STATISTICS IN MEDICINE, VOL. 12, 1703–1722 (1993)

CONDITIONAL INDEPENDENCE MODELS FOR EPIDEMIOLOGICAL STUDIES WITH COVARIATE MEASUREMENT ERROR

SYLVIA RICHARDSON

Institut National de la Santé et de la Recherche Médicale, U. 170, 16 Avenue Paul Vaillant-Couturier, 94807 Villejuif Cedex, France

AND

WALTER R. GILKS

Medical Research Council Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 2SR, U.K.

SUMMARY

We construct a unifying representation of the structure of measurement error problems with particular reference to situations commonly encountered in epidemiological studies, and outline how estimation of the parameters of interest can be carried out in a Bayesian framework using Gibbs sampling. We show how this approach can be implemented for designs involving continuous measurement errors assessed through a validation substudy, and discuss our results on simulated data.

1. INTRODUCTION

Quantification of the association between covariates and response is a major aim of many statistical applications: the accuracy of the quantification depends critically on the quality of measurement of the covariates. However there are many situations, for instance in dietary or environmental epidemiology, in which covariates are difficult to measure directly and can at best be assessed in only a small fraction of the study population. In these circumstances, investigators must rely on recording imperfect surrogate measures of the true covariates. Here we focus on the epidemiological context where much of the methodology for dealing with measurement error has been developed.

It is well kinown that, in general, analyses which ignore measurement errors or misclassification and treat the surrogates as the true covariates may produce biased results (see the overviews given by Willett[1] or Gail[2]). Most statistical methods for correcting measurement errors have been developed in the context of non-differential measurement error, that is, under the hypothesis that the measurement error is independent of the response. Non-differential measurement error of a covariate usually attenuates relative risk estimates.[3] Moreover, the loss of information created by measurement errors reduces the precision of the estimates as well as the power of the statistical tests. In a multivariate model, measurement error on one covariate can bias the association between the response and the other covariates, even if these latter variables are measured without error. The direction of the bias will depend on the correlation amongst covariates.[4]

0277–6715/93/181703–20$15.00

Received September 1992
Revised December 1992

In view of these difficulties there has been a growing interest in the development of statistical methods for correcting the effects of measurement errors. Existing methods can be characterized by the type of covariate considered (continuous or categorical); by the type of model linking covariate and surrogate; and by the estimation framework employed (maximum likelihood or quasi-likelihood, and whether or not approximations are used). Recent reviews of these methods can be found in particular in Liu and Liang[5] for misclassification of categorical covariates; in Caroll and Wand[6] for continuous covariates; and in the comprehensive overview of Gail.[2]

In this paper we aim at presenting a unifying approach to analyse common situations of measurement errors encountered in epidemiological studies. This approach is based on the formulation of conditional independence relationships between relevant subsets of variables.[7] Once the structure of the model has been defined, estimation is carried out in a Bayesian framework using Gibbs sampling.[8,9] A Gibbs sampling approach to analyse a particular case of measurement error has been proposed by Thomas *et al.*[10] Stephens and Dellaportas[11] discuss how Gibbs sampling can be used in a more general setting of measurement error. Our work develops that of Stephens and Dellaportas[11] in that we structure the information available in general epidemiological settings in terms of conditional independence models.

The structure of the measurement problem in epidemiology can be formulated as follows. Risk factors (covariates) for each individual are to be related to the disease status (response variable) Y of that individual. However, for many or all individuals, while some risk factors C are truly known, other risk factors X are unknown, although one or several surrogate measures Z of X are recorded. To model this situation we shall distinguish three submodels (following the terminology of Clayton[12]):

1. *a disease model*, which expresses the relationship between the risk factors C and X and the disease status Y;
2. *a measurement model*, which expresses the relationship between the surrogate measures Z and the true unknown risk factor X;
3. *an exposure model*, which specifies the distribution of the unknown risk factor X in the general population.

For valid inference in the disease model, traditionally several epidemiological designs have been considered, which may include either or both of:

(a) *a validation group*, that is a group of individuals for whom X can be supposed known (which assumes the existence of a 'gold standard' that is an error-free method of measuring X) and for whom Z has also been recorded;
(b) *repeated determinations* of X by one or more surrogate measurement methods.

The relative merits and the type of errors that can be addressed by these two strategies, validation versus repeated measures, have been discussed by Willett[1] and Marshall.[13] Methods for deattenuation for continuous covariates, using information on the error structure given for example by a validation group, have been proposed by a number of authors;[14-16] some methods make specifically the assumption of small error variance.[17-19] Quasi-likelihood estimation of parameters for a broad class of epidemiological designs including both validation and repeated measures has been considered by Caroll and Stefanski.[20] For discrete data, different sampling schemes and methods for adjusting for misclassification errors can be formulated in the framework of log-linear models.[21,22] In a recent paper, Duffy *et al.*[23] compare the precision of risk estimates for a misclassified binary risk factor derived either from a validation substudy or from repeated determinations.

In Section 2, we present the structure of a broad class of models encompassing all the epidemiological designs commonly used to assess measurement errors. Section 3 is concerned with estimation and the description of the stochastic algorithm (Gibbs sampling) which enables estimation to be carried out. The implementation of our method in particular cases of a logistic regression with measurement error and fairly complex designs is given in Section 4. The estimates derived by Gibbs sampling are also compared, for a relatively simple design, with those obtained using the method proposed by Rosner *et al.*[24] The final section discusses the relative merits of our conditional independence Bayesian approach with those of more standard methods.

2. CONDITIONAL INDEPENDENCE MODELLING OF MEASUREMENT ERROR DESIGNS

In this section we first write down general model statements corresponding to the disease, measurement and exposure submodels presented in Section 1.

The underlying structure of these three submodels can be fundamentally characterized by expressing the following conditional independence assumptions:

disease model: $$[Y_i|X_i, C_i, \beta] \quad (1)$$

measurement model: $$[Z_i|X_i, \lambda] \quad (2)$$

exposure model: $$[X_i|C_i, \pi], \quad (3)$$

where i denotes the individual, β, λ and π denote model parameters, and $[U|V]$ denotes the conditional distribution of U given V. The variables in (1), (2) and (3) can be scalar or vector. Equations (1), (2) and (3) are called *model conditional distributions* (*model conditionals* for short). Since we are in a Bayesian framework, prior distributions for β, λ and π are also required, which we denote respectively by $[\beta]$, $[\lambda]$ and $[\pi]$.

Conditional independence assumptions

By asserting equations (1)–(3) as *model conditionals*, we imply far more than the conditional dependencies themselves: we also imply additional conditional *independence* assumptions which follow from the *directed Markov assumption.*[25] This states that the joint distribution of all the variables can be written as the product of all the model conditionals:

$$[\beta][\lambda][\pi]\prod_i [X_i|C_i, \pi]\prod_i [Z_i|X_i, \lambda]\prod_i [Y_i|X_i, C_i, \beta]. \quad (4)$$

In particular:

1. Equation (1) states that the disease status of individual i, Y_i, is *only* dependent on its true exposure X_i, on known covariates C_i and on unknown parameters β. We are thus in the classical case where, conditionally on the true exposure being known, the surrogate measures Z_i do not add any information on the disease status. Also as we have not included in equation (1) a dependence of Y_i on the risk factors $X_{i'}$ for other individuals i', we assert that Y_i is independent of $\{X_{i'}, i' \neq i\}$, conditionally on X_i, C_i and β. Equally, since we have not stated any dependence between the Ys conditionally on the Xs and β, we imply conditional independence.
2. Equation (2) states that by conditioning on appropriately defined parameters λ and the true exposure X_i, the surrogate measures Z_i are independent amongst individuals. The construction of λ requires careful attention and will be detailed at a second stage.

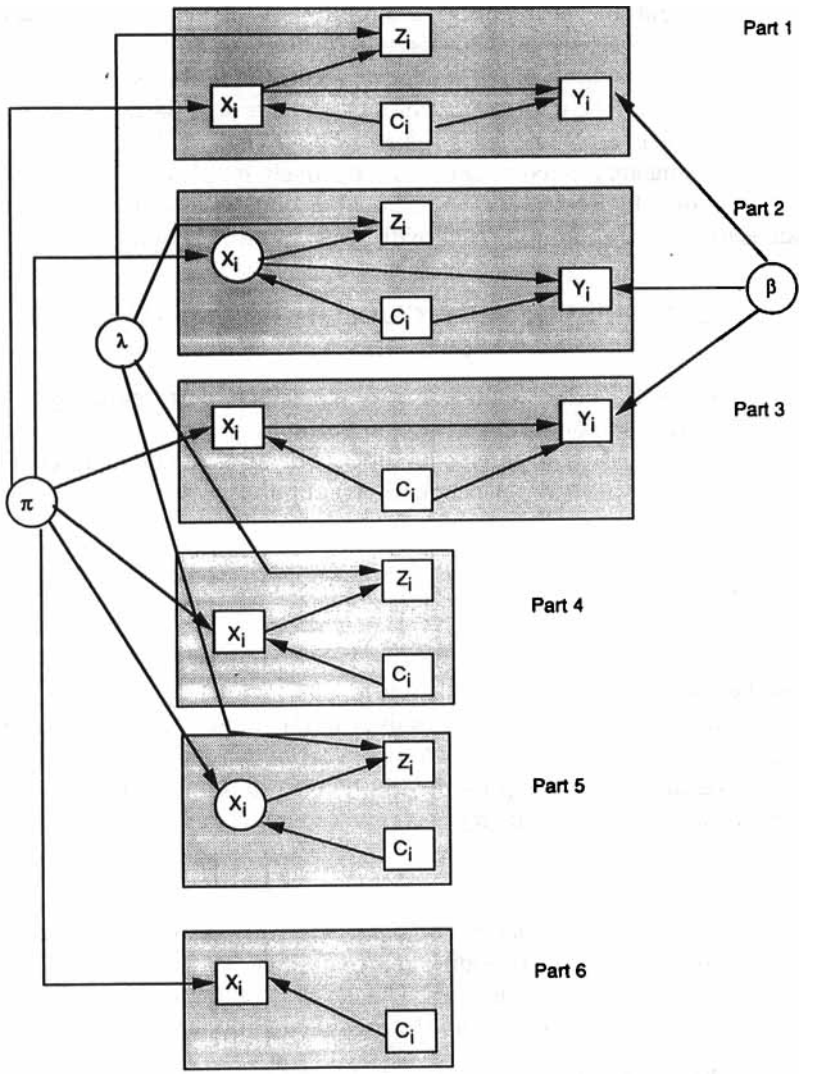

Figure 1. Graph corresponding to model equations (1), (2) and (3), illustrating the six parts of the design

3. Equation (3) models the population distribution of unknown risk factors amongst individuals in terms of parameters π. Dependence between the different components of vector X_i can be accommodated, through parameters contained in π; possible dependence between the covariates C_i and X_i is also acknowledged through conditioning on known covariates C_i.

Graph

We now set up an influence diagram or graph corresponding to equations (1) to (3) encompassing different epidemiological designs (Figure 1). This graphical representation of our model conveys conditional independence assumptions visually and helps to clarify structural relationships between different components of the model. The 'local' nature of the model, that is, that each

variable is only directly related to a few other variables, can be seen clearly from the graph. In this diagram, circular nodes denote unknown quantities (unknown risk factors and model parameters), square nodes denote known quantities (data) and edges (arrows) reflect dependencies in the model conditionals. For example, the edges entering node Y_i show that Y_i depends on X_i, C_i and β; the absence of an edge between Y_i and Z_i shows that they are conditionally independent given X_i. Each node drawn stands generically for the family generated by the subscripts, (for example, X_i represents nodes X_1, X_2, X_3, etc.) and Z_i represents all the surrogate measures (including repeats of different measuring instruments) of X_i. On a fuller graph, each individual would be represented individually and the implicit conditional assumptions between individuals would appear more clearly. As we want to concentrate on a synthetic representation of different designs, we chose this more concise graph which exhibits better the structure of the various designs.

Designs

Possible designs of epidemiological studies for dealing with measurement errors can be distinguished according to whether either or both of X_i and Z_i have been recorded and whether Y_i is known or not. Combining these designs leads to an overall design containing six parts which are generically represented in Figure 1.

Parts 1 and 4 of Figure 1 are validation studies, in which both the true exposure X_i and the surrogates Z_i are known. Validation studies are either internal, if disease status Y_i is also known for the same individuals (part 1); or external, if there is no information on the disease status (part 4). Part 2 represents the common situation where only surrogates and disease status are known. Part 3 represents a subgroup in which only the true exposure and disease status are known. Clearly if all the individuals were in this group, there would be no measurement error problem. Hence in general the number of individuals included in parts 1 or 3 will be small with respect to that of part 2. Finally parts 5 and 6 represent 'survey' situations in which information is obtained only on surrogates or on the true exposure.

The global influence diagram illustrates how information can flow from one part of the graph to another. For instance one can see that part 6 would contribute some information on π which could then be used in part 2 say, whilst part 5 contribute information on λ. This flow of information will appear clearly in the next sections when parameter estimation is considered.

Note that the nodes representing X_i are always drawn on Figure 1, whether known or unknown, whereas when Y_i or Z_i are not known, they are not represented on the graph. This is because nodes at the end of a chain of arrows (like Y_i or Z_i) can be integrated out in (4) without destroying the conditional independence structure of the graph, whilst this is not the case for intermediate nodes like X_i. Precisely, if X is not known for individual i_0, then integrating out X_{i_0} from (4) leads to a term

$$\int [X_{i_0}|C_{i_0}, \pi][Z_{i_0}|X_{i_0}, \lambda][Y_{i_0}|X_{i_0}, \beta]\, dX_{i_0}$$

which is a complicated function of $(\beta, \lambda, \pi, C_{i_0}, Z_{i_0}, Y_{i_0})$, destroying the conditional independence structure of (4).

Further structure

So far the measurement process, as formulated in equation (2), has only specified generically the relationship between the unknown covariates X_i and the surrogate measures. To use our

conditional modelling approach for a given epidemiological study, one must write down model equations corresponding to the specific measurement instruments of the study. For example in studies on diet and cancer, the covariates X_i might correspond to nutrient intakes for which surrogates have been constructed from weekly diaries of food intake and/or from a quantitative dietary history questionnaire. This particular example of a measurement error model is quite intricate and requires careful modelling. We now detail specific aspects of the measurement model relevant to commonly used epidemiological designs.

Let $X_i = (X_{i1}, \ldots, X_{ij}, \ldots, X_{iJ})'$ represent the J-dimensional vector of unknown covariates for individual i. We suppose that information on X_{ij} is available through several methods or instruments (indexed by m), each carrying out repeated measurements (indexed by r): Z_{ijmr}. The number of instruments or of repeats need not be the same for each individual. Further levels of conditional independence now need to be modelled, corresponding to the newly introduced subscripts $\{j, m, r\}$. (We retain the conditional independence amongst individuals i.) Consider the more structured equation

$$[Z_{ijmr} | X_{ij}, \lambda_{jm}] \tag{2'}$$

This equation states firstly that the surrogates indexed by j are related only to the jth unknown covariate. Secondly, for each covariate and for each instrument, there is some common structure between repeats parameterized by λ_{jm}. Thirdly, conditionally on the true value of the jth covariate and on λ_{jm}, there is independence of the surrogates Z_{ijmr} between repeats and between instruments; and finally, Z_{ijmr} is also independent of all unknown covariates $\{X_{ij'}, j \neq j'\}$. These conditional independence assumptions are illustrated in Figure 2.

Even though the formulation of equation (2') is quite general, it does not fit all measurement processes. For example the case of a multivariate link between a set of unknown covariates and a set of surrogates, as in factor analysis, is not included; nor is that of the same measuring instruments being used on several covariates. Appropriate models can easily be defined for each of these particular cases but for clarity of exposition we stay with the fairly general setting of equation (2').

In equation (2') the parameters of the measurement model were expressed globally as λ_{jm}. We now indicate a more detailed parameterization of the measurement model, decomposing λ_{jm} into a vector of parameters, each expressing different aspects of the measurement process (see Willett[1] for a discussion of types of errors in epidemiological studies). To simplify the exposition, in the remainder of this section we shall restrict ourselves to the case where only one covariate is unknown ($J = 1$), and omit subscript j.

A useful distinction is to separate the components of λ_m which are solely instrument dependent from those which are also related to the individual. Consider first the instrument-dependent parameters. If X_i and Z_{imr} are both continuous variables, instrument-dependent parameters could include, for each instrument, scaling parameters relating the continuous measures as well as precision parameters. If X_i and Z_{imr} are discrete variables then the instrument-dependent parameters would comprise the misclassification probabilities for each instrument.

The other component of measurement error is related to what might be called a systematic within-person bias. Some individuals might consistently underdeclare their exposure to a risk factor, for instance their alcohol consumption in a nutritional questionnaire. This would create a bias η_{im} which is both instrument dependent and individual dependent. Further, for two different instruments m and m' one might expect some correlation between η_{im} and $\eta_{i,m'}$. Hence this component of the measurement process can be thought of as a set of random effects $\eta_m = (\eta_{1m}, \ldots, \eta_{im}, \ldots)'$ which are related by overall parameters ω, that is through $[\eta_m | \omega]$, which model their joint distribution. It will be also necessary to specify $[\omega]$, the prior distribution

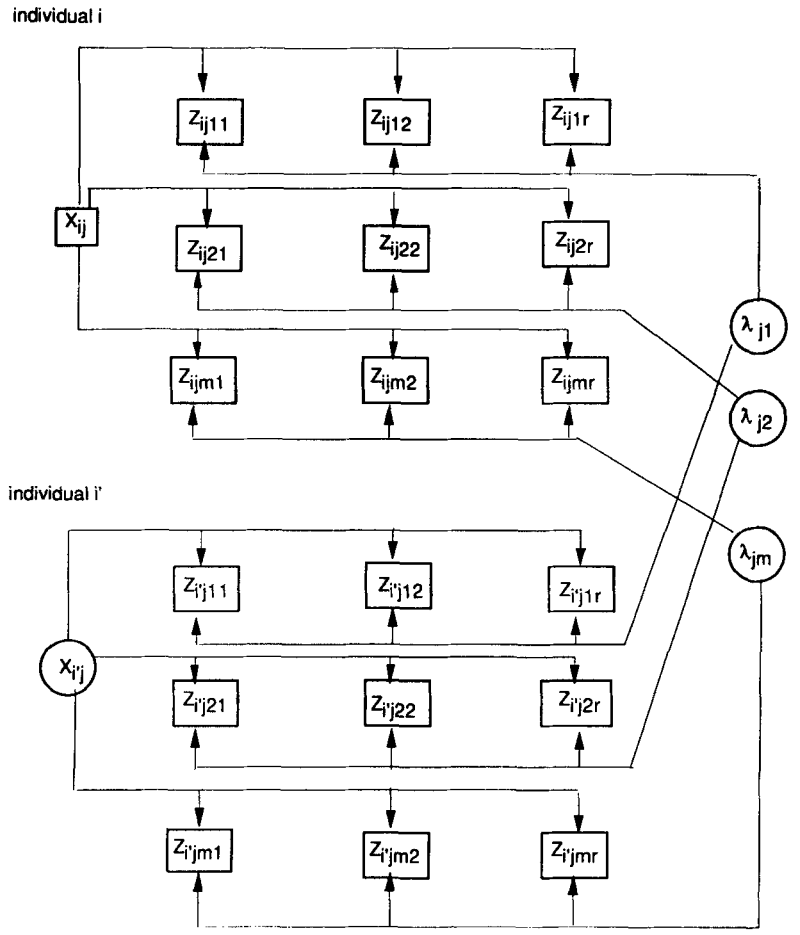

Figure 2. Graph corresponding to model equation (2')

of the parameters ω, thus introducing a further hierarchical level. For example if there are M methods, ω might correspond to an $M \times M$ covariance matrix and $[\omega]$ could be chosen as an inverse Wishart distribution.

Example

We illustrate the formulation of the measurement model by discussing the prospective study of dietary factors in breast cancer risks in a cohort of nurses analysed by Rosner *et al.*,[24] a classical example of measurement error related to dietary assessment. In this study the cohort was composed of $N = 89{,}538$ women nurses who had completed a self-administered semi-quantitative food frequency questionnaire. This group forms what we have described previously as the 'main study' (part 2). The dietary risk factors investigated were the daily saturated fat consumption (unobserved) X_1, the daily total intake of calories (unobserved) X_2, and the daily consumption of alcohol (unobserved) X_3. The intake of these three nutrients which are imperfectly apprehended by the food frequency questionnaire will be denoted correspondingly by Z_j, $1 \leqslant j \leqslant 3$.

The food frequency questionnaire was subjected to a detailed validation study conducted among 173 cohort women (part 1). In the validation group, it was assumed that the true values of the dietary intakes (X_j, $1 \leqslant j \leqslant 3$) were observed. There were no repeated measures. The women's age, denoted by C and perfectly recorded, was also introduced in the measurement model. A natural way to model this setup is to consider the simple additive error model

$$Z_{ij} = \phi_j + \psi_j X_{ij} + \gamma C_i + \varepsilon_{ij}, \tag{5}$$

with ε_{ij} independent of X_{ij}, $1 \leqslant i \leqslant N, j = 1, 2, 3$ and $\tilde{X}_i = [X_{i1}, X_{i2}, X_{i3}]'$ following a multivariate normal distribution with mean μ and variance–covariance $\sum$ for each i.

In equation (5), ϕ_j and ψ_j express the linear relationship between the true intakes and their surrogates given by the dietary questionnaire. The term ε_{ij} represents the intrinsic random fluctuation of the dietary questionnaire, which is supposed independent of X_{ij} and which could be parameterized say by a variance θ_j^{-1}, which may or may not be the same for the three nutrients. The covariance matrix $\sum$ models the associations between the nutrients, associations which are to be expected.[24]

So far we have assumed (as in Rosner *et al.*[24]) that, in the validation group, true intakes were assessed. In reality, these 'true intakes' correspond rather to measurements from another instrument: the one-week diet record, which is thought to be more precise and also possibly unbiased. Introducing the subscript m, where $m = 1$ corresponds to the food frequency questionnaire and $m = 2$ to the one-week diet record, the measurement model now becomes

$$Z_{ijm} = \phi_{jm} + \psi_{jm} X_{ij} + \gamma_m C_i + \varepsilon_{ijm}, \tag{6}$$

with ε_{ijm} independent of X_{ij}, $1 \leqslant i \leqslant N, j = 1, 2, 3, m = 1, 2$; $\phi_{j2} = 0$ and $\psi_{j2} = 1, j = 1, 2, 3$, if the second measuring instrument is thought to be unbiased; and the same hypotheses on the $\tilde{X}_i$ as for equation (5).

Further structure can still be introduced. In the design carried out by Rosner *et al.*,[24] the one-week diet record was repeated four times. Model equation (6) can be generalized to include these repeats as well as possible random effects:

$$Z_{ijmr} = \phi_{jm} + \psi_{jm} X_{ij} + \gamma_m C_i + \eta_{ijm} + \varepsilon_{ijmr}. \tag{7}$$

The term ε_{ijmr} is again independent of X_{ij} and its variance θ_{jm}^{-1} would be common to all repeats r. The term η_{ijm} is the within-person bias for individual i and covariate j of instrument m which could be modelled as a random effect dependent on a 2×2 covariance matrix ω_j. In our previous notation:

$$\lambda_{jm} = (\phi_{jm}, \psi_{jm}, \gamma_m, \theta_{jm}, \eta_{jm}),$$

where $\eta_{jm} = (\eta_{1jm}, \ldots, \eta_{ijm}, \ldots)'$ and the model conditionals corresponding to equation (2′) are

$$[Z_{ijmr} | X_{ij}, \phi_{jm}, \psi_{jm}, \gamma_m, \theta_{jm}, \eta_{ijm}].$$

Moreover, by introducing ω_j, a further hierarchical level in the modelling of the prior distribution of λ_{jm} has been specified, with $[\lambda_{jm}]$ now equal to the product $[\phi_{jm}][\psi_{jm}][\theta_{jm}][\gamma_m][\eta_{jm}|\omega_j][\omega_j]$.

Identifiability

In a Bayesian analysis, evidence provided by the data is combined with information contained in the prior distribution of the parameters to derive the posterior distribution of the parameters given the data. There are situations where scientists are unable or reluctant to formulate strong priors and choose to carry out the analysis using non-informative priors for the parameters. To be

able to estimate all the model parameters with only non-informative prior information, the data alone must contain information on all the parameters. In this situation, we say that the model is 'fully identifiable'. We now examine identifiability in particular examples of designs.

The model given in equation (7) is only fully identifiable in the case of a single instrument if the overall design of the study includes a validation group (that is, part 1 or part 4 of Figure 1) as well as repeated measurements (that is, repeats for individuals in part 1, 2, 4 or 5). In less complete designs, some parameters might not be identifiable. For example parameters ϕ_{jm} and ψ_{jm} cannot be estimated without a validation group; parameters ω_j modelling the random effects cannot be estimated without repeated measures, and this is also true for θ_{jm} in general, except if we suppose that there is no within-person bias ($\eta_{ijm} \equiv 0$) in which case θ_{jm} can be also estimated from a validation group.

If more than one instrument is present, some information can be transferred from one instrument to another and the identifiability conditions might be relaxed. Placing ourselves again in the case where there is no within-person bias for any instrument, one can see for example that parameters ϕ_{jm} and ψ_{jm} for one instrument m can be estimated *without a validation group* if another unbiased instrument m' (that is with $\phi_{m'} = 0$ and $\psi_{m'} = 1$) with repeated measures is included in the design for some individuals either in part 2 or in part 5. We can thus specify some designs, where all the measurement parameters would be identifiable even though no 'gold standard' would ever be presumed or measured.

In general validation and repeats are not necessarily on the same individuals but repeats of the different instruments on the same individuals are necessary in some cases, for instance to fully assess correlation between the random effects for different instruments.

3. ESTIMATION USING GIBBS SAMPLING

In the previous section, we described a general framework which accommodates a large class of epidemiological designs concerned with measurement error. This conditional independence modelling framework is directly linked with a unifying computational method, Gibbs sampling, which enables estimation to be successfully carried out in situations which were previously considered intractable.

Bayesian methods for estimating model parameters are based on the posterior distribution of the parameters given in the data. Gibbs sampling is a Markov chain Monte Carlo method for generating samples from the joint posterior distribution of the model parameters. It was originally proposed by Hastings[9] and rediscovered by Geman and Geman.[8] The wide applicability of the algorithm to general statistical modelling was brought out by Gelfand and Smith[26] and Gelfand *et al.*,[27] and has since been demonstrated by many authors. For reviews of work in this field, see Yiannoutsos and Gelfand,[28] Smith and Roberts[29] and Gilks *et al.*[30]

Gibbs sampling proceeds by visiting in turn each unknown variable (circular node) in the graph, and replacing the value of the variable by a new value, drawn at random from its current *full conditional distribution* (see below). Recall that unknown variables can be either model parameters or unknown exposures. A cycle of the Gibbs sampler is completed when all the unknown variables in the model have been updated once. The updating cycle is repeated a large number of times. It has been shown, under weak regularity conditions, that this process generates a Markov chain whose equilibrium distribution is the distribution of interest, that is, the joint posterior distribution of the unknown variables in the graph.[8,29,31] The current full conditional distribution of a variable U, denoted $[U|.]$, is its conditional distribution (or density) given the

current values of all other data and parameters in the model. For example the full conditional distribution for β_k, the kth element of the logistic regression parameter vector β, is

$$[\beta_k|.] \propto [\beta] \prod_{i \in \mathscr{P}_1, \mathscr{P}_2, \mathscr{P}_3} [Y_i|X_i, \beta],$$

where $\mathscr{P}_p$ denotes the set of individuals represented in part p of the graph in Figure 1. Full conditionals for a particular design are given in detail in the Appendix.

In any implementation of the Gibbs sampler, virtually the entire computational effort is in sampling from full conditional distributions. It is therefore essential that such random variate generation is done efficiently. In some situations full conditional distributions reduce analytically (through conjugacy of prior and likelihood) to standard forms, for which efficient sampling algorithms are available. For example, if both $[Y_i|X_i, \beta]$ and $[\beta]$ are normal distributions, then the full conditional $[\beta_k|.]$ will also be a normal distribution and therefore easily sampled. For full conditional distributions where no analytic reduction is possible, such as when $[Y_i|X_i, \beta]$ is of logistic form (as is commonly the case in epidemiological studies), one can use efficient black-box random-variate generation techniques such as adaptive rejection sampling[32,33] or the ratio of uniforms method.[34] Tierney[35] has proposed using the Metropolis algorithm[36] within Gibbs sampling, to avoid sampling directly from awkward full conditional distributions.

At each iteration of the Gibbs sampler, current values of parameters of interest are output to a file. This output file is used both to assess distributional convergence of the Gibbs sampler and for Bayesian inference on the parameters once 'convergence' is judged to have been reached. Methods for assessing convergence can be broadly distinguished as those monitoring ergodic averages within one run of the Gibbs sampler[37,38] or those comparing several shorter parallel runs.[39]

These issues and several other issues of practical importance when implementing the Gibbs sampler have been discussed by Geyer,[40] Besag and Green,[31] and Tierney.[35]

4. ANALYSIS OF SIMULATED DATA

In this section, we analyse several simulated data sets to illustrate some of the characteristic features of the Gibbs sampling estimation technique. These data sets reproduce some common epidemiological study designs. We start by giving a parametric formulation of the three submodels (disease, measurement and exposure) composing our design and by detailing the computations needed to run the Gibbs sampler on our examples.

Design setup

We consider a study design including only part 2 (main study) and part 4 (external validation group) of our overall design. Throughout we suppose that there are 1000 individuals in the main study and 50 individuals in the validation group. Three risk factors are involved in the disease model: two of them, X_1 and X_2, are measured with error, and the third one, C, is precisely known. We have the flexibility to suppose that the number of instruments and the number of repeated measures could vary for the two covariates and between the subjects (see Appendix), but for our illustrative purpose we shall consider in this section a simple case where the number M of instruments is the same for each covariate and the number R_m of repeated measurements of instrument m, $1 \leqslant m \leqslant M$, is the same for each individual and each part of the design.

Model conditionals

We consider the case of a logistic link between risk factors and disease status, an additive measurement error model similar to the one stated in equation (7) but without random effects, and a multivariate normal distribution for the exposure model. Precisely, corresponding to equations (1)–(3), we suppose

(i) $Y_i \sim$ Bernoulli (α_i), where logit $(\alpha_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 C_i$
(ii) $Z_{ijmr} \sim$ normal $(\phi_{jm} + \psi_{jm} X_{ij}, \theta_{jm}^{-1})$ for $j = 1, 2$, $1 \leqslant m \leqslant M$, $1 \leqslant r \leqslant R_m$
(iii) $(X_{i1}, X_{i2}, C_i) \sim \text{MVN}(\mu, \tau^{-1})$, where MVN denotes a multivariate normal, μ is a three-dimensional vector and τ^{-1} is a 3×3 variance–covariance matrix.

Prior distributions

Whenever appropriate, conjugate prior distributions were chosen. We specified $\beta_k \sim$ normal (a_k, A_k), $\phi_{jm} \sim$ normal (b_{jm}, B_{jm}), $\psi_{jm} \sim$ normal (c_{jm}, C_{jm}), $\theta_{jm} \sim$ gamma (d_{jm}, D_{jm}), $\mu \sim$ MVN (**e**, **E**) and $\tau \sim$ Wishart (f, **F**), where the parameters of these distributions are constants and are detailed below. From the model conditionals and the prior distributions, full conditional distributions for each parameter can be written (see Appendix).

The simulation study

We consider three cases, differing only in the measurement model:

Simulation 1. Each covariate is measured by one instrument (instrument 1) only and there is no replication ($M = 1, R_1 = 1$).
Simulation 2. Each covariate is measured by one instrument (instrument 1) only and there are two replications giving altogether three surrogate measures for each covariate ($M = 1, R_1 = 3$).
Simulation 3. Each covariate is measured by two instruments (instruments 1 and 2) and there is no replication ($M = 2, R_1 = R_2 = 1$).

For each of these three cases, a data set was generated by forward sampling from the model conditionals, using values of $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ given in Table I (column 'True value'), with

$$\mu = (0{\cdot}5, -0{\cdot}5, 0{\cdot}8); \qquad \tau = \begin{bmatrix} 1{\cdot}4 & -0{\cdot}1 & 0{\cdot}2 \\ -0{\cdot}1 & 1{\cdot}5 & -0{\cdot}5 \\ 0{\cdot}2 & -0{\cdot}5 & 1{\cdot}1 \end{bmatrix}.$$

The same values of Y_i and of $\{X_{ij}, j = 1, 2\}$ were used in each data set for all individuals, but surrogate measures $\{Z_{ijmr}\}$ were generated anew for each case.

We thus simulated a situation with one mildly protective risk factor X_1 (with risk ratio 0·82) and two detrimental risk factors X_2 and C (with risk ratios of 2·46 and 3·32). There was a positive correlation between X_2 and $C(r = 0{\cdot}38)$, no correlation between X_1 and $X_2(r = 0{\cdot}04)$ and a small negative correlation between X_1 and $C(r = -0{\cdot}12)$. Note also that the precisions (θ_{12} and θ_{22}) associated with instrument 2 were substantially larger than those of instrument 1 (θ_{11} and θ_{21}), whence instrument 2 is more accurate than instrument 1.

Methods of analysis

For each case we ran the Gibbs sampler for 5000 iterations this number of iterations being sufficient to ensure convergence (see Section 3). We used adaptive rejection sampling[32] to sample

Table I. Gibbs sampling analysis of the simulated data corresponding to simulations 1, 2 and 3

Parameter	True value	Simulation 1 Posterior mean	(S.D)	Simulation 2 Posterior mean	(S.D)	Simulation 3 Posterior mean	(S.D)	Analysis on true covariates Posterior mean	(S.D)
β_0	−0·8	−0·95	(0·50)	−0·54	(0·24)	−0·60	(0·18)	−0·64	(0·14)
β_1	−0·2	0·18	(0·63)	−0·24	(0·23)	−0·34	(0·14)	−0·30	(0·09)
β_2	0·9	0·90	(0·27)	1·02	(0·16)	0·99	(0·13)	1·02	(0·11)
β_3	1·2	1·09	(0·17)	1·01	(0·10)	1·02	(0·10)	1·00	(0·09)
ϕ_{11}	0·8	0·88	(0·10)	0·75	(0·08)	0·77	(0·07)		
ψ_{11}	0·4	0·18	(0·17)	0·42	(0·06)	0·52	(0·08)		
θ_{11}	0·7	0·62	(0·05)	0·72	(0·02)	0·71	(0·04)		
ϕ_{12}	0·1					0·07	(0·08)		
ψ_{12}	0·9					0·90	(0·08)		
θ_{12}	2·5					2·16	(0·31)		
ϕ_{21}	−0·3	−0·23	(0·15)	−0·12	(0·10)	−0·27	(0·09)		
ψ_{21}	1·3	1·25	(0·15)	1·33	(0·09)	1·21	(0·08)		
θ_{21}	0·6	0·68	(0·12)	0·59	(0·02)	0·61	(0·04)		
ϕ_{22}	0·0					0·02	(0·07)		
ψ_{22}	1·0					0·93	(0·06)		
θ_{22}	3·0					3·50	(0·50)		

from the full conditionals where these dit not reduce analytically to easily sampled distributions. The hyperparameters were chosen so as to assume only vague prior information, specifically: $\{a_0 = -1,\ a_k = 0$ for $k = 1, 2, 3,\ A_k = 1$ for $k = 1, \ldots, 4\}$; $\{b_{jm} = 0,\ c_{jm} = d_{jm} = 1, B_{jm} = C_{jm} = D_{jm} = 1$ for $j = 1, 2$ and $1 \leqslant m \leqslant M\}$; $\mathbf{e} = \mathbf{0}$ and $\mathbf{E} = \mathbf{I}_3$, the 3×3 identity matrix; $f = 5$ and $\mathbf{F} = 0{\cdot}2\ \mathbf{I}_3$.

Simulation 1 was also analysed with the method proposed by Rosner *et al.*[24] using software supplied by these authors. This method provides corrected estimates of the parameters $\boldsymbol{\beta}$ based on correction terms estimated from the validation group.

Results

Table I presents the results from the Gibbs sampling analysis of simulations 1, 2 and 3. Convergence was rapid and we have summarized the posterior distribution of the parameters by reporting the empirical mean and standard deviations over the first 5000 iterations. In the last column of Table I, we have given, as benchmarks, the estimate of β obtained by a Bayesian logistic regression analysis based on the knowledge of the *true values* of X_{i1} and X_{i2} for all the 1000 individuals in the main study. It is worth noticing that the estimate of β_0 is somewhat higher than its 'true value'. The results for simulation 1 show that our estimation method performs satisfactorily with all the estimated values lying well within one posterior standard deviation of the true values.

A surprising result concerns the precision of the estimates of the regression coefficients β_1 and β_2. We see that the posterior standard deviation for β_1 is much larger than for β_2. One might expect the precision of the estimate of a regression coefficient to depend critically on how well its covariate has been measured, but this does not explain our finding since the true precision of the measurements of X_1 and X_2 are roughly equal ($\theta_{11} \approx \theta_{21}$). The additional information on

β_2 would appear to derive from Y, via Y's stronger linkage (relative risk) with X_2 than with X_1. Thus information is seen to flow from Y to X_2 to β_2. Hence, in the presence of measurement error, large relative risks tend to be better estimated than those close to 1·0

In simulation 2, three surrogate values, corresponding to repeats of instrument 1, are available for each individual. We see that all posterior standard deviations have been reduced in comparison to simulation 1. Through the additional data both slopes ψ_{11} and ψ_{21} are now reasonably well estimated, leading to good estimates of all the regression coefficients and in particular to an improvement in the estimation of β_1. The estimate of β_0 is a little farther from its 'true value' than in simulation 1 but quite similar to the benchmark estimate given in the last column.

In simulation 3, we return to the situation of simulation 1 with no repeats on instrument 1, but another measuring instrument, instrument 2, has been added, which is more accurate than instrument 1. In comparison to simulations 1 and 2, there is an improvement in the posterior standard deviations which are now approaching those given in the final column. The improvement is particularly noticeable for β_1. It is worth noting that some information has passed from instrument 2 to instrument 1 with the posterior standard deviations of ϕ_{j1}, ψ_{j1} and θ_{j1}, $j = 1, 2$, smaller than those in simulation 1. This is due to better knowledge of X_{i1} and X_{i2} in simulation 3, which is then reflected in good estimates of the measurement parameters for both methods. From our results, it seems that the more precise measuring instrument has been more beneficial than repeats of the less accurate instrument.

We also simulated a combination of simulations 2 and 3 (three measurements from instrument 1 and one measurement from instrument 2). The results showed little improvement in the posterior standard deviations of the regression coefficients compared with simulation 3 (results not shown).

To compare the results from our Bayesian analysis with an alternative classical approach, we reanalysed the data from simulation 1 using the method of Rosner *et al.*,[24] which estimates the measurement model parameters (ϕ, ψ, and θ) from the validation study (part 4), and substitutes these estimates into the logistic regression in the main study (part 2), with an approximate adjustment to standard errors for the regression coefficients. It is worth noting at this point that this widely referenced approach does not ideally suit our situation: in particular it assumes a rare disease (whereas 44 per cent of individuals are diseased in our data). Moreover, Rosner's method does not generalize to include both repeat and validation data (as in simulations 2 and 3).

The analysis of simulation 1 by Rosner's method is given in Table II, which also gives crude estimates obtained by a classical (maximum likelihood) logistic regression on Z_1, Z_2 and C, ignoring measurement error. The crude estimates are seriously biased, being many standard deviations from their true values. The corrected estimates from Rosner's method have much larger standard errors, so that true values lie well within 95 per cent confidence intervals of the corrected estimates. We see nevertheless that β_1 is poorly estimated.

The final column of Table II gives estimates from a classical logistic regression on the true values X_1, X_2 and C. These estimates closely resemble those from the Bayesian analysis of the same data, reported in the final column of Table I. This is to be expected since the Bayesian analysis did not use strong prior information.

5. DISCUSSION

In this paper we have constructed a unifying representation of the structure of measurement error problems with particular reference to situations commonly encountered in epidemiological studies. Our approach emphasizes the structural aspects of the problem, that is the conditional dependence and conditional independence assumptions. Only after having posited these do we

Table II. Logistic regression analysis of the simulated data corresponding to simulation 1: crude and corrected estimates given by Rosner *et al.*'s[24] method

Regression coefficient	True value	Analysis of simulation 1		Analysis on true covariates*
		Crude estimate (SE)	Corrected estimate (SE)	Estimate (SE)
β_1	−0·2	0·02 (0·06)	0·84 (0·86)	−0·30 (0·09)
β_2	0·9	0·29 (0·05)	0·67 (0·23)	1·03 (0·09)
β_3	1·2	1·12 (0·09)	1·13 (0·21)	1·00 (0·09)

* Logistic regression analysis of simulation 1 based on the true values of X_1 and X_2 for alfi individuals.

then go on to consider detail distributional forms. We have outlined how estimation of the parameters of interest can be carried out in a Bayesian framework using Gibbs sampling. We have shown how this approach can be implemented for designs involving continuous measurement errors assessed through a validation substudy, and have compared in a simple case our results with those given by the method proposed by Rosner *et al.*[24]

There is an enormous literature on methods for obtaining correct estimation of parameters of the response function in the presence of measurement error on the covariates. Possible approaches, together with their relative merits, have been discussed and characterized by Caroll[41] in the following way: 'maximum likelihood for a specified parametric family; quasi-likelihood generalized least squares; and exploitation of special features of certain models', with further distinction according to whether Taylor series approximations are used or not. Conditional independence modelling, coupled with estimation via Gibbs sampling, possesses distinct advantages over methods previously proposed. It is conceptually straightforward and follows exactly the structure of the error problem without recourse to artificial assumptions or to approximations; it utilizes all the information in the model specification (not just first and second moments); it deals faithfully with all sources of errors and propagates the resulting uncertainty through to the parameter estimates; it is flexible, treats symmetrically continuous or discrete covariates, can handle missing data and ancillary data, and can combine different data sources. One disadvantage of this method is that it is computationally intensive. We now briefly discuss each of these points.

Structure of the error model

Any method proposed for correcting estimates in the presence of measurement error relies heavily on assumptions about the structure of the error model. Two types of additive error model linking a true covariate X and its surrogate Z have traditionally been considered. In the 'classical error model', Z is modelled as a function of X and the measurement error is assumed independent of the true covariate X. In the 'Berkson error model', it is X which is modelled as a function of Z and the measurement error is assumed independent of the surrogate Z.

It is not always straightforward to choose which formulation (classical or Berkson) is more suited to the particular problem which is being modelled. Berkson models arise mostly in experimental settings where the variable Z is controlled and it is then reasonable to suppose that the distribution of X is entirely specified by that of Z. This strong assumption is less tenable in the epidemiological context, where we would argue that true cases of Berkson error are quite rare. In dietary studies where measurement error is a major concern, the error model is naturally classical but Berkson error models have sometimes been used for computational convenience.[16] Caroll[41]

points to some difficulties which could arise from this contrived formulation. Berkson error models have also been used in environmental epidemiology, for example in modelling individual exposure to indoor pollution by functionally relating individual absorption X of nitrogen dioxide (NO_2) to observed mean levels Z of NO_2 in the surrounding rooms.[42] But the correct error model in this case should relate X not directly with Z but rather to unobserved population parameters of the distribution of Z. In this way, fluctuations of Z around its population parameters would also enter in the modelling of the error. This complex type of error structure also arises in occupational epidemiology: see Gilks and Richardson[43] and Armstrong.[44]

As classical and Berkson error models apply in different situations, methods should be able to tackle both. Nonetheless most maximum likelihood or quasi-likelihood methods have been formulated in the Berkson setting, requiring knowledge of the conditional distribution of X given Z or of some of its moments. Bayarri and DeGroot[45] note that in measurement error models, ambiguities in the definition of the likelihood function can arise. Indeed, there might be no consensus on whether and how to include the unobserved true covariates X in the likelihood.

In the Bayesian framework, an analogous distinction between the two measurement models can also be made. The difference between the two measurement models is now related to their specific conditional independence assumptions. Thus for the classical error case, the distribution $[Z|X, \lambda]$ is modelled, and additional conditional independence assumptions imply that the joint distribution of the measurement variables and parameters can be written as: $[X][\lambda][Z|X, \lambda]$. For the Berkson case, $[X|Z, \lambda^*]$ is modelled and additional conditional independence assumptions imply that the joint distribution can be written as $[Z][\lambda^*][X|Z, \lambda^*]$. We have formulated our approach in the classical error setting but we could have done it as easily in the Berkson setting; arrows would now point from Z to X and there would be no need to specify a prior distribution for X.

Exact inference

The development of our approach does not require any approximations; hence there are no restrictions for its implementation in varied epidemiological settings. For instance, the rare disease assumption required by the approximations made in Rosner's method limits its applicability to cohort studies of rare chronic diseases. It is also worth emphasizing that exact inference can also be made on any functions (for example, ratios) of the parameters of interest. Indeed, the posterior distribution of any function of the parameters is immediately available from the samples generated by the Gibbs sampler.

Computational considerations

Although conceptually straightforward, the Gibbs sampler is less straightforward to implement, owing to the current lack of general purpose software. To date, all implementations of Gibbs sampling have involved writing computer programs in low-level languages such as Fortran or C. This for the time being might inhibit wider applications of the methodology. However, Thomas *et al.*[46] have reported on a project to develop software for implementing the Gibbs sampler, allowing natural model specification, and containing an expert system to decide on appropriate methods for sampling from full conditionals such as conjugacy or adaptive rejection sampling. The model conditionals that we have stated in our paper would be accommodated by this software. Another problem arising in the implementation of the Gibbs sampler is the assessment of convergence. Determining in practice when the Gibbs sampler has converged is currently an area of debate. The main area of the debate concerns whether one should assess convergence on one long run of the Gibbs sampler, or by comparing several shorter parallel runs. Another issue

concerns whether one should assess convergence on one univariate summary of each iteration (for example, one parameter) or by considering the parameters jointly. These issues and several others have been discussed by several authors.[31,35,40]

Parametric approach

At each step of our modelling approach, conditional distributions need to be explicitly specified. These conditional distributions are then fully used in the Gibbs sampling algorithm. The choice of distributional forms is not constrained by computational considerations. Indeed conjugacy is not an issue since efficient methods exist for sampling from most full conditional distributions commonly encountered.[32] Thus distributional assumptions can be tailored naturally to the application. On the other hand, there might be cases where specifying an appropriate parametric model is awkward. The need to relax the fully parametric approach has been discussed by several authors. Caroll and Wand[6] propose a non-parametric kernel regression on the validation subset to estimate the probability function linking the outcome variable and the surrogate Z. Thomas *et al.*[10] discuss some of the difficulties linked to the specification of the population distribution of X and consider a non-parametric exposure model which is assumed to be concentrated at a finite number of points. This type of non-parametric approach can be combined in principle with Gibbs sampling, as has been shown by Thomas *et al.*,[10] but there is scope for further work in this direction. Another interesting aspect which deserves further attention is the robustness of the results to a wrong parametric choice for the exposure model.

Propagation of uncertainty

There are different levels of uncertainty in our model: use of surrogate instead of true values, uncertainty on the parameters, and uncertainty on the hyperparameters. Gibbs sampling takes into account all these sources of uncertainty without resorting to approximations, and propagates them through to the parameters of interest (in contrast to the E-M algorithm, for example, which assumes fixed hyperparameters). The graph itself nicely brings out the flow of information from the validation substudies to the main study and vice versa. Hence imprecision on the parameters λ of the measurement model is fully relayed to the variability of the regression parameters. Furthermore our model gives as natural framework for combining information from several measuring instruments, a design which is particularly relevant in cases where 'the true exposure' can never be measured.

Generalizability

Misclassification of discrete covariates has traditionally been treated in the literature quite separately from the case of measurement error of continuous covariates. The flexibility of the graphical modelling approach allows it to treat both these cases, as well as any mixture of mismeasured discrete and continuous covariates, in the same conceptual framework. The only modifications required for implementation concern the detailed specification of the measurement model. Besides, more complex error models like those occurring in occupational epidemiology, resulting in a combination of classical and Berkson error models, can also be handled.[43] Globally this approach allows one to integrate different data sources, for example several validation groups or ancillary data arising from an independent survey, so long as one is prepared to explicate their common structure. An added benefit is that in doing so the understanding of the structure of the underlying problem is enhanced. It is also easy to see that missing data, which can be treated as additional unknown parameters, present no particular problem. In addition, the Bayesian

framework allows prior information on regression or measurement parameters to be accommodated naturally.

By adopting the natural causal ordering between exposure and disease, we have cast our graphical model in the framework of cohort studies. Thomas *et al.*[10] have developed graphical models for measurement error for matched case-control studies. Adaptation of our methodology to case-control studies can be envisaged but requires further work.

Another issue which deserves attention is how to use our methodology in *designing* studies. As the posterior distribution of the parameters is usually not tractable analytically, this is a challenging question.

APPENDIX

We consider here the general six-part design represented in Figure 1, giving full conditionals for parameters which are updated during a cycle of the Gibbs sampling algorithm. We use the same notation as in the text, but to accommodate greater generality we define

$$\mathbf{X}_i^* = (\mathbf{X}_i', \mathbf{C}_i')',$$

where $\mathbf{X}_i$ is a vector of J_X potentially unknown covariates for individual i, and $\mathbf{C}_i$ is a vector of J_C known covariates; M_{jp} is the number of instruments used to measure the jth covariate of any individual in part p; R_{jmp} is the number of measurements performed with instrument m to measure the jth covariate for individuals is part p.

The full conditional for X_{ij} for an individual i in part 2 is

$$\ln[X_{ij}|.] \propto -\tfrac{1}{2}(\mathbf{X}_i^* - \boldsymbol{\mu})'\boldsymbol{\tau}(\mathbf{X}_i^* - \boldsymbol{\mu}) + Y_i \ln(\alpha_i) + (1 - Y_i)\ln(1 - \alpha_i)$$

$$-\frac{1}{2}\sum_{m=1}^{M_{j2}} \theta_{jm} \sum_{r=1}^{R_{jm2}} (Z_{ijmr} - (\phi_{jm} + \psi_{jm}X_{ij}))^2$$

for $j = 1, \ldots, J_X$, where logit $\alpha_i = [\beta'(1, \mathbf{X}_i^*)]$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_{J_X+J_C})'$.

For an individual i belonging to part 5, this full conditional simplifies to

$$\ln[X_{ij}|.] \propto -\frac{1}{2}(\mathbf{X}_i^* - \boldsymbol{\mu})'\boldsymbol{\tau}(\mathbf{X}_i^* - \boldsymbol{\mu}) - \frac{1}{2}\sum_{m=1}^{M_{j5}} \theta_{jm} \sum_{r=1}^{R_{jm5}} (Z_{ijmr} - (\phi_{jm} + \psi_{jm}X_{ij}))^2.$$

Let n be the number of individuals in all six parts of the study. The full conditionals for $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ are given by

$$\boldsymbol{\mu} \sim \text{MVN}\left((\mathbf{E} + n\boldsymbol{\tau})^{-1}\left(\mathbf{E}\mathbf{e} + \boldsymbol{\tau} \sum_{i \in \mathscr{P}_1 \text{ to } \mathscr{P}_6} \mathbf{X}_i^*\right), (\mathbf{E} + n\boldsymbol{\tau})^{-1}\right)$$

$$\boldsymbol{\tau} \sim \text{Wishart}\left(f + n, \mathbf{F}^{-1} + \sum_{i \in \mathscr{P}_1 \text{ to } \mathscr{P}_6} (\mathbf{X}_i^* - \boldsymbol{\mu})(\mathbf{X}_i^* - \boldsymbol{\mu})'\right).$$

The full conditional for the logistic regression parameter β_k is

$$\ln[\beta_k|.] \propto -\frac{1}{2}(\beta_k - a_k)^2/A_k + \sum_{i \in \mathscr{P}_1, \mathscr{P}_2, \mathscr{P}_3} Y_i \ln(\alpha_i) + (1 - Y_i)\ln(1 - \alpha_i)$$

for $k = 0, \ldots, J_X + J_C$.

Let n_p denote the number of individuals in part p. The full conditionals for the measurement parameters are given by

$$[\phi_{jm}|.] \sim \mathrm{N}(\mu_{\phi jm}, \sigma_{\phi jm})$$

with

$$\sigma_{\phi jm} = \left(B_{jm}^{-1} + \theta_{jm} \sum_{p=1,2,4,5} n_p R_{jmp} \right)^{-1}$$

and

$$\mu_{\phi jm} = \sigma_{\phi jm} \left[b_{jm} B_{jm}^{-1} + \theta_{jm} \sum_{p=1,2,4,5} \sum_{i \in \mathscr{P}_p} \sum_{r=1}^{R_{jmp}} (Z_{ijmr} - \psi_{jm} X_{ij}) \right]$$

$$[\psi_{jm}|.] \sim \mathrm{N}(\mu_{\psi jm}, \sigma_{\psi jm})$$

with

$$\sigma_{\psi jm} = \left(C_{jm}^{-1} + \theta_{jm} \sum_{p=1,2,4,5} \sum_{i \in \mathscr{P}_p} R_{mjp} X_{ij}^2 \right)^{-1}$$

and

$$\mu_{\psi jm} = \sigma_{\psi jm} \left[c_{jm} C_{jm}^{-1} + \theta_{jm} \sum_{p=1,2,4,5} \sum_{i \in \mathscr{P}_p} \sum_{r=1}^{R_{jmp}} (Z_{ijmr} - \phi_{jm}) \right]$$

$$[\theta_{jm}|.] \sim \mathrm{gamma}\left(d_{jm} + \frac{1}{2} \sum_{p=1,2,4,5} n_p R_{jmp}, D_{jm} + \frac{1}{2} \sum_{p=1,2,4,5} \sum_{i \in \mathscr{P}_p} \sum_{r=1}^{R_{jmp}} (Z_{ijmr} - (\phi_{jm} + \psi_{jm} X_{ij}))^2 \right).$$

REFERENCES

1. Willett, W. 'An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies', *Statistics in Medicine*, **8**, 1031–1040 (1989).
2. Gail, M. H. 'A bibliography and comments on the use of statistical models in epidemiology in the 1980s', *Statistics in Medicine*, **10**, 1819–1885 (1991).
3. Kleinbaum, D. G., Kupper, L. L. and Morgenstern. H. *Epidemiologic Research*, Lifetime Learning Publications, 1982.
4. Greenland, S. 'The effect of misclassification in the presence of covariates', *American Journal of Epidemiology*, **112**, 564–569 (1980).
5. Liu, X. and Liang, K. J. 'Adjustment for non-differential misclassification error in the generalized linear model', *Statistics in Medicine*, **10**, 1197–1211 (1990).
6. Caroll, R. J. and Wand, M. P. 'Semiparametric estimation in logistic measurement error models', *Journal of the Royal Statistical Society, Series B*, **53**, 573–585 (1991).
7. Whittaker, J. *Graphical Models in Applied Multivariate Statistics*, Wiley, New York, 1990.
8. Geman, S. and Geman, D. 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *Institute of Electrical and Electronic Engineers Transactions in Pattern Analysis and Machine Intelligence*, **6**, 721–741 (1984).
9. Hastings, W. K. 'Monte-Carlo sampling methods using Markov chains and their applications', *Biometrika*, **57**, 97–109 (1970).
10. Thomas, D. C., Gauderman, J. and Kerber, R. 'A non-parametric Monte Carlo approach to adjustment for covariate measurement errors in regression analysis', *Biometrics* (in press).
11. Stephens, D. A. and Dellaportas, P. 'Bayesian analysis of generalised linear models with covariate .measurement error', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Bayesian Statistics 4*', Oxford University Press, Oxford, 1992.
12. Clayton, D. G. 'Models for the analysis of cohort and case-control studies with inaccurately measured exposures', in Dwyer, Feinleib, Lippert and Hoffmeister, (eds.), *Statistical Models for Longitudinal Studies of Health*, Oxford University Press, New York, 1992.
13. Marshall, J. R. 'The use of dual or multiple reports in epidemiologic studies', *Statistics in Medicine*, **8**, 1041–1049 (1989).

14. Armstrong, B. 'Measurement error in the generalized linear model', *Communication in Statistics, Series B*, **14**, 529–544 (1985).
15. Schafer, D. W. 'Covariate measurement error in generalised linear models', *Biometrika*, **74**, 385–391 (1987).
16. Rosner, B., Spiegelman, D. and Willett, W. C. 'Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error', *Statistics in Medicine*, **8**, 1051–1069 (1989).
17. Stefanski, L. A. and Caroll, R. J. 'Covariate measurement error in logistic regression', *The Annals of Statistics*, **13**, 1335–1351 (1985).
18. Whittemore, A. S. and Keller, J. B. 'Approximations for regression with covariate measurement error', *Journal of the American Statistical Association*, **83**, 1057–1066 (1988).
19. Chesher, A. 'The effect of measurement error', *Biometrika*, **78**, 451–462 (1991).
20. Caroll, R. J. and Stefanski, A. 'Approximation quasi-likelihood estimation in models with surrogate predictors', *Journal of the American Statistical Association*, **85**, 652–663 (1990).
21. Espeland, M. and Hui, S. L. 'A general approach to analysing epidemiological data that contain misclassification error', *Biometrics*, **43**, 1001–1012 (1987).
22. Chen, T. T. 'A review of methods for misclassified categorical data in epidemiology', *Statistics in Medicine*, **8**, 1095–1106 (1989).
23. Duffy, S. W., Maximovitch, D. and Day, N. E. 'External validation, repeat determination and precision of risk estimation in misclassified exposure data in epidemiology', *Journal of Epidemiology and Community Health*, **46**, 620–624 (1992).
24. Rosner, B., Spiegelman, D. and Willett, W. C. 'Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariate measured with error', *American Journal of Epidemiology*, **132**, 734–745 (1990).
25. Lauritzen, S. L., David, A. P., Larsen, B. N. and Leimer, H. G. 'Independence properties of directed Markov fields', *Networks*, **20**, 491–505 (1990).
26. Gelfand, A. E. and Smith, A. F. M. 'Sampling based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1990).
27. Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. 'Illustration of Bayesian inference in normal data models using Gibbs sampling', *Journal of the American Statistical Association*, **85**, 972–985 (1990).
28. Yiannoutsos, C. T. and Gelfand, A. E. 'Subgraph approximations for large Directed graph models', to be published in *Statistical Decision Theory and Related Topics, V.*
29. Smith, A. F. M. and Roberts, G. O. 'Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society, Series B*, **55**, 3–23 (1993).
30. Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best., N. G., McNeil, A. J. Sharples, L. D. and Kirby, J. 'Modelling complexity: applications of Bibbs sampling in medicine', *Journal of the Royal Statistical Society, Series B*, **55**, 39–52 (1993).
31. Besag, J. and Green, P. J. 'Spatial statistics and Bayesian computation', *Journal of the Royal Statistical Society, Series B*, **55**, 25–37 (1993).
32. Gilks, W. R. and Wild, P. 'Adaptive rejection sampling for Gibbs sampling', *Applied Statistics*, **41**, 337–348 (1992).
33. Gilks, W. R. 'Derivative-free adaptive rejection sampling for Gibbs sampling', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992.
34. Wakefield, J., Gelfand, A. E. and Smith, A. F. M. 'Efficient generation of random variates via the ratio-of-uniforms method', *Statistics and Computing*, **1**, 129–133 (1991).
35. Tierney, L. 'Exploring posterior distributions using Markov chains', *Computer Science and Statistics*, Proceedings of the 23rd symposium on the interface. Karamidas, E. M. (ed.), Interface Foundation, Fairfax Station, VA, pp 563–570.
36. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. 'Equations of state calculations by fast computing machine', *Journal of Chemical Physics*, **21**, 1087–1091 (1953).
37. Raftery, A. and Lewis, S. 'How many iterations of the Gibbs sampler?', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Baysian Statistics 4*, Oxford University Press, Oxford, 1992.
38. Geweke, J. 'Evaluating the accuracy of sampling based approaches to the calculation of posterior moments', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992.

39. Gelman, A. and Rubin, D. B. '*Inferences from Iterative Simulation using multiple sequences*', Statistical Science, **7**, 457–472 (1992).
40. Geyer, C. J. 'Practical Markov Chain Monte-Carlo', *Statistical Science*, **7**, 473–511 (1992).
41. Caroll, R. J. 'Covariance analysis in generalized linear measurement error models', *Statistics in Medicine*, **8**, 1075–1093 (1989).
42. Tosteson, T. D., Stefanski, L. A. and Schafer, D. W. 'A measurement-error model for binary and ordinal regression', *Statistics in Medicine*, **8**, 1139–1147 (1989).
43. Gilks, W. and Richardson, S. 'Analysis of disease risks using ancillary risk factors, with application to job-exposure matrices', *Statistics in Medicine*, **11**, 1443–1463 (1992).
44. Armstrong, B. G. 'The effects of measurement errors on relative risk regression', *American Journal of Epidemiology*, **132**, 1176–1184 (1990).
45. Bayarri, M. J. and DeGroot, M. B. 'Difficulties and ambiguities in the definition of a likelihood function', *Journal of the Italian Statistical Society*, **1**, 1–15 (1992).
46. Thomas, A., Spiegelhalter, D. and Gilks, W. R. 'BUGS: a program to perform Bayesian inference using Gibbs sampling', in Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds.), *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992.