Bayesian Analysis of Misclassified Binary Data from a Matched Case-Control Study with a Validation Sub-Study

Gordon J Prescott[1] **and Paul H Garthwaite**[2*]

[1]Department of Public Health, University of Aberdeen, AB25 2ZD, U.K.

[2]Department of Statistics, The Open University, Milton Keynes, MK7 6AA, U.K.

Short title: Misclassification in a matched study

SUMMARY

Bayesian methods are proposed for analysing matched case-control studies in which a binary exposure variable is sometimes measured with error, but whose correct values have been validated for a random sample of the matched case-control sets. Three models are considered. Model 1 makes few assumptions other than randomness and independence between matched sets, while Models 2 and 3 are logistic models, with Model 3 making additional distributional assumptions about the variation between matched sets. With Models 1 and 2 the data are examined in two stages. The first stage analyses data from the validation sample and is easy to perform; the second stage analyses the main body of data and requires MCMC methods. All relevant information is transferred between the stages by using the posterior distributions from the first stage as the prior distributions for the second stage. With Model 3, a hierarchical structure is used to model the relationship between the exposure probabilities of the matched sets, which gives the potential to extract more information from the data. All the methods that are proposed are generalized to studies in which there is more than one control for each case. The Bayesian methods and a maximum likelihood method are applied to a data set for which the exposure of every patient was measured using both an imperfect measure that is subject to misclassification, and a much better measure whose classifications may be treated as correct. To test methods, the latter information was suppressed for all but a random sample of matched sets.

Key Words: Errors in variables; Matched case-control; Measurement error; Misclassification; Odds ratio.

* Correspondence to: Paul H Garthwaite, Department of Statistics, Open University, Milton Keynes, MK7 6AA, UK. E-mail: p.h.garthwaite@open.ac.uk

1. INTRODUCTION

Matched case-control studies are frequently used to examine the association between a binary exposure variable and the presence or absence of disease. Misclassification of an exposure variable due to forgetfulness or mis-reporting is a problem in many studies, and can be particularly important where information on exposure is obtained retrospectively. If a non-trivial level of misclassification is anticipated then this should be allowed for in the design and analysis of the study, as ignoring the misclassification can lead to biased estimates and inaccurate standard errors[1,2].

If a perfect, but expensive "gold-standard" measure is available for the exposure variable, then this can be compared to the ordinary measure for a subset of the study group. This enables adjustments to be made in the statistical analysis for misclassification in the main part of the study. Situations where a gold-standard measure is available can arise in various fields. For example, in epidemiology a cheap and low-effort source of retrospective information is a postal questionnaire, while a case-note review provides a more expensive but reliable source of information. Obtaining the gold-standard information from a case-note review may be too time-consuming to perform for the complete study group but may be viable for subset of it, while, for most of the study group, information might be obtained only through postal questionnaires. Similarly, in the case of occupational exposures to chemicals, an individual may not recall with complete accuracy the chemicals they were working with, while employers' records may hold this information. A full search of this information may be too demanding for the entire number of people surveyed and might be performed for just a subset.

In the early 1980s Greenland[3] showed that misclassification bias may be increased or decreased by matching and noted that bias due to misclassification is a function of the closeness of the matching. He presented a simple frequentist correction procedure, using classification rates to estimate a corrected odds ratio. Greenland and Kleinbaum[4] applied the correction for misclassification to matched case-control data using an example constructed to have similar misclassification rates as they had found with unmatched data. Many other authors have used these data. Greenland and Kleinbaum[4] highlighted the dangers of using misclassification rates from external sources where the study design may have influenced them away from population misclassification rates. An estimate of an asymptotic covariance for the true matched pair

frequencies was developed later,[5] from which an approximate confidence interval for the odds ratio can be calculated. However this simple matrix method does not allow for mis-specification of the misclassification rates from the validation study without the variance estimates for the parameters becoming substantially more complex.

Bayesian methods have been applied to *unmatched* case-control studies involving measurement error[6–8] but work on applying Bayesian methods to matched studies with measurement error has been very limited. Some methods that have been developed for cohort studies[9–13] are relatively general and, in principle, they could be extended to matched case-control studies, but this has yet to be done. York *et al.*[14] examined a matched case-control example using a Bayesian approach, but they assumed no gold standard measure and instead they placed great restrictions on the relationship between information from two sources that were both subject to error. Thurigen *et al.*[15] gave a wide ranging review of frequentist and Bayesian methods in which a validation sample is used to correct for misclassification.

In this paper we suppose a matched case control study has been conducted to relate a binary exposure variable to an outcome, such as presence (for cases) or absence (for controls) of a disease. For the whole study, exposure has been classified using a measure that is subject to error and, for a random sample of matched sets, it has also been classified using a gold standard measure. In the first instance we suppose a matched set is just a matched pair, but results are extended to the situation where there is more than one control for each case. With the Bayesian methods, the procedures that give point estimates of the effect of exposure also give interval estimates with no further effort. It is much harder to use frequentist methods when there is misclassification and more than one matched control for each case; the only method we have found in the literature makes the strong assumption that misclassification rates are known.[5]

Misclassification rates are said to be non-differential if they are the same for both cases and controls; otherwise they are differential. With frequentist methods based on matrix methods[3,4], the analysis for differential misclassification is only slightly harder than for non-differential misclassification. However, for more complex frequentist approaches it is more difficult to model differential misclassification and the assumption of non-differential misclassification is often made. The review of Thurigen *et al.*[15] lists only five non-Bayesian papers that allowed for differential misclassification and only one of these was for matched case-control data.[16] With

Bayesian methods, though, differences between the differential and non-differential cases can be handled with minor modification. For the most part, we address the more complex case where misclassification is differential, but modifications needed for the non-differential case are given.

In Section 2 we suppose that misclassification is not present. Even for this simpler situation, the appropriate analysis for a matched case-control study is not clear-cut and we consider three models that have been advocated. The models differ in the assumptions they make. In Section 3 we develop Bayesian methods for them that use data from a validation study to allow for misclassification. Methods for the first two models are a natural extension of a method developed by Prescott and Garthwaite[7] that allows for misclassification in an unmatched study. The third model is a hierarchical model that requires a markedly different approach. In Section 3 it is assumed that there is a single control for each case, but the methods are extended in Section 4 so that there can be several controls for each case. In Section 5 various methods are applied to a data set and their results compared. Non-differential misclassification is considered briefly in Section 6 and concluding comments are given in Section 7. The methods presented in this paper are analytically intractable and were implemented using MCMC methods and WinBUGS software[17].

2. OVERVIEW OF MODELS WITHOUT MISCLASSIFICATION

In this section we introduce some notation and describe the three models that are considered, without allowing for misclassification at this stage. Let N denote the total number of case-control pairs in the study and suppose that a random sample of n_1 of these pairs form a validation sub-study group and the remaining (n_2) pairs form the main study group. The gold standard measure, known only in the validation group, determines true exposure for an individual, E, where $E = 1$ indicates exposure and $E = 0$ indicates non-exposure. A case (disease present) is indicated by $D = 1$ and a control (disease absent) by $D = 0$.

The primary aim in most case-control studies is to estimate the odds ratio,

$$OR = \frac{\Pr[D = 1 \,|\, E = 1]}{\Pr[D = 0 \,|\, E = 1]} . \frac{\Pr[D = 0 \,|\, E = 0]}{\Pr[D = 1 \,|\, E = 0]}$$

which may be expressed as

$$OR = \frac{\Pr[E = 1 \,|\, D = 1]}{\Pr[E = 0 \,|\, D = 1]} . \frac{\Pr[E = 0 \,|\, D = 0]}{\Pr[E = 1 \,|\, D = 0]}. \tag{1}$$

The logarithm of the odds ratio is also often estimated.

Model 1

Suppose a matched case-control pair is picked at random and classified according to their exposure. Let (E_1, E_2) denote their true exposure status, where E_1 represents the exposure status of the case and E_2 that of the control. Then (E_1, E_2) equals (1, 1), (1, 0), (0, 1) or (0, 0), and we denote the probabilities of each of these outcomes by θ_{ij}, where

$$\theta_{ij} = \Pr[(E_1, E_2) = (i, j)] \quad \text{for } i = 0, 1;\ j = 0, 1. \tag{2}$$

Model 1 is the simplest model and makes no assumptions about any relationships between the θ_{ij} probabilities, except that they must sum to 1. For this model, the odds ratio in equation (1) is given by

$$OR = \frac{(\theta_{10} + \theta_{11})}{(\theta_{00} + \theta_{01})} . \frac{(\theta_{00} + \theta_{10})}{(\theta_{01} + \theta_{11})}. \tag{3}$$

Models 2 and 3

Models 2 and 3 are forms of the logistic model; the most common model for analysing matched case-control studies. For these models, individual pairs must be identified. Let (E_{1k}, E_{2k}) denote the value of (E_1, E_2) for the kth pair $(k = 1, \ldots, N)$. Analogous to equation (2), define

$$\theta_{ijk} = \Pr[(E_{1k}, E_{2k}) = (i, j)] \quad \text{for } i = 0, 1;\ j = 0, 1;\ k = 1, \ldots, N. \tag{4}$$

Under the logistic model, for $k = 1, \ldots, N$,

$$\Pr[E_{1k} = 1] = \frac{\exp(\beta_k + \delta_1)}{1 + \exp(\beta_k + \delta_1)} \tag{5}$$

and

$$\Pr[E_{2k} = 1] = \frac{\exp(\beta_k + \delta_2)}{1 + \exp(\beta_k + \delta_2)}. \tag{6}$$

To make the parameters identifiable a constraint must be imposed. For Bayesian models we assume $\mathcal{E}(\beta_k) = 0$, where $\mathcal{E}$ denotes expectation over all pairs. For frequentist models we make the usual assumption that $\delta_2 = 0$. The logistic model also assumes conditional independence (given β_k) within a pair,

$$\Pr[(E_{1k}, E_{2k} = (i, j) \,|\, \beta_k)] = \Pr(E_{1k} = i \,|\, \beta_k).\Pr(E_{2k} = j \,|\, \beta_k) \tag{7}$$

for $i = 0, 1$; $j = 0, 1$; $k = 1, \ldots, N$. From equation (1), the odds ratio for the kth pair is

$$OR = \frac{\exp(\beta_k + \delta_1)/\{1 + \exp(\beta_k + \delta_1)\}}{1/\{1 + \exp(\beta_k + \delta_1)\}} \cdot \frac{1/\{1 + \exp(\beta_k + \delta_2)\}}{\exp(\beta_k + \delta_2)/\{1 + \exp(\beta_k + \delta_2)\}}$$
$$= \exp(\delta_1 - \delta_2). \tag{8}$$

Equation (8) implies that it is meaningful to talk about 'the odds-ratio', as the ratio does not vary with k.

The θ_{ij} probabilities defined in equation (2) can be related to the odds ratio given in equation (8). Taking expectations,

$$\theta_{10} = \mathcal{E}[\theta_{10k}] = \mathcal{E}[\Pr(E_{1k} = 1).\Pr(E_{2k} = 0)]$$
$$= \exp(\delta_1)\, \mathcal{E}\left[\frac{\exp(\beta_k)}{1 + \exp(\beta_k + \delta_1)} \cdot \frac{1}{1 + \exp(\beta_k + \delta_2)}\right]$$

and, similarly,

$$\theta_{01} = \mathcal{E}[\theta_{01k}] = \exp(\delta_2)\, \mathcal{E}\left[\frac{1}{1 + \exp(\beta_k + \delta_1)} \cdot \frac{\exp(\beta_k)}{1 + \exp(\beta_k + \delta_2)}\right].$$

Hence,

$$OR = \exp(\delta_1 - \delta_2) = \theta_{10}/\theta_{01}. \tag{9}$$

For Model 2, no distributional assumptions are made about the β_k. Altham[18] shows that Model 2 is a special case of Model 1 in which the constraint is imposed that $\theta_{11} \geq (\theta_{11} + \theta_{10})(\theta_{11} + \theta_{01})$ or, equivalently,

$$\theta_{11}\theta_{00} \geq \theta_{10}\theta_{01}. \tag{10}$$

For Model 2, we fit Model 1 subject to this constraint and use the posterior distribution of θ_{10}/θ_{01} to estimate the OR from equation (9).

Ghosh *et al.*[19] do not consider misclassified data, but they fit a logistic distribution in a Bayesian hierarchical structure to matched case-control data. Their model assumes the β_k take values from a normal distribution with a mean of zero and an unknown variance. With this model β_k $(k = 1, \ldots, N)$ and δ_i $(1 = 1, 2)$ are estimable. For Model 3 we assume the same structure as Ghosh *et al.*,[19] but allow for recall bias. The odds ratio for this model is estimated from equation (8).

There are a variety of frequentist methods for analyzing matched pairs data. The most common approach is based on the logistic regression structure of Model 2 and substitutes estimates of θ_{10} and θ_{01} into equation (9), yielding the Mantel-Haenszel estimate of the OR.[20]

Mantel and Haenszel[20] also proposed a frequentist estimator derived from Model 1, and they have used it in practice.[21] The frequentist model that is most similar to the hierarchical structure of Model 3 is a mixed-effects logistic regression model, with random effects $\{\beta_k\}$ and $\delta_2 = 0$.[22] The $\{\beta_k\}$ are assigned a distribution and integrated out; the resulting marginal likelihood of δ_1 is then maximized to give an estimate $\widehat{\delta}_1$ of the log odds ratio. However, Neuhaus et al.[23] show that if the resulting estimator of the OR is consistent, then it will be the same as the Mantel-Haenszel estimator. A hierarchical generalised linear model has been used in a frequentist framework by Lee[24], but his approach requires the use of a Poisson-gamma model as an approximation to the logistic model.

The intuitive drawback of the estimate given by Model 2 is that it ignores those pairs for which the case and control have the same exposure [i.e. those pairs for which (E_1, E_2) equals (1, 1) or (0, 0)] and hence sometimes it can be based on only a small portion of the data. However, equations (3) and (9) show that the odds ratios given by Models 1 and 2 are either both above 1.0 (if the estimate of θ_{10} exceeds that of θ_{01}) or both below 1.0 (if the estimate of θ_{01} exceeds that of θ_{10}). Hence Models 1 and 2 give identical probabilities that the OR exceeds 1.0, so Model 2 seems to use the main features of the data, as Model 1 uses all the data to estimate the OR. Moreover, from equations (3) and (9) it follows that the OR from Model 2 is greater than that from Model 1 if and only if

$$\theta_{00}\theta_{11}(\theta_{10} - \theta_{01}) > \theta_{01}\theta_{10}(\theta_{10} - \theta_{01}).$$

When $\theta_{00}\theta_{11} > \theta_{01}\theta_{10}$, so that the condition in (10) holds for Model 1 as well as Model 2, then the OR estimate of Model 2 is more extreme (*i.e.* further from 1.0) than the OR estimate of Model 1, which suggests that it is better to use Model 2 when the assumptions that underlie the logistic model hold. However, if the logistic model is inappropriate then, in general, $\theta_{10}\theta_{01}$ is not an estimator of the odds ratio and equation (3) should be used to estimate it.

A disadvantage of Model 1 is that the estimator it yields makes only limited use of the matching. If the θ_{ij} are estimated by maximum likelihood, then equation (3) yields the same point estimate of the OR as would be obtained if the data were treated as coming from an unmatched case-control study. This does not mean that Model 1 is equivalent to an unmatched case-control study; they generally differ in the standard error they attach to the OR estimate, as the unmatched study assumes $\theta_{00}\theta_{11} = \theta_{01}\theta_{10}$. Model 3 has been proposed only recently

and it is somewhat untested. However, potentially it can exploit information from all pairs (unlike Model 2) and take account of the matched-pairs structure (unlike Model 1). To achieve this, though, it must make stronger assumptions than the other models.

Rothman and Greenland[25] point out that matching in case-control studies can lead to bias and note (page 150) that " . . . if the matching factors are associated with the exposure in the source population, matching in a case-control study requires control by matching factors in the analysis, even if the matching factors are not risk factors for the disease." As an example, suppose the matching factor is perfectly correlated with exposure, so that the exposure level of a control is identical to the case it matches. Then clearly the experiment provides no useful information about the OR. With frequentist methods, the odds ratio will be 0/0 for Model 2, while with Bayesian methods the posterior distribution will be improper if a suitable uninformative prior distribution is used, and the variance of the estimated odds ratio will be infinite. Thus the inadequacy of the data can be recognised when it is analysed as a matched pairs study. In contrast, if matching is ignored then 1.0 will be the estimate of the odds ratio and the standard estimate of this ratio will be finite.

3. THEORY AND IMPLEMENTATION OF THE BAYESIAN MODELS

3.1. Models 1 and 2

For these models, the Bayesian method proposed here examines the data in two stages. In the first stage the data from the validation sub-study is combined with a non-informative prior distribution. The posterior distribution this yields is used as the prior in the second stage, where it is combined with data from the main part of the study. It is assumed that the validation sub-study is a random sample of matched pairs from the whole study.

First stage: validation sub-study

The primary purpose of the validation sub-study is to learn about the relationship between true exposure and apparent exposure. We assume observations are random and independent and, in particular, that misclassification of the exposure state of a case is independent of misclassification of the exposure of its matched control. However, our model allows cases ($D =$

1) and controls ($D = 0$) to have different misclassification rates. Analogous to the definition of the true exposure (E), denote the *apparent* exposure of an individual by A, where $A = 1$ indicates apparently exposed and $A = 0$ indicates apparently not-exposed. For $i = 0, 1$; $l = 0, 1$, define ϕ_{il} by

$$\phi_{il} = \Pr(A = 1 \,|\, E = i, D = l). \tag{11}$$

These are the probabilities of an individual being classified as apparently exposed given their disease and true exposure state.

Let (E_1, E_2, A_1, A_2) denote the true and apparent exposure states of the case and control in a matched pair, where E_1 and A_1 are the true and apparent exposure states for the case, respectively, and E_2 and A_2 are those for the control. Then

$$\begin{aligned}
&\Pr[(E_1, E_2, A_1, A_2) = (i, j, l, m)] \\
&\quad = \Pr[(E_1, E_2) = (i, j)] \;.\; \Pr[(A_1 = l \,|\, E_1 = i; D = 1] \;.\; \Pr[(A_2 = m \,|\, E_2 = j; D = 0] \\
&\quad = \theta_{ij} \;.\; \phi_{i1}^{l}(1 - \phi_{i1})^{1-l} \;.\; \phi_{j0}^{m}(1 - \phi_{j0})^{1-m}.
\end{aligned} \tag{12}$$

It is helpful to tabulate the data from the validation sub-study in two different ways. To summarise information about misclassification rates, the sub-study data are arranged as in Table I, where x_{il} and y_{il} $(i = 0, 1; l = 0, 1)$ are counts in the different cells of the table. The validation sub-study also provides information about the θ_{ij}, which relate exposure and disease. To summarise information about the θ_{ij}, the data may be arranged to give counts of the number of (exposed, exposed), (exposed, not-exposed), (not-exposed, exposed), and (not-exposed, not-exposed) pairs in the validation study. Notation for these counts is given in Table II, where corresponding information about *apparent* exposure in the main study group is also given.

Let $\mathcal{D}$ denote the data from the validation sub-study. Under the assumption of independence between pairs, the likelihood (L) from $\mathcal{D}$ is obtained by multiplying together the probabilities for each matched pair in the sub-study. Each of these probabilities is given by equation (12) so, summing the powers for each term in the likelihood gives

$$L = \theta_{11}^{t_{11}}\, \theta_{10}^{t_{10}}\, \theta_{01}^{t_{01}}\, \theta_{00}^{t_{00}} \prod_{i=0}^{1} \prod_{l=0}^{1} \phi_{il}^{x_{il}}(1 - \phi_{il})^{y_{il}}. \tag{13}$$

Letting $\underline{\theta} = (\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})'$, a prior distribution for the ϕ_{il} and $\underline{\theta}$ must be specified. For Model 1 we suppose that, *a priori*, the ϕ_{il} are independent and identically distributed with

beta distributions, $\phi_{il} \sim \text{beta}(\alpha_0, \alpha_1)$, and that $\underline{\theta}$ is independent of the ϕ_{il} and has a Dirichlet distribution, $\underline{\theta} \sim \text{Dirichlet}(\gamma,\, \gamma,\, \gamma,\, \gamma)$. Thus the prior distribution specifies

$$f(\underline{\theta},\, \phi_{11},\, \phi_{01},\, \phi_{10},\, \phi_{00}) \;\propto\; (\theta_{11}\,\theta_{10}\,\theta_{01}\,\theta_{00})^{\gamma-1} \prod_{i=0}^{1} \prod_{l=0}^{1} \phi_{il}^{\alpha_0-1}(1-\phi_{il})^{\alpha_1-1}. \tag{14}$$

Combining (13) and (14), the posterior distribution is

$$f(\underline{\theta},\, \phi_{11},\, \phi_{01},\, \phi_{10},\, \phi_{00} \,|\, \mathcal{D})$$
$$\propto\; \theta_{11}^{\gamma+t_{11}-1}\, \theta_{10}^{\gamma+t_{10}-1}\, \theta_{01}^{\gamma+t_{01}-1}\, \theta_{00}^{\gamma+t_{00}-1} \prod_{i=0}^{1} \prod_{l=0}^{1} \phi_{il}^{\alpha_0+x_{il}-1}(1-\phi_{il})^{\alpha_1+y_{il}-1}. \tag{15}$$

Thus, in the posterior distribution the ϕ_{il} are mutually independent with

$$\phi_{il} \,|\, \mathcal{D} \sim \text{beta}(\alpha_0 + x_{il}, \alpha_1 + y_{il}) \tag{16}$$

for $i = 0, 1$; $l = 0, 1$ and, independently of the ϕ_{il},

$$\underline{\theta} \,|\, \mathcal{D} \;\sim\; \text{Dirichlet}(\gamma + t_{11}, \gamma + t_{10}, \gamma + t_{01}, \gamma + t_{00}). \tag{17}$$

For Model 2, $\underline{\theta}$ must satisfy the constraint given in (10). Let $I(\underline{\theta})$ be an indicator function that equals 1 for values of $\underline{\theta}$ that satisfy the constraint; 0 otherwise. Analogous to (14), we suppose the prior distribution for Model 2 is given by

$$f(\underline{\theta},\, \phi_{11},\, \phi_{01},\, \phi_{10},\, \phi_{00}) \;\propto\; I(\underline{\theta})\,.\,(\theta_{11}\,\theta_{10}\,\theta_{01}\,\theta_{00})^{\gamma-1} \prod_{i=0}^{1} \prod_{l=0}^{1} \phi_{il}^{\alpha_0-1}(1-\phi_{il})^{\alpha_1-1}.$$

Then the posterior distribution for the ϕ_{il} is again given by equation (16) and the posterior distribution of $\underline{\theta}$ is given by

$$f(\underline{\theta} \,|\, \mathcal{D}) \;\propto\; I(\underline{\theta})\,.\,\theta_{11}^{\gamma+t_{11}-1}\, \theta_{10}^{\gamma+t_{10}-1}\, \theta_{01}^{\gamma+t_{01}-1}\, \theta_{00}^{\gamma+t_{00}-1}. \tag{18}$$

In the practical application of Bayesian statistics, uninformative prior distributions are commonly used. For an uninformative prior distribution, we suggest setting $\alpha_0 = \alpha_1 = \gamma = 0$. This form of prior distribution is improper, but it has good invariance properties[26] and the posterior distribution will be proper unless one of the counts x_{il}, y_{il} or t_{il} is zero. The writers' view is that the use of an uninformative prior distribution for the model in (12) is inappropriate if one of the x_{il}, y_{il} or t_{il} is zero. (If $x_{il} = 0$, for example, then the posterior variance of ϕ_{il} is $\alpha_0(\alpha_1 + y_{il})/\{(\alpha_0 + \alpha_1 + y_{il} + 1)(\alpha_0 + \alpha_1 + y_{il})^2\}$, which is highly sensitive to the value chosen for α_0 in the prior distribution, so α_0 should not just be given an arbitrary value.) However, in the example in Section 5 we also consider setting $\alpha_0 = \alpha_1 = \gamma = 1$ as a vague prior, which

gives a prior distribution that is flat and proper. For the example, the posterior distribution is insensitive as to whether α_0, α_1 and γ are set equal to 0 or 1.

Second stage: main study

The second stage of the method for Models 1 and 2 analyzes the data from the main study group, for which only the apparent exposures are known. Analogous to equation (2), we define $p_{lm} = \Pr[(A_1,\, A_2) = (l, m)]$ for $l = 0, 1;\, m = 0, 1$. Then,

$$\begin{aligned} p_{lm} &= \sum_{i=0}^{1}\sum_{j=0}^{1} \theta_{ij} \,.\, \Pr[(A_1 = l \,|\, E_1 = i; D = 1)] \,.\, \Pr[(A_2 = m \,|\, E_2 = j; D = 0)] \\ &= \sum_{i=0}^{1}\sum_{j=0}^{1} \theta_{ij} \,.\, \phi_{i1}^{l}(1-\phi_{i1})^{1-l} \,.\, \phi_{j0}^{m}(1-\phi_{j0})^{1-m} \end{aligned} \tag{19}$$

for $l = 0, 1;\, m = 0, 1$.

The posterior distribution from the first stage forms the prior distribution for the second stage, which transfers all relevant information between the two stages. Thus, the prior distribution of $(\underline{\theta},\, \phi_{11},\, \phi_{01},\, \phi_{10},\, \phi_{00})$ for the second stage is given by equations (16) and (17) for Model 1 and equations (16) and (18) for Model 2. These determine the prior distribution of p_{lm} $(l = 0, 1;\, m = 0, 1)$ as the p_{lm} are functions of the ϕ_{il} $(i = 0, 1;\, j = 0, 1)$ and $\underline{\theta}$. The relevant sampling model for the second stage is

$$(T_{11},\, T_{10},\, T_{01},\, T_{00}) \sim \text{multinomial}(n_2;\, p_{11},\, p_{10},\, p_{01},\, p_{00}). \tag{20}$$

The posterior distribution that results from the second stage is analytically intractable, but samples from the posterior distribution are readily obtained using MCMC methods, such as that implemented in WinBUGS. A separate chain is run for each model. The simplest way of implementing Model 2 is to specify the same prior distribution as for Model 1 and add the constraint given in (10), which is straightforward in WinBUGS. Each iteration of a chain gives an observation of $(\theta_{11},\, \theta_{10},\, \theta_{01},\, \theta_{00})$ from which an estimate of the odds ratio is calculated using equation (3) for Model 1 or (9) for Model 2. This gives a sequence of observations from the posterior distribution of the odds ratio for each model, which may be used for estimation and inference.

3.2. Model 3

For this model each pair must be considered separately. Also, data from both the validation substudy and the main study must be considered simultaneously. This is because the information

gained from the sub-study cannot be expressed as a tractable distribution, so there is no simple way of transferring information from the sub-study to the analysis of the main study. However, the likelihood has a structure similar to the likelihood in equation (13), in that it factorises into one factor for each ϕ_{il} and further factors for the other parameters. Hence, the part of the model that relates to misclassification rates remains unchanged, with information about the ϕ_{il} still given by equation (16).

The logistic model assumes conditional independence of exposure status within a pair, given β_k, so the distribution for the cases and controls may be specified separately. From equations (5) and (6), the sampling model for the n_1 cases and controls in the validation sub-study is

$$\Pr[E_{ik} = j] = \frac{\{\exp(\beta_k + \delta_i)\}^j}{1 + \exp(\beta_k + \delta_i)} \tag{21}$$

for $i = 1, 2$; $j = 0, 1$; $k = 1, \ldots, n_1$. For pairs in the main study, only apparent exposure is observed, which we denote by A_{1k} for the case in the kth pair and A_{2k} for the control. Let

$$\begin{aligned} \eta_{lk} &= \sum_{i=0}^{1} \Pr(E_{1k} = i) \,.\, \Pr(A_{1k} = l \,|\, E_{1k} = i) \\ &= \sum_{i=0}^{1} \Pr(E_{1k} = i) \,.\, \phi_{i1}^{l}(1 - \phi_{i1})^{1-l} \end{aligned} \tag{22}$$

and

$$\begin{aligned} \psi_{mk} &= \sum_{j=0}^{1} \Pr(E_{2k} = j) \,.\, \Pr(A_{2k} = m \,|\, E_{2k} = j) \\ &= \sum_{j=0}^{1} \Pr(E_{2k} = j) \,.\, \phi_{j0}^{m}(1 - \phi_{j0})^{1-m} \end{aligned} \tag{23}$$

for $l = 0, 1$; $m = 0, 1$; $k = n_1 + 1, \ldots, N$. Then the sampling model for the main study is

$$\Pr(A_{1k} = l) = \eta_{lk} \tag{24}$$

for cases and

$$Pr(A_{2k} = m) = \psi_{mk} \tag{25}$$

for controls ($l = 0, 1$; $m = 0, 1$; $k = n_1{+}1, \ldots, N$). Equations (21)–(25) determine the likelihood for the main study.

Prior distributions for the ϕ_{il}, δ_1, δ_2 and $\beta_1, \ldots, \beta_N$ must be specified. The part of the model that relates to misclassification is the same as Models 1 and 2 so, from equation (14),

the prior distribution for the ϕ_{il} is taken as

$$f(\phi_{11},\, \phi_{01},\, \phi_{10},\, \phi_{00}) \propto \prod_{i=0}^{1} \prod_{l=0}^{1} \phi_{il}^{\alpha_0 - 1}(1 - \phi_{il})^{\alpha_1 - 1}. \tag{26}$$

Following Ghosh *et al.*,[19] diffuse prior distributions are used for δ_1 and δ_2:

$$\delta_i \sim \mathrm{N}(0.0, 10^5) \quad \text{for } i = 1,\, 2. \tag{27}$$

Again following Ghosh *et al.*,[19], the β_k are given the hierarchical prior structure:

$$\beta_k \sim \mathrm{N}(0, \tau^{-1}) \quad \text{for } k = 1, \ldots, N \tag{28}$$

and

$$\tau \sim \mathrm{gamma}(a, b). \tag{29}$$

The number of parameters in Model 3 grows in direct proportion to the sample size, giving a likelihood that is not well-behaved; for example, the maximum likelihood estimator of $\delta_1 - \delta_2$ is inconsistent.[27] A consequence is that choosing an uninformative prior distribution for τ is difficult. An equivalent way of expressing (29) is to say that $2b\tau$ has a χ^2 distribution on $2a$ degrees of freedom. Hence, the value of a and b should be small if the prior distribution is to convey little information and a common choice (the standard choice in examples in WinBUGS[17]) is to set $a = b = 0.001$. However, the example in Section 5 illustrates that the estimate of the odds ratio and, in particular, the posterior precision of τ are sensitive to the values of a and b, even when these are both small. We defer discussion of the choice of a and b to Section 5.

As with Models 1 and 2, the posterior distribution cannot be determined but MCMC methods can be used to generate a sequence of observations from the posterior distribution. The odds ratio is $\exp(\delta_1 - \delta_2)$ (equation (8)) and this is calculated at each iteration.

4. THE 1:M MATCHED CASE-CONTROL MODEL

In some matched case-control studies, more than one control is matched with each case. In this section, suppose each case is matched with M controls. Extending earlier definitions, let E_1 and E_2 denote the number of cases and controls in a matched set who have been truly exposed, so E_1 equals 0 or 1 while E_2 equals 0, 1 ... or M. Also, let A_1 and A_2 denote the corresponding numbers for apparent exposure. Table III gives notation for the data, where t_{ij} is the number

of matched sets in the validation subgroup for which $(E_1,\, E_2) = (i,\, j)$ and T_{ij} is the number of matched sets in the main study group for which $(A_1,\, A_2) = (i,\, j)$, $[i = 0, 1;\, j = 0, 1, \ldots, M]$. Notation for data on misclassification is unchanged and the data are summarised in Table I. (As there are now M controls in each matched set, $x_{10} + x_{00} + y_{10} + y_{00} = MN$.)

4.1. Model 1

Let

$$\theta_{ij} = \Pr[(E_1,\, E_2) = (i,\, j)] \quad \text{for } i = 0, 1;\, j = 0, 1, \ldots, M.$$

Analogous to equation (3), the odds ratio for Model 1 is

$$\text{OR} \;=\; \frac{\sum_{j=0}^{M} \theta_{1j}}{\sum_{j=0}^{M} \theta_{0j}} \,.\, \frac{\sum_{j=0}^{M} (M-j)(\theta_{0j} + \theta_{1j})}{\sum_{j=0}^{M} j(\theta_{0j} + \theta_{1j})}. \tag{30}$$

The counts from the validation study follow a multinomial distribution,

$$(t_{1M}, \ldots, t_{10}, t_{0M}, \ldots, t_{00}) \sim \text{multinomial}(n_1;\, \theta_{1M}, \ldots, \theta_{10}, \theta_{0M}, \ldots, \theta_{00}) \tag{31}$$

and an uninformative (improper) prior distribution for the θ_{ij} is assumed,

$$f(\theta_{1M}, \ldots, \theta_{10}, \theta_{0M}, \ldots, \theta_{00}) \propto \prod_{i=0}^{1} \prod_{j=0}^{M} \theta_{ij}^{-1}.$$

Given the data from the validation sub-study, $\mathcal{D}^*$ say, the posterior distribution for the θ_{ij} is

$$\theta_{1M}, \ldots, \theta_{10}, \theta_{0M}, \ldots, \theta_{00} \,|\, \mathcal{D}^* \sim \text{Dirichlet}(t_{1M}, \ldots, t_{10}, t_{0M}, \ldots, t_{00}), \tag{32}$$

which is proper provided $t_{ij} \geq 1$ for $i = 0, 1;\, j = 0, \ldots, M$. The analysis of misclassification rates is essentially unchanged. As before, we assume the ϕ_{il} are independent and identically distributed in the prior distribution, with $\phi_{il} \sim \text{beta}(\alpha_0, \alpha_1)$ for $i = 0, 1;\, l = 0, 1$. Then (16) gives the posterior distribution of the ϕ_{il} and they are mutually independent and independent of the θ_{ij}.

For the second stage of the analysis, extending equation (19) we put

$$p_{lm} \;=\; \sum_{i=0}^{1} \sum_{j=0}^{M} \theta_{ij} \;.\; \Pr[(A_1 = l \,|\, E_1 = i; D = 1)] \;.\; \Pr[(A_2 = m \,|\, E_2 = j; D = 0)]$$

$$= \sum_{i=0}^{1} \sum_{j=0}^{M} \sum_{h=0}^{j} \theta_{ij} \;.\; \phi_{i1}^{l} (1 - \phi_{i1})^{1-l} \;.\; \binom{j}{h} \phi_{10}^{h} (1 - \phi_{10})^{j-h} \binom{M-j}{m-h} \phi_{00}^{m-h} (1 - \phi_{00})^{M-j-m+h}, \tag{33}$$

for $l = 0, 1$; $m = 0, 1, \ldots, M$, where

$$\binom{M-j}{m-h} = 0$$

for $m - h < 0$ or $M - j < m - h$. The prior distribution for the second stage is given by the product of the distributions in equations (16) and (32), and the sampling model is

$$(T_{1M}, \ldots, T_{10}, T_{0M}, \ldots, T_{00}) \sim \text{multinomial}(N - n_1;\, p_{1M}, \ldots, p_{10}, p_{0M}, \ldots, p_{00}). \quad (34)$$

Samples from the posterior distribution are obtained using MCMC methods. The values of the θ_{ij} $(i = 0, 1;\, j = 0, \ldots, M)$ and the ϕ_{il} $(i = 0, 1;\, l = 0, 1)$ are stored at each iteration of the chain and, from them, estimates of the odds ratio at each iteration are determined using equation (30).

4.2. Model 2

For the logistic model, individual matched sets must be identified. For the kth set (*c.f.* equation (4)), define

$$\theta_{ijk} = \Pr[(E_{1k},\, E_{2k}) = (i,\, j)] \quad \text{for } i = 0, 1;\, j = 0, 1, \ldots, M;\, k = 1, \ldots, N.$$

Extending the notation in equation (6), let $\exp(\beta_k+\delta_2)/(1+\exp(\beta_k+\delta_2))$ denote the probability that *any* specified control in the kth set is exposed. Then, assuming conditional independence (given β_k),

$$\begin{aligned}\theta_{ijk} &= \frac{\{\exp(\beta_k + \delta_1)\}^i}{1 + \exp(\beta_k + \delta_1)} \cdot \frac{M!}{j!\,(M-j)!} \cdot \frac{\{\exp(\beta_k + \delta_2)\}^j}{\{1 + \exp(\beta_k + \delta_2)\}^M} \\ &= \exp(i\delta_1 + j\delta_2)\, \frac{M!}{j!\,(M-j)!}\, \Lambda \exp\{(i+j)\beta_k\} \end{aligned} \quad (35)$$

for $i = 0, 1$; $j = 0, 1, \ldots, M$; $k = 1, \ldots, N$, where $\Lambda = [\{1+\exp(\beta_k+\delta_1)\}\{1+\exp(\beta_k+\delta_2)\}^M]^{-1}$. Letting $\mathcal{E}$ denote expectation over β_k,

$$\theta_{ij} = \mathcal{E}[\theta_{ijk}] = \exp(i\delta_1 + j\delta_2) \cdot \frac{M!}{j!\,(M-j)!} \cdot \mathcal{E}\left[\Lambda \exp\{(i+j)\beta_k\}\right] \quad (36)$$

for $i = 0, 1$; $j = 0, 1, \ldots, M$.

The standard frequentist estimate of the OR for 1:M matched studies is the Mantel-Haenszel estimate, which is obtained by replacing the θ_{ij} by their frequentist estimates[28] in the equation,

$$OR = \frac{\sum_{j=0}^{M} (M-j)\theta_{1j}}{\sum_{j=0}^{M} j\theta_{0j}}. \quad (37)$$

For Model 2, we estimate the odds ratio from equation (37) with the θ_{ij} defined in equation (36). Now, from equation (36),

$$\exp(\delta_1 - \delta_2) = \frac{(M-j)\theta_{1j}}{(j+1)\theta_{0(j+1)}} \quad \text{for } j = 0, \ldots, M-1. \tag{38}$$

Also, $(M-j)\theta_{1j} = 0$ for $j = M$ and $j\theta_{0j} = 0$ for $j = 0$. Hence the odds ratio for Model 2 will satisfy $OR = \exp(\delta_1 - \delta_2)$, as it should.

The θ_{ij} must satisfy various constraints if the logistic model holds. From equation (38), we may put

$$\lambda = \frac{M\theta_{10}}{\theta_{01}} = \ldots = \frac{(M-j)\theta_{1j}}{(j+1)\theta_{0(j+1)}} = \ldots = \frac{\theta_{1(M-1)}}{M\theta_{0M}}. \tag{39}$$

Note that λ is the estimator of the odds ratio given by equation (37). Also, from Schwartz' inequality,

$$\mathcal{E}\left[\Lambda \exp(j\beta_k)\right] . \mathcal{E}\left[\Lambda \exp\{(j+2)\beta_k\}\right] \geq \left(\mathcal{E}\left[\Lambda \exp\{(j+1)\beta_k\}\right]\right)^2,$$

so equation (36) implies that $\theta_{0j}\theta_{1(j+1)} \geq \theta_{0(j+1)}\theta_{1j}$ for $j = 0, \ldots, M-1$. Hence, for the logistic model the θ_{ij} must also satisfy the constraints,

$$\frac{\theta_{00}}{\theta_{10}} \geq \ldots \geq \frac{\theta_{0j}}{\theta_{1j}} \geq \ldots \geq \frac{\theta_{0M}}{\theta_{1M}}. \tag{40}$$

The equalities in equation (39) enable Model 2 to be reparameterised with fewer parameters. Letting $\zeta_j = \theta_{0j} + j\lambda\theta_{0j}/(M-j+1)$ for $j = 0, 1, \ldots, M$ and $\zeta_{M+1} = \theta_{1M}$ gives

$$\theta_{0j} = \frac{(M-j+1)}{M-j+1+j\lambda}\zeta_j \qquad \text{for } j = 0, 1, \ldots, M. \tag{41}$$

and

$$\theta_{1(j-1)} = \frac{j\lambda}{M-j+1+j\lambda}\zeta_j \qquad \text{for } j = 1, \ldots, M+1. \tag{42}$$

Also, we have that $\sum_{j=0}^{M+1}\zeta_j = \sum_{i=0}^{1}\sum_{j=0}^{M}\theta_{ij} = 1$. Let $\underline{\zeta} = (\zeta_0, \ldots, \zeta_{M+1})$. From (41) and (42), Model 2 can be parametrized in terms of λ and $\underline{\zeta}$ and straightforward algebra shows that the constraints

$$\frac{(M-j+1+j\lambda)^2}{\{M-j+2+(j-1)\lambda\}\{M-j+(j+1)\lambda\}} \cdot \frac{j+1}{j} \cdot \frac{M-j+2}{M-j+1} \geq \frac{\zeta_j^2}{\zeta_{j-1}\zeta_{j+1}} \tag{43}$$

for $j = 1, \ldots, M$ are equivalent to the constraints given in (40).

A disadvantage of these constraints is that, for some values of $\underline{\zeta}$, they restrict the values that λ can take to the union of two disjoint intervals. This seems undesirable, especially as

the intervals would not include $\lambda = 1$, which is *a priori* a plausible value of λ, as λ is the odds ratio. Rather than (43), we impose the constraint

$$\frac{j+1}{j} \cdot \frac{M-j+2}{M-j+1} \geq \frac{\zeta_j^2}{\zeta_{j-1}\zeta_{j+1}}, \tag{44}$$

for $j = 1, \ldots, M$. This is more restrictive on the values which $\underline{\zeta}$ can take than the constraint in (43), since

$$(M-j+1+j\lambda)^2 \,/\, [\{M-j+2+(j-1)\lambda\}\{M-j+(j+1)\lambda\}] \geq 1.$$

When $\underline{\zeta}$ satisfies (44), the values that λ can take are unrestricted.

In the first stage of the analysis the data are from the validation sub-study. From equation (31), the likelihood (L^*) is

$$L^* \;=\; \zeta_0^{t_{00}} \zeta_{M+1}^{t_{1M}} \prod_{j=1}^{M} \left[\frac{M-j+1}{M-j+1+j\lambda}\,\zeta_j\right]^{t_{0j}} \left[\frac{j\lambda}{M-j+1+j\lambda}\,\zeta_j\right]^{t_{1(j-1)}} \tag{45}$$

Let $J(\underline{\zeta})$ be an indicator function that equals 1 when $\underline{\zeta}$ satisfies the constraint in (44); $J(\underline{\zeta}) = 0$ otherwise. We assume an improper uninformative prior distribution,

$$f(\lambda, \underline{\zeta}) \;\propto J(\underline{\zeta})\, \lambda^{-1} \prod_{j=0}^{M+1} \zeta_j^{-1}, \tag{46}$$

which is $J(\underline{\zeta})$ multiplied by the standard diffuse prior distribution for positive-valued parameters. Combining (45) and (46), in the posterior distribution,

$$f(\underline{\zeta} \,|\, \mathcal{D}^*) \;\propto\; J(\underline{\zeta})\, \zeta_0^{t_{00}-1} \left[\prod_{j=1}^{M} \zeta_j^{t_{0j}+t_{1(j-1)}-1}\right] \zeta_{M+1}^{t_{1M}-1}, \tag{47}$$

which is $J(\underline{\zeta})$ multiplied by the kernel of a Dirichlet distribution. Independently of $\underline{\zeta}$,

$$f(\lambda \,|\, \mathcal{D}^*) \;\propto\; \lambda^{-1} \prod_{j=1}^{M} \left[\frac{M-j+1}{M-j+1+j\lambda}\right]^{t_{0j}} \left[\frac{j\lambda}{M-j+1+j\lambda}\right]^{t_{1(j-1)}}. \tag{48}$$

In the second stage of the analysis, the data are the apparent exposures in the main study. The prior marginal distribution for the misclassification rates (ϕ_{il}) are given in equation (16), as for Model 1, and the prior marginal distributions for $\underline{\zeta}$ and λ are given by (47) and (48). The product of these marginal distributions forms the prior distribution. To obtain the likelihood, the p_{lm} in equation (33) are expressed in terms of $\underline{\zeta}$ and λ by substituting for the θ_{ij} using equations (41) and (42). Then the likelihood is obtained from (34). Combining the likelihood and prior distribution is again intractable analytically but a random sample can be

obtained from the posterior distribution using MCMC methods. While running the MCMC sampler, the constraints in (44) are imposed. At each iteration of the chain, an estimate of the OR is determined from equations (37), (41) and (42).

Implementation of the above method using software such as WinBUGS is a little tricky because the distribution in (48) is not a standard form. The method we use to handle this problem is outlined in Appendix 1.

4.3. Model 3

The analysis for Model 3 is unchanged. There is a single β_k for each matched set ($k = 1, \ldots, N$) and, conditional on the β_k, the data for each individual are independent. The likelihood is given by equations (21)–(25), the prior distribution is given by (26)–(29), and the odds ratio by equation (8).

5. EXAMPLES AND COMPARISONS

In this section, use of the methods proposed here is illustrated by applying them to data about smoking and myocardial infarct (MI). The data set has not previously been examined in relation to recall bias, but it has been reported elsewhere.[29] The original study looked at MI and smoking in 103 cases and 309 controls matched on date of birth (within 6 months). A control had to be under observation when the case was diagnosed with the infarct, have no history of MI at this time and be recruited by a different GP to the case. Owen-Smith *et al.*[29] compared the effects of smoking at recruitment (in 1968), from doctors' records, to smoking at follow-up (in 1995). Subjects' recall of whether they smoked at the time of recruitment was ascertained from questioning them at follow-up. Therefore we have full information on smoking according to both the doctor's record and the patient's recall for all 412 patients. The doctor's record is treated as a gold standard measure and we suppose a patient's recall is potentially misclassified.

The data were first used to obtain a 1:1 matched-pairs study by taking the case and the first of the three controls available in the original data. To create a data set in which recall bias is relevant, we randomly selected 50 of the 103 case-control sets and assigned them to be the validation sub-study, and assigned the rest to be the main study. We ignored the doctors' records of smoking for the 53 sets in the main study as if these data were unknown.

The resulting data set is summarised in Tables IV and V, and shows a modest amount of misclassification, with a greater degree of misclassification for the controls than for the cases. Table V indicates that matching has not been very effective as there is little association between the exposure states within a pair.

The three models described in Section 3 were fitted to the data using the WinBUGS software for MCMC.[17] The priors used in the models were varied in order to assess the sensitivity of the posterior distribution to different choices of 'uninformative' prior distribution. For Models 1 and 2, the uninformative prior distributions that were used are $\alpha_0 = \alpha_1 = \gamma = 0$, which gives improper prior distributions, and $\alpha_0 = \alpha_1 = \gamma = 1$, which gives prior distributions that are flat and proper. For Model 3, the same choice of prior distributions were used for the ϕ_{il}, but parameters for the prior distribution of τ must still be specified. The latter prior distribution is assumed to have the form, $\tau \sim \text{gamma}(a, b)$, and in the first instance we put $a = 0.5$ and $b = 0.5$, which is reasonably uninformative. For Model 1, a burn-in phase of 1000 iterations was clearly adequate for convergence and a further 10,000 iterations were run. For Model 2, 20,000 iterations were run after a burn in of 10,000. For Model 3, convergence was sluggish and 10,000 iterations were used as a burn-in phase, followed by a further 50,000 iterations. For each model the mean of the log odds ratio (LOR) was determined, together with a 95% credible interval for the LOR.

For comparison with the Bayesian methods, the Mantel-Haenszel (M-H) maximum likelihood method proposed by Greenland[5] was also applied to the data, using the same selected 50 pairs as an external validation study. All 103 pairs were included in the main study as Greenland's method re-uses the validation data in the analysis of the main study. Let $\underline{p} = (p_{11}, p_{01}, p_{10}, p_{00})'$ and let Φ be a 4×4 matrix whose elements are the ϕ_{il} arranged so that $\underline{p} = \Phi\underline{\theta}$ is equivalent to equation (19). Greenland[5] refers to Φ as a classification matrix; let $\mathcal{C}$ denote its maximum likelihood estimate given by the validation sub-study data. Let $\mathcal{T} = (T^*_{11}, T^*_{01}, T^*_{10}, T^*_{00})'$, where T^*_{ij} is the combined total number of pairs in the sub-study and main study for which $(A_1, A_2) = (i, j)$. Also, let $\underline{t} = N\underline{\theta}$; N is the number of pairs in the whole study. Then, assuming the classification matrix is invertible (which it should be for a reasonable classification method), an estimate of $\underline{t}$ is $\widehat{\underline{t}} = \mathcal{C}^{-1}\mathcal{T}$. An asymptotic first-order estimator of $\text{var}(\widehat{\underline{t}})$ was derived by Greenland. The formula for the matrix estimator appears simple, but contains complex functions of the true classification rates and the covari-

ances among the estimated classification rates. If the uncertainty in the estimated classification rates is negligible then the asymptotic estimator reduces to a simple function of the classification matrix and the estimated variance-covariance of the disease-apparent exposure cell-counts, $\text{var}\,(\widehat{\underline{t}}) = \mathcal{C}^{-1}(\text{var}(\mathcal{T}))\mathcal{C}^{-1}$, asymptotically. The log odds ratio is estimated using equation (9) and its variance is readily obtained from $\text{var}\,(\widehat{\underline{t}})$.

The M-H method was applied using matrix multiplication in S-Plus, and the maximum likelihood of the LOR and a 95% confidence interval for it were calculated. The M-H maximum likelihood method was designed to be used with an external validation study and the version developed by Greenland[5] which we have used for the confidence intervals does not allow for any uncertainty in the misclassification probabilities from the validation study. Allowing for this uncertainty would greatly increase the complexity of the method.

Results are presented in Table VI. The first two rows of the table relate to analysis for the situation described above, with a sub-study of 50 pairs and a main study of the remaining 53 pairs. The first row is for the improper prior ($\alpha_0 = \alpha_1 = \gamma = 0$) and its results are similar to those given in the second row for the flat prior ($\alpha_0 = \alpha_1 = \gamma = 1$), suggesting that the precise form of the uninformative prior distribution for the ϕ_{il} and $\underline{\theta}$ is of little importance if there are adequate sample data. Methods were also applied to (a) the actual data, so that gold-standard values were available for all pairs; and (b) to the data from patient's recall, but treating these data as if they were not subject to misclassification. Results for these two data sets are given in the bottom two rows of Table VI.

The table indicates that the choice of model had limited effect on the point estimate of the LOR. When the prior for τ with Model 3 is $\tau \sim$ gamma(0.5, 0.5), that model gave estimates higher than Model 1, which in turn gave estimates higher than Model 2, but the observed differences between models are small. The estimates from Model 2 and M-H maximum likelihood are both based just on those pairs in which the case smoked but the control did not, or vice-versa. Hence, differences between these estimates result largely from the use the methods make of the gold-standard measurements. Model 2 uses them to help estimate the LOR, while the maximum likelihood method uses them only to estimate misclassification rates and uses the apparent exposures from both the validation sub-study and the main study to estimate the OR. To incorporate the gold-standard measurements into the estimation of the LOR within the M-H maximum likelihood method would require substantial revision of the

method previously presented by Greenland. The last two rows of Table VI show that the gold-standard measurements gave a slightly higher estimate of the LOR than the measurements derived from the recall of subjects. The widths of credible intervals were also similar for all models, which is perhaps surprising, since stronger assumptions often lead to narrower credible intervals, and Model 3 makes stronger assumptions than Model 2, which in turn makes stronger assumptions than Model 1. The "true" LOR, obtained with full knowledge of the true exposures, was always comfortably within both the credible intervals from the Bayesian analyses and the confidence intervals from maximum likelihood.

A sensitivity analysis was conducted for Model 3 to examine the impact of varying the prior distribution of τ. This prior distribution is gamma(a, b) and values (a, b)= (0.001, 0.001), (0.1, 0.1), (0.5, 0.5), (1.0, 1.0), (0.5, 2.0), and (1.5, 5) were examined. Results are presented in Table VII for both the cases $\alpha_0 = \alpha_1 = \gamma = 0$ and $\alpha_0 = \alpha_1 = \gamma = 1$. The table shows that the estimate of the Log OR and its credible interval are little affected by the values of a and b, although there is a detectable difference between $\tau \sim$ gamma(0.001, 0.001) and $\tau \sim$ gamma(0.1, 0.1). This difference is perhaps surprising, as these gamma distributions both ostensibly convey little information - they are equivalent to χ^2 distributions on 0.002 and 0.2 degrees of freedom, respectively.

The choice of a and b has a major impact on the posterior estimate of τ, with the posterior mean varying from 0.77 to 157.7. The width of the credible interval for τ is also alarmingly large when a and b are very small. These results show that the sample data provide very little information about the value of τ, otherwise the data would swamp the small amount of information contained in the prior distribution. Consequently, in implementing Model 3 it seems sensible to use an informative prior distribution for τ that reflects background knowledge and to examine the sensitivity of estimates to a range of reasonable choices for this prior distribution. Ghosh *et al.*[19] chose $\tau \sim$ gamma(2.5, 1.5) and $\tau \sim$ gamma(1.5, 5), describing the latter as a diffuse distribution. Like us, they found the choice of prior distribution for τ had a marked effect on its posterior distribution and little effect on the estimate of Log OR, which is the quantity of most interest.

The results in Table VII also examine the sensitivity of estimates to the values of α_0, α_1 and γ, the parameters of the prior distribution of $(\theta, \phi_{11}, \phi_{01}, \phi_{10}, \phi_{00})$. Whether their values were set equal to 0 or 1 seldom had much effect on the posterior estimate of the Log

OR or any parameters. An exception arose when the prior distribution for τ specified $\tau \sim$ gamma(0.001, 0.001), when the posterior distribution of τ is very volatile and was affected substantially by changes in α_0, α_1 and γ. In summary, the sensitivity analysis suggests that which prior distribution is chosen for τ is important, while which diffuse distribution is taken as an uninformative prior distribution for $(\theta, \phi_{11}, \phi_{01}, \phi_{10}, \phi_{00})$ generally matters little.

Bayesian and frequentist methods were also applied to the full data set, in which there are *three* controls for each case. Fifty sets were again randomly selected as the validation group and the remaining 53 sets formed the main study group. The resulting data are summarized in Table VIII. The three Bayesian models were fitted using the methods described in Section 4 and in the prior distribution we set $\alpha_0 = \alpha_1 = 0$ as a non-informative prior. For Model 3, τ was given the prior distribution $\tau \sim$ gamma(0.5, 0.5). Burn-in phases of 10,000 iterations were used for each model, followed by a further 20,000 iterations for Model 1 and a further 50,000 iterations for each of Models 2 and 3. The M-H maximum likelihood estimates of the LOR were also calculated, again using the method given by Greenland[5]. As well as the actual data, the methods were also applied to the full set of gold-standard data, and to the data based on patients' recall, ignoring the possibility of misclassification with the latter. Results are given in Table IX.

The table shows that Models 1 and 3 and M-H maximum likelihood gave similar log odds ratios to each other and Model 2 gave a slightly higher Log OR, but differences could be due to random variation in the data. There are many ways of partitioning the 103 matched sets into a validation sub-study of 50 sets and a main study of 53 sets, and simulations were repeated for a further three partitions. For these other partitions, Model 2 gave Log OR estimates of 1.244, 1.116, and 1.366, which are noticeably smaller than the value of 1.515 that was found with the first partition. The greater number of controls result in credible intervals that are about 30% narrower than in Table VI, where only one control was used for each case. Again, the widths of credible intervals are very similar for the three models and comfortably contain the estimates of LOR that were obtained when gold-standard information on all individuals was used.

6. NON-DIFFERENTIAL MISCLASSIFICATION

For some data sets it is appropriate to assume misclassification is non-differential, so that cases and controls have identical misclassification rates. The methods developed in Sections 3 and 4

address the slightly more complicated situation where misclassification is differential, but they are readily adapted to handle the non-differential case, as follows.

With non-differential misclassification, $\phi_{i0} = \phi_{i1}$ for each i (i=0,1). Hence, the definitions given by equation (11) are replaced by

$$\phi_i = \Pr(A = 1 \,|\, E = i, D = 1) = \Pr(A = 1 \,|\, E = i, D = 0)$$

for $i = 0, 1$. In addition, equation (16) is replaced by

$$\phi_i \,|\, \mathcal{D} \sim \text{beta}(\alpha_0 + x_{i0} + x_{i1},\, \alpha_1 + y_{i0} + y_{i1})$$

for $i = 0, 1$, and (26) is replaced by

$$f(\phi_0,\, \phi_1) \,\propto\, \prod_{i=0}^{1} \phi_i^{\alpha_0 - 1}(1 - \phi_i)^{\alpha_1 - 1}.$$

Also, we replace ϕ_{i0} and ϕ_{i1} by ϕ_i $(i = 0, 1)$ in equations (19), (22), (23) and (33). The methods are otherwise unchanged.

7. CONCLUDING COMMENTS

Three different Bayesian models have been described for analyzing matched case-control studies. The models differ in the assumptions they make and which of them is the more appropriate to use in practice will depend upon characteristics of the data being analyzed. When there is a single control for each case, then the condition given in equation (10) should be compared with the data. If this condition does not seem to hold, it suggests that the logistic model does not fit the data and that Model 1 should be preferred to Models 2 and 3. Similarly, when there is more than one control for each case, then the conditions given in equations (39) and (40) should appear to hold if a logistic model fits the data; otherwise Model 1 is again to be preferred. In contrast, when the logistic model seems to fit the data well, then either (or both) Model 2 and Model 3 should be used. Model 3 requires an informative prior distribution for τ to be specified and can be sensitive to the distribution that is chosen. Hence, Model 3 should only be used if prior knowledge enables a realistic prior distribution for τ to be specified, or if a set of plausible prior distributions can be specified and inferences about the odds ratio are insensitive to which distribution from the set is chosen. Otherwise, Model 2 is to be preferred to Model 3. Model 2 also has the advantage that it is the model that is

most similar to the standard frequentist model, while avoiding the computational difficulties in handling misclassification that arise with the more sophisticated frequentist methods.

The approaches described in this paper could be applied to more complex situations, such as those in which there are covariates. In principle, a perfectly measured binary or categorical covariate could easily be accommodated in Models 1 or 2. For a binary covariate, this would double the number of cell combination counts required for the summary validation and main study data. The complexity of the model would not greatly increase. The limiting factor on the number of covariates that could be included would be ensuring sufficient observations in each combination so that the estimates of misclassification rates were not based on very small or zero cells. If the cell counts become very small, misclassification for certain disease and covariate combinations would not occur in the validation study and so would only be poorly addressed in the main study. Model 3 has the potential to be extended to more complex situations than the other two models because it considers each individual separately, rather than using summary statistics. The introduction of both categorical and continuous covariates should be reasonably straightforward, but further work is needed in this area.

The methods given in this paper can all be implemented using WinBUGS software and programs that implement the methods for matched pairs are given at http://mcs-notes1.open.ac.uk:8080/repository/StatsPublications1.nsf/$$ViewTemplate%20for%20Supporting%20Material?OpenForm, where programs used in Section 5 for three controls per case are also available. Where data values are needed in the programs, those given in Tables V and VIII are used. For the data set in Table V, the computer run-times for Model 1, Model 2 and Model 3 were 27 seconds, 81 seconds and 638 seconds, respectively, while for the data in Table VIII they were 557 seconds, 1231 seconds and 1300 seconds. These times are not prohibitively long, and the examples in Section 5 illustrate that the proposed methods are viable ways of analyzing matched case-control data.

ACKNOWLEDGEMENTS

We are grateful to two referees for constructive comments that improved this paper. We must also thank Phil Hannaford for allowing the use of data from the Royal College of General Practitioners' Oral Contraceptive Study. The original postal questionnaire was funded by an unconditional grant from Wyeth-Ayerst International Inc. Additional support for the Oral

Contraceptive Study was received from the Royal College of General Practitioners, Schering AG, and Schering Health Care.

Appendix 1

Implementation of Model 2 when there are multiple controls

In Winbugs, using $f(\lambda \,|\, \mathcal{D}^*)$ in equation (48) as a prior distribution for stage 2 of the analysis is difficult because $f(\lambda \,|\, \mathcal{D}^*)$ is a non-standard distribution. For $j = 1, \ldots, M$ define

$$\xi_j \sim \text{bin}[t_{0j} + t_{1(j-1)},\, (M - j + 1)/(M - j + 1 + j\lambda)]$$

and suppose that the observed value of ξ_j is t_{0j}. If $\mathcal{D}^{\#}$ is the data set $\{\xi_1, \ldots, \xi_M\}$, then $f(\lambda \,|\, \mathcal{D}^*) = f(\lambda \,|\, \mathcal{D}^{\#})$. Hence, $f(\lambda \,|\, \mathcal{D}^*)$ is equivalent to the combination of the prior distribution $f(\lambda) \propto \lambda^{-1}$ and the (hypothetical) data $\mathcal{D}^{\#}$. However this cannot itself be implemented in WinBUGS because improper prior distributions such as $f(\lambda) \propto \lambda^{-1}$ are not allowed.

Instead, pick a value of j, say j^*, for which t_{0j} and $t_{1(j-1)}$ are both positive (we actually choose j^* so that the minimum of t_{0j} and $t_{1(j-1)}$ is as large as posible.). Define w by

$$w = \frac{M - j^* + 1}{M - j^* + 1 + j^*\lambda}$$

and suppose w has a beta distribution, $w \sim \text{beta}(c_0, c_1)$. Then $f(\lambda) \propto \lambda^{-1}$ if $c_0 \to 0$ and $c_1 \to 0$. As $c_0 \to 0$ or $c_1 \to 0$, the distribution of w becomes improper but the distribution of $w \,|\, \xi_j^* = t_{0j^*}$ is proper and has the standard form

$$(w \,|\, \xi_j^* = t_{0j*}) \sim \text{beta}(t_{0j*},\, t_{1(j*-1)}).$$

Let $\mathcal{D}^{\#}_{\setminus j^*}$ denote the data $\mathcal{D}^{\#}$ with information about ξ_{j^*} deleted. Then, for WinBUGS, we specify that w has a prior distribution beta(t_{0j^*}, $t_{1(j^*-1)}$ and add the hypothetical data $\mathcal{D}^{\#}_{\setminus j^*}$ to the real data from the second stage of the analysis. This is equivalent to using $f(\lambda \,|\, \mathcal{D}^*)$ (with no hypothetical data) as the prior distribution for the second stage.

REFERENCES

1. Diamond E. L. and Lillenfeld A. M. 'Effects of errors in classification and diagnosis in various types of epidemiological studies', *American Journal of Public Health,* **52**, 1137–1144 (1962).

2. Reade-Christoper, S. J. and Kupper, L. L. 'Effects of exposure misclassification on regression analyses of epidemiologic follow-up study data', *Biometrics,* **47**, 535–548 (1991).
3. Greenland, S. 'The effect of misclassification in matched-pair case-control studies', *American Journal of Epidemiology,* **116**, 402–406 (1982).
4. Greenland, S. and Kleinbaum, D. G. 'Correcting for misclassification in two-way tables and matched-pair studies', *International Journal of Epidemiology,* **12**, 93–97 (1983).
5. Greenland, S. 'On correcting for misclassification in twin studies and other matched pair studies', *Statistics in Medicine,* **8**, 825–829 (1989).
6. Dellaportas, P. and Stephens, D. A. 'Bayesian analysis of errors-in-variables regression models', *Biometrics,* **51**, 1085–1095 (1995).
7. Prescott, G. J. and Garthwaite, P. H. 'A simple analysis of misclassified binary data with a validation sub-study', *Biometrics,* **58**, 454–458 (2002).
8. Muller, P. and Roeder, K. A. 'Bayesian semiparametric model for case-control studies with errors in variables', *Biometrika,* **84**, 523-537 (1997).
9. Kuha, J. 'Estimation by data augmentation in regression models with continuous and discrete covariates measured with error', *Statistics in Medicine,* **16**, 189–201 (1997).
10. Richardson, S. and Gilks, W. R. 'A Bayesian approach to measurement error problems in epidemiology using conditional-independence models', *American Journal of Epidemiology,* **138**, 430–442 (1993).
11. Richardson, S. and Gilks, W. R. 'Conditional-independence models for epidemiologic studies with covariate measurement error', *Statistics in Medicine,* **12**, 1703–1722 (1993).
12. Richardson, S. and Leblond, L. 'Some comments on misspecification of priors in Bayesian modelling of measurement error problems', *Statistics in Medicine,* **16**, 203–213 (1997).
13. Schmid, C. H. and Rosner, B. 'A Bayesian-approach to logistic regression models measurement error following a mixture distribution', *Statistics in Medicine,* **12**, 1141–1153 (1993).
14. York, J., Madigan, D., Heuch, I. and Lie, R. T. 'Birth-defects registered by double sampling - a Bayesian-approach incorporating covariates and model uncertainty', *Applied Statistics,* **44**, 227–242 (1995).
15. Thurigen, D., Spiegelman, D., Blettner, M., Heuer, C. and Brenner, H. 'Measurement error correction using validation data: a review of methods and their applicability in case-control studies', *Statistical Methods in Medical Research,* **9**, 447–474 (2000).

16. Armstrong B. G., Whittemore, A. S. and Howe, G. R. 'Analysis of case-control data with covariate measurement error: application to diet and colon cancer', *Statistics in Medicine,* **8**, 1151–1163 (1989).
17. Spiegelhalter, D. J., Thomas, A., and Best, N. G. *WinBUGS Version 1.2.,* Technical report, University of Cambridge: MRC Biostatistics Unit, 1999.
18. Altham, P. M. E. 'The analysis of matched proportions', *Biometrika,* **58**, 561–576 (1971).
19. Ghosh, M., Chen, M.-H., Ghosh, A. and Agresti, A. 'Hierarchical Bayesian analysis of binary matched pairs data', *Statistica Sinica,* **10**, 647–657 (2000).
20. Mantel, N. and Haenszel, W. 'Statistical aspects of the analysis of data from retrospective studies of disease', *Journal of the National Cancer Institute,* **22**, 719–748 (1959).
21. Haenszel, W., Shimkin, M. B. and Mantel, N. 'A retrospective study of lung cancer in women', *Journal of the National Cancer Institute,* **21**, 825–842 (1958).
22. Pierce, D. 'A random effects model for matched pairs of binary data', Technical Report 55, Department of Statistics, Oregon State University (1975).
23. Neuhaus, J. M., Kalbfleish, J. D. and Hauck, W. W. 'Conditions for consistent estimation in mixed-effects models for binary matched-pairs data', *Canadian Journal of Statistics,* **22**, 139–148 (1994).
24. Lee, Y. 'Can we recover information from concordant pairs in binary matched pairs?', *Journal of Applied Statistics,* **28**, 239–246 (1971).
25. Rothman, K. J. and Greenland, S. *Modern epidemiology.* Philadelphia: Lippincott-Raven, 1998.
26. Seaman, S. R. and Richardson, S. 'Bayesian analysis of case-control studies with categorical covariates', *Biometrika,* **88**, 1073–1088 (2001).
27. Andersen, E. B. 'Asymptotic properties of conditional maximum likelihood estimators', *Journal of the Royal Statistical Society, Series B,* **32**, 283–301 (1970).
28. Breslow, N. E. and Day, N. E. *Statistical methods in cancer research. Volume 1: The analysis of case=control studies.* Lyon: International Agency for Research in Cancer, 1980; pp 169–171.
29. Owen-Smith, V., Hannaford, P., Warskyj, M., Ferry, S. and Kay, C. R. 'Effects of changes in smoking status on risk estimates for myocardial infarction among women recruited for the Royal College of General Practitioners' Oral Contraceptive Study in the UK', *Journal of Epidemiology and Community Health,* **52**, 420–424 (1998).

Table I. Counts for a Matched Pairs Study with Misclassified Exposure Data

Apparent exposure	Validation study group			
	Cases ($D=1$)		Controls ($D=0$)	
(A)	$E=1$	$E=0$	$E=1$	$E=0$
1	x_{11}	x_{01}	x_{10}	x_{00}
0	y_{11}	y_{01}	y_{10}	y_{00}

Table II. Counts for a Matched Pairs Study with Misclassified Exposure Data

	Validation study group			Main study group		
	True	Controls ($D=0$)		Apparent	Controls ($D=0$)	
	exposure	$E_2=1$	$E_2=0$	exposure	$A_2=1$	$A_2=0$
Cases	$E_1=1$	t_{11}	t_{10}	$A_1=1$	T_{11}	T_{10}
(D=1)	$E_1=0$	t_{01}	t_{00}	$A_1=0$	T_{01}	T_{00}

Table III. Counts for a 1:M Matched Study with Misclassified Exposure Data

	Validation study group					Main study group				
	True	Controls ($D=0$)				Apparent	Controls ($D=0$)			
	exposure	$E_2=M$	...	$E_2=1$	$E_2=0$	exposure	$A_2=M$	...	$A_2=1$	$A_2=0$
Cases	$E_1=1$	t_{1M}	...	t_{11}	t_{10}	$A_1=1$	T_{1M}	...	T_{11}	T_{10}
(D=1)	$E_1=0$	t_{0M}	...	t_{01}	t_{00}	$A_1=0$	T_{0M}	...	T_{01}	T_{00}

Table IV. Counts for OCS 1:1 Matched Study with Misclassified Smoking Exposure

Apparent exposure	Validation study group			
	Cases ($D = 1$)		Controls ($D = 0$)	
(A)	$E = 1$	$E = 0$	$E = 1$	$E = 0$
1	27	1	14	4
0	2	20	3	29

Table V. Counts for OCS 1:1 Matched Study with Misclassified Smoking Exposure

	Validation study group			Main study group		
	True exposure	Controls ($D = 0$)		Apparent exposure	Controls ($D = 0$)	
		$E_2 = 1$	$E_2 = 0$		$A_2 = 1$	$A_2 = 0$
Cases (D=1)	$E_1 = 1$	9	20	$A_1 = 1$	12	26
	$E_1 = 0$	8	13	$A_1 = 0$	5	10

Table VI. LOR Estimates and 95% Credible/Confidence Intervals for the OCS Matched Pairs Example

Prior/ Data Set	Model 1 11 000 iterations		Model 2 30 000 iterations		Model 3[a] 60 000 iterations		M-H Maximum Likelihood	
	Log OR	cred. int.	Log OR	cred. int.	Log OR	cred. int.	Log OR	conf. int.
B(0, 0)[b]	1.383	0.678–2.102	1.477	0.750–2.256	1.499	0.769–2.276	1.272[f]	0.506–2.038
B(1, 1)[c]	1.278	0.592–1.996	1.369	0.651–2.129	1.481	0.733–2.278		
Gold Stan.[d]	1.289	0.876–1.702	1.357	0.930–1.798	1.438	0.862–2.076	1.242[g]	0.625–1.859
Recall[e]	1.247	0.823–1.678	1.299	0.887–1.726	1.363	0.706–1.978	1.121[g]	0.538–1.703

[a] Prior distribution used for the precision: $\tau \sim$ gamma(0.5,0.5).
[b] $\alpha_0 = \alpha_1 = \gamma = 0$ in the prior; 50 pairs in the validation sub-study and 53 pairs in the main study.
[c] $\alpha_0 = \alpha_1 = \gamma = 1$ in the prior; 50 pairs in the validation sub-study and 53 pairs in the main study.
[d] Using 103 gold standard values and $\alpha_0 = \alpha_1 = \gamma = 0$.
[e] Treating apparent values as true values and $\alpha_0 = \alpha_1 = \gamma = 0$.
[f] Calculated using ML with fixed differential misclassification estimates taken from the validation study.
[g] From traditional Mantel-Haenszel formula with no misclassification.

Table VII. Sensitivity analysis for Model 3: LOR Estimates and 95% Credible/Confidence Intervals for the OCS Matched Pairs Example

Prior	$\alpha_0 = \alpha_1 = \gamma = 0$				$\alpha_0 = \alpha_1 = \gamma = 1$			
for τ[a]	Log OR	cred. int.	τ	cred. int.	Log OR	cred. int.	τ	conf. int.
(0.001, 0.001)	1.410	0.722–2.139	76.07	0.97–487.0	1.383	0.665–2.126	157.7	1.21–1058
(0.1, 0.1)	1.458	0.747–2.202	5.56	0.74–20.1	1.441	0.704–2.211	5.35	0.62–19.8
(0.5, 0.5)	1.499	0.769–2.276	2.70	0.61–7.35	1.481	0.733–2.278	2.74	0.55–7.86
(1.0, 1.0)	1.526	0.767–2.313	2.00	0.55–4.90	1.510	0.750–2.312	2.01	0.52–5.11
(0.5, 2.0)	1.622	0.836–2.433	1.13	0.37–2.62	1.581	0.793–2.434	1.17	0.37–2.71
(1.5, 5.0)	1.685	0.888–2.521	0.79	0.34–1.54	1.655	0.828–2.533	0.77	0.31–1.52

[a] Parameters of gamma distribution. *e.g.* the first row is gamma(0.001, 0, 001).

Table VIII. Counts for OCS 1:3 Matched Study with Misclassified Smoking Exposure

	Validation study group					Main study group				
	True	Controls ($D = 0$)				Apparent	Controls ($D = 0$)			
	exposure	$E_2 = 3$	$E_2 = 2$	$E_2 = 1$	$E_2 = 0$	exposure	$A_2 = 3$	$A_2 = 2$	$A_2 = 1$	$A_2 = 0$
Cases	$E_1 = 1$	1	8	9	11	$A_1 = 1$	1	8	13	16
(D=1)	$E_1 = 0$	1	5	11	4	$A_1 = 0$	1	2	7	5

Table IX. LOR Estimates and 95% Credible/Confidence Intervals for the OCS 1:3 Matched Sets Example

Data Set	Model 1 30 000 iterations		Model 2 60 000 iterations		Model 3[a] 60 000 iterations		M-H Maximum Likelihood	
	Log OR	cred. int.	Log OR	cred. int.	Log OR	cred. int.	Log OR	conf. int.
Actual data	1.343	0.794–1.909	1.515	0.898–2.162	1.427	0.867–1.989	1.398[b]	0.807–1.988
Gold Stan.[c]	1.302	0.963–1.655	1.259	0.982–1.637	1.381	0.898–1.874	1.231[d]	0.761–1.701
Recall[e]	1.394	1.051–1.735	1.385	1.021–1.756	1.505	1.008–2.009	1.351[d]	0.873–1.828

[a] Prior distribution used for the precision: $\tau \sim$ gamma(0.5,0.5).
[b] Calculated using ML with fixed differential misclassification estimates taken from the validation study.
[c] Using 103 gold standard values.
[d] From traditional Mantel-Haenszel formula with no misclassification.
[e] Treating apparent values as true values.