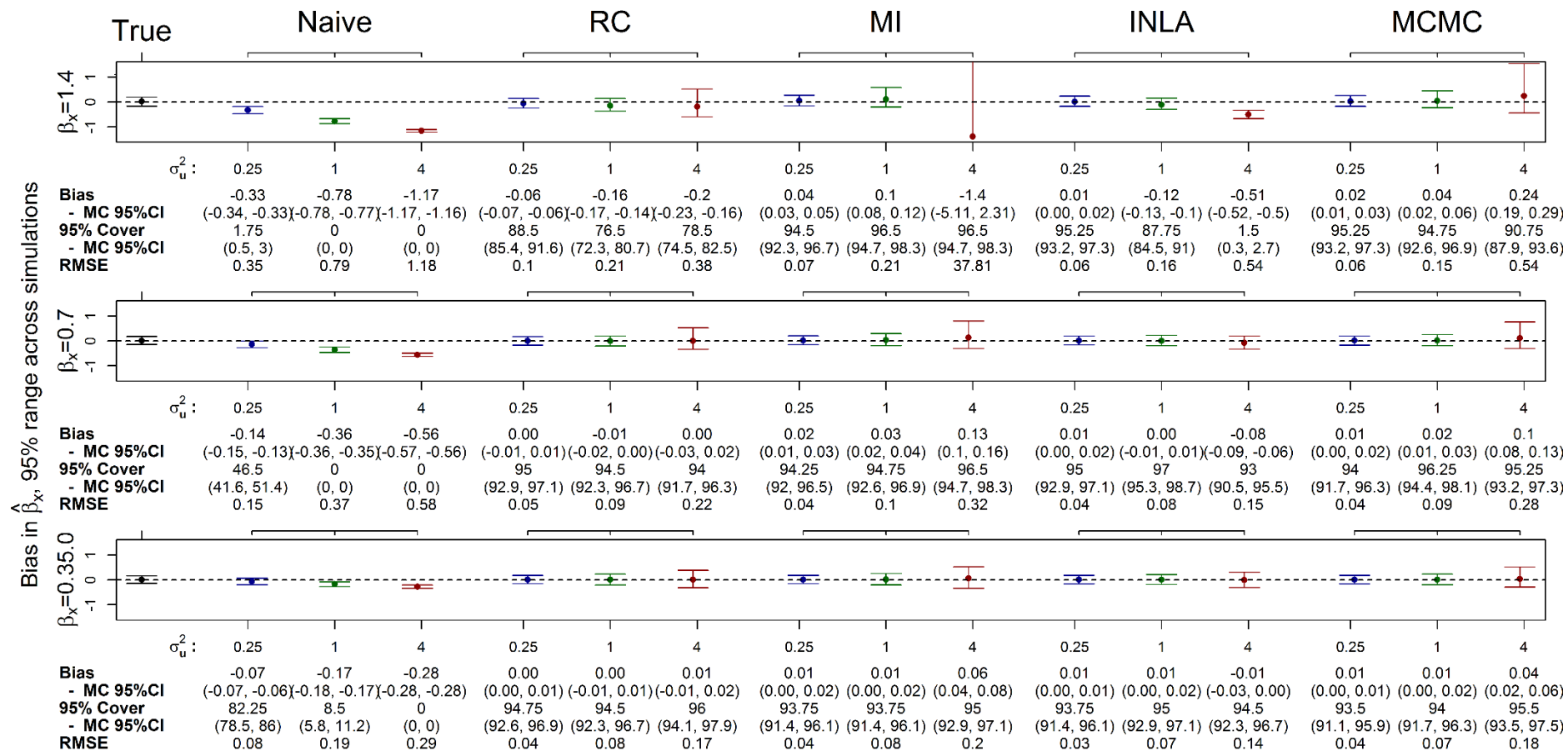


Figure G.6. Simulation study results in the logistic regression setting with sample size ( $N$ ) of 2000 and a replicate study in which a replicate measure is available for 30% ( $P = 0.3$ ) of the study participants. Correction methods applied to 400 simulated data sets including regression calibration (RC); multiple imputation (MI); integrated nested Laplace approximations (INLA); and Markov chain Monte Carlo sampling from a Bayesian model (MCMC). The association between the true predictor and the outcome,  $\beta_x$ , is varied by row (bottom, middle, and top) and the measurement error variance,  $\sigma_u^2$ , is varied by column (blue, green, and red) underneath each method. For each method,  $\beta_x$ , and  $\sigma_u^2$ , the following results are displayed: 1) the bias in the estimate of  $\beta_x$ ,  $\hat{\beta}_x$ , is displayed in the plot with the 95% range of values across all simulations indicated by the error bars; 2) specific estimated bias values and the 95% Monte Carlo error confidence interval (MC 95% CI) for the bias; 3) estimated coverage of the nominal 95% confidence intervals (95% Cover) and MC 95% CI for the coverage; and 4) root mean square error (RMSE).

## H. SENSITIVITY ANALYSIS FOR REGRESSION CALIBRATION WITH MODEL SELECTION INCLUDED IN THE BOOTSTRAPPED 95% CONFIDENCE INTERVALS

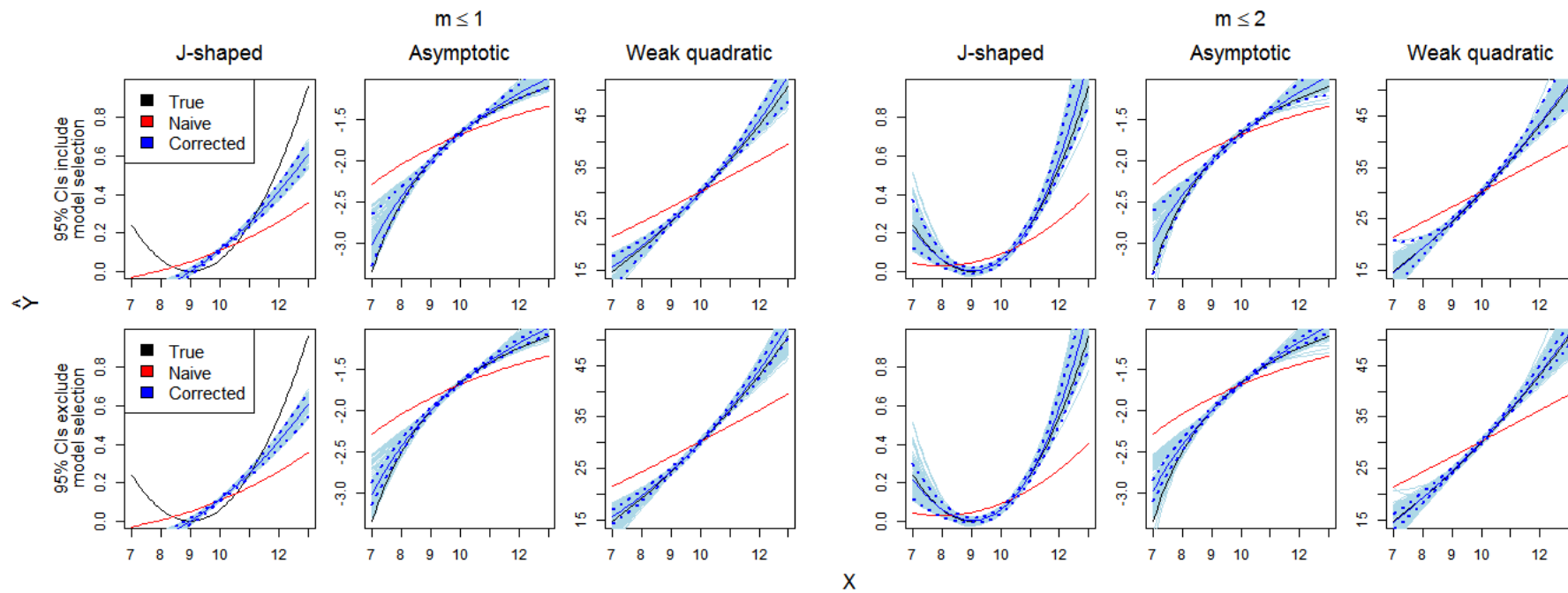


Figure H.1 The chosen curve fits after regression calibration (RC) applied in the context of the fractional polynomial method where the 95% confidence bounds were estimated by bootstrapping either inclusive or non-inclusive of bootstrapping. Data was generated with a validation study and classical measurement error variance equal to the variance of  $X$  for 200 simulations of sample size 2000 (blue lines). A maximum degree of 1 ( $m \leq 1$ ) and 2 ( $m \leq 2$ ) was applied. The black line represents the true curve used to generate the simulated data. The red line is the mean naïve curve fit applying the fractional polynomial method with a single error-prone measure. Darker solid blue line is the mean chosen curve fit (the point-wise mean of all curve fits), and the dashed blue lines are the mean point-wise 95% confidence bounds as determined by use of bootstrapping. For 95% confidence intervals inclusive of model selection, the percentile bootstrap method was used; for 95% confidence bounds exclusive of model selection, the basic bootstrap method was used [77].



