

Birth Defects Registered by Double Sampling: A Bayesian Approach Incorporating Covariates and Model Uncertainty

Jeremy York, Department of Statistics, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213 USA

David Madigan¹, Department of Statistics GN-22,
University of Washington, Seattle, Washington 98195 USA

Ivar Heuch

Department of Mathematics, University of Bergen,
N-5007 Bergen, Norway

and Rolv Terje Lie

Medical Birth Registry of Norway, University of Bergen,
Armauer Hansen Building, Haukeland Hospital, N-5021 Bergen, Norway

June 27, 1994

¹Correspondence author

Abstract

In double sampling schemes, a large sample is classified using one method, and a subsample is also classified with a supplementary method. In the application discussed here, we are attempting to identify infants in Norway born with Down's syndrome using both a national birth registry and a regional registry. Usual methods for analysing such data assume that one classification method is perfect, which is not the case here. We develop a Bayesian approach that allows for error in both registries, includes covariates (here, the age of the mother) and explicitly accounts for our lack of knowledge about the complexity of the relationships between the variables considered. Markov chain Monte Carlo methods are used to approximate the posterior. In the data considered here, the error rates of the two registries appear to be substantial. Despite a strong relationship between maternal age and risk of Down's syndrome, inclusion of the maternal age covariate does not substantially change the overall estimates.

Keywords : Double sampling; Markov chain Monte Carlo; Down's syndrome; model uncertainty; Bayesian inference

1 Introduction

1.1 What is Double Sampling?

Suppose that we wish to estimate the proportion of infants born with Down’s syndrome nationwide. For every birth during a certain time period, the midwife or obstetrician has noted whether or not the child has Down’s syndrome. This classification is based on a visual inspection of each child and may miss some of the true cases. For a small subsample of births, results of more accurate but expensive cytogenetic tests are also available.

Estimation of the required proportion using only the cytogenetic data would give an estimate which is unaffected by measurement error but has a large variance because of the small sample size. Use of only the visual data would provide an estimate that is biased but less variable. Neither estimate utilises information about the accuracy of the visual test provided by a cross-classification of the two data sources. A preferable approach would use *all* of the data to estimate both the desired proportion and the accuracy of the visual test. This whole scenario exemplifies “double sampling”.

A straightforward maximum likelihood approach to double sampling was presented by Tenenbein (1970, 1972). Extensive generalisations of Tenenbein’s work have been reported by Chen (1979, 1989), Hochberg (1977), Ekholm and Palmgren (1987), Ekholm (1991), Espeland and Hui (1987), Espeland and Odoroff (1985) and others. A Bayesian approach was presented by Geng and Asano (1989). Lie *et al.* (1994) and Nedelman (1988) addressed the statistical problems arising when the results reported from the secondary test are themselves subject to error.

Here we present an approach to double sampling which is based on Bayesian graphical models. This allows us to account for model uncertainty, incorporate informative expert opinion and handle the case where no infallible classifier is available.

1.2 Down's Syndrome Proportions in Norway

We will use a Bayesian graphical model approach to estimate Down's syndrome proportions among newborns in Norway. Since 1970, epidemiological surveillance of congenital malformations has been carried out in Norway on the basis of data in the nationwide Medical Birth Registry (MBR). These data are collected at birth by the midwife or obstetrician and correspond to the visual inspection in the simplified description above. In principle, any pregnancy with at least 16 weeks of gestation should be reported, regardless of whether any birth defect is seen or not, including pregnancies which are terminated for any reason, as well as stillbirths.

It is well-known that the reporting to the MBR is far from perfect for certain birth defects, although little has been known about the extent of underreporting. Nationwide reporting to the MBR is required by law. For some birth defects considerably more detailed information may be available at an individual level, but accumulation of information in a nationwide registry is only allowed for summarized data such as those reported to the MBR. Detailed results from prenatal tests, for example, cannot be included directly in these registries, although such results could form the basis of some of the information supplied to the registries.

Because of growing concerns about incomplete ascertainment in the nationwide registration, a new notification system entitled "Melding om Fosterindiserte Aborter og Medfødte Misdannelser" (MIA) was introduced in 1985 in the county of Hordaland. The MIA registration was based on prenatal diagnostics and pediatric follow-up including results from cytogenetic tests and was itself subject to error. The MIA registry was not designed to be superior to the MBR but was rather based on an entirely different reporting routine. In contrast to the MBR, a particular child was reported to the MIA registry only if a birth defect was observed. The target population within the county of Hordaland was the same as for the MBR, namely all births occurring inside the county. This represents about 15% of all births in Norway. Hordaland, situated on the West Coast, is a fairly typical Norwegian county, and there is no prior reason to believe that reporting patterns to the MBR should

Table 1: about here

differ essentially from those in other parts of the country.

The purpose of the secondary registry MIA was to provide a basis for an evaluation of the main registry MBR, both with regard to unbiased prevalence estimates of birth defects as well as ascertainment probabilities. Potential ascertainment bias can involve failure to detect abnormalities as well as failure to report a detected case. In this work we only examine Down's syndrome; the overall data for 1985–1988 are given in Table 1, where R_1 represents a case ascertained through the MBR, R_2 refers to MIA, and the vinculum denotes negation. Further details were given by Lie *et al.* (1991,1994).

Since neither registry is infallible, traditional double sampling approaches cannot be used. Lie *et al.* (1994) describe a frequentist approach for analysing such data, assuming uniformity of reporting practices over the entire country, and they fit two different models to data from these registries. In the first (due to Nedelman, 1988), it is hypothesised that ascertainment is independent in the two registration systems. In the second model, a strong negative association is hypothesised between the registration systems, namely that all doubly sampled cases (i.e., the cases of Down's syndrome in Hordaland) are ascertained by at least one system. Under both models it is assumed that no false positives were recorded. This assumption is quite reasonable in this context, as those responsible for the registration would be very cautious not to indicate a defect of this kind on the registration forms without proper justification.

In this paper we address two difficulties which arise in the analysis of Lie *et al.* (1994). Firstly, although both of their models provide a reasonable fit to the data, estimates of Down's syndrome prevalence at birth and the corresponding asymptotic standard errors are quite different under the two models. The Bayesian framework used here explicitly accounts for this model uncertainty. Secondly, Lie *et al.* (1994) did not consider any covariates such as maternal age in their analysis. Because of the strong association between Down's syndrome and maternal age, as reviewed by Cuckle *et al.* (1987), a complete study of this defect should include age as a covariate. The complexity of the calculations in Lie *et al.* (1994), in

particular the determination of asymptotic variances, suggests that such extensions would be difficult. The Bayesian graphical model framework greatly facilitates the incorporation of covariates.

1.3 Bayesian Graphical Models: Basic Ideas

Here we sketch the elements of the approach; a more complete description is provided in the next section.

Let Δ denote a quantity of interest about which we wish to make inference, and let Y denote the data. Adopting a Bayesian approach, our objective is to derive a posterior distribution for Δ , $\text{pr}(\Delta | Y)$, combining all the available sources of information about Δ , i.e. prior knowledge and the data. Typically, such a posterior distribution is conditioned upon a single “best” model chosen via some model selection procedure. However, any approach (Bayesian or not) which chooses a single model ignores a major component of uncertainty, namely uncertainty about the model itself (Draper, 1994, Madigan and Raftery, 1994). As a consequence, uncertainty about quantities of interest can be underestimated. Striking examples of this are given in Regal and Hook (1991) and Draper (1994). Indeed, this problem is widely acknowledged so it is somewhat surprising that *practical* methodologies for accounting for model uncertainty are only now emerging.

A Bayesian solution to this problem accounts for model uncertainty by averaging over *all* the models in the class being considered, to give:

$$\text{pr}(\Delta | Y) = \sum_{k=1}^K \text{pr}(\Delta | M_k, Y) \text{pr}(M_k | Y). \quad (1)$$

In equation (1), M_1, M_2, \dots, M_K are the models considered, the posterior probability of model M_k is given by

$$\text{pr}(M_k | Y) = \frac{\text{pr}(Y | M_k) \text{pr}(M_k)}{\sum_{l=1}^K \text{pr}(Y | M_l) \text{pr}(M_l)},$$

where

$$\text{pr}(Y | M_k) = \int \text{pr}(Y | \theta, M_k) \text{pr}(\theta | M_k) d\theta,$$

θ is a vector of parameters, $\text{pr}(\theta \mid M_k)$ is the prior for θ under model M_k , $\text{pr}(Y \mid \theta, M_k)$ is the likelihood, and $\text{pr}(M_k)$ is the prior probability that M_k is the true model.

Madigan and Raftery (1994) demonstrate that averaging over *all* the models in this fashion provides better predictive ability, as measured by a logarithmic scoring rule, than using any single model M_j .

The models we consider are the directed graphical models (sometimes called recursive causal models or Bayes networks). They are fully defined by a set of conditional independence relationships which can be represented by a directed acyclic graph. In the graph, nodes represent random variables and directed links (from ‘parent nodes’ to ‘child nodes’) represent direct influences, and each node is independent of all of its non-descendants given its parent nodes. More detailed accounts are given by Whittaker (1990), Madigan and York (1993), and York (1992b).

To fully specify the model, the probability of each child node given all configurations of its parent nodes must be specified. One strength of the graphical model approach is that these parameters are easy to interpret, so that informative prior distributions are easily elicited. The graph itself is a powerful model representation medium: model assumptions become entirely transparent.

Figure 1: about here

As an example, a simple graphical model for the Norwegian Down’s syndrome data is shown in Figure 1. Here S is a binary random variable representing the true Down’s syndrome state (which is never actually observed), while R_1 and R_2 represent the classifications from the MBR and MIA registries, respectively. The graph embodies the assumption that R_1 and R_2 are conditionally independent given S .

2 Bayesian Graphical Models for Double Sampling

In this section we outline some of the technical details of the Bayesian graphical model approach to double sampling.

The orientations of the edges in the models we consider are constrained in order to rule out nonsense models. For instance, presence of Down’s syndrome (S) influences the registries (R_1 and R_2), not vice versa, and so edges between these variables should point from S to the registries. The likelihood of the data conditional on a model M is parameterised by a vector θ which gives the probability of each node conditioned on each of the states of the node’s parents. For the example of Figure 1 we have:

$$\theta = \{\text{pr}(R_1 | S), \text{pr}(R_1 | \bar{S}), \text{pr}(R_2 | S), \text{pr}(R_2 | \bar{S}), \text{pr}(S)\}.$$

We assume *a priori* that the components of θ are independent. An informative conjugate prior distribution for θ can be found by eliciting beta distributions for each of its components in turn. The variables considered here are binary; if they had multiple categories, we would use Dirichlet priors on the components of θ . Since we assume that there are no false positives, $\text{pr}(R_1 | \bar{S})$ and $\text{pr}(R_2 | \bar{S})$ are fixed at zero.

The calculations would be much easier if, in addition to the observed data Y , we had hypothetical data Z which provided the true value of S for all children not identified by the registries as having Down’s syndrome. Under the no false positive assumption, S is known for all cases identified by the registries. The posterior $\text{pr}(\Delta | Y, Z, M_k)$ could be easily calculated (Spiegelhalter and Lauritzen, 1990), as could the posterior model probability $\text{pr}(M_k | D)$ since the prior is conjugate. For more general double sampling problems, Z would include latent variables and missing values for various observed variables.

The fact that only Y is observed complicates the calculations considerably. Exact calculation of the expressions in (1) requires summing over all possible values of the unobserved data, which is not feasible. Numerical methods are required to evaluate functionals of the posterior.

A variety of computational methods are presented in Madigan and York (1993); see also Kass and Raftery (1994). Here, we adopt a version of the Gibbs sampler (Smith and Roberts, 1993; York, 1992a; Besag *et al.*, 1991), in which we simulate both the unobserved data and the correct model.

The posterior of the quantity of interest can be re-expressed as follows :

$$\text{pr}(\Delta | Y) = \mathbf{E}(\text{pr}(\Delta | M, Y, Z) | Y),$$

where the expectation is with respect to $\text{pr}(Z, M | Y)$. This posterior can be numerically approximated by simulating a process $\{Z(t), M(t)\}, t = 1, 2, \dots$, with stationary distribution $\text{pr}(Z, M | Y)$. An empirical average of any function of the process converges to its expectation under the stationary distribution :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \text{pr}(\Delta | Y, Z(t), M(t)) \rightarrow \text{pr}(\Delta | Y) \quad a.s. \quad (2)$$

for any fixed value of Δ ; see Wei and Tanner (1990).

There are many ways to implement a Gibbs sampler approach for simulating the process $\{Z(t), M(t)\}$. We do so here by also including θ in the simulation scheme. At each iteration, we first draw $M(t)$ from the conditional distribution

$$\text{pr}(M|Y, Z(t-1)) \propto \text{pr}(M) \int \text{pr}(Y, Z(t-1)|M, \theta) \text{pr}(\theta|M) d\theta; \quad (3)$$

the choice of a conjugate prior for θ makes this integration straightforward. Next, we draw $\theta(t)$ from $\text{pr}(\theta|Y, Z(t-1), M)$, which amounts to simulating values from independent beta distributions. These two steps together provide us with values of M and θ that are drawn jointly from $\text{pr}(M, \theta|Y, Z(t-1))$. Finally, we draw from the distribution $\text{pr}(Z|Y, M(t), \theta(t))$ to obtain $Z(t)$. See York (1992b) for technical details. This Gibbs sampling approach to missing or unobservable data is a variation of the IP algorithm (Tanner and Wong 1987).

We use the simulated values of Z and M to calculate an estimate of the posterior of Δ as in (2); expectations of functions of Δ can be readily calculated from this approximate posterior. Since the quantity Δ is a function of the model and its parameters, we can also use the simulated values of $\theta(t)$ and $M(t)$ to estimate the expectation of Δ :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \Delta(\theta(t), M(t)) \rightarrow \mathbf{E}(\Delta | Y) \quad (4)$$

by the same argument as above. In the independent sampling case, Gelfand and Smith (1991) have noted that this method will have a higher Monte Carlo variance than would

be obtained by calculating the expectation of the approximate posterior (2). Thus, we rely upon the approximation (2) whenever possible, and only use (4) when $\text{pr}(\Delta | Y, Z(t), M(t))$ does not take a convenient form (such as the posterior of the error rate of the MIA registry under some models). Posterior model probabilities can be estimated by taking empirical averages of the quantities calculated in (3) above.

3 Application to Down’s Syndrome

In this section, we present results of Bayesian analyses of the Norwegian Down’s syndrome data, both with and without the mother’s age as a covariate. We have selected particular priors based on the information available to us; the philosophy of this approach is that no prior can be truly non-informative, and that analysts and decision makers are not completely ignorant. We do not claim that our priors are “correct”; we *do* claim that they are reasonable, and have verified that reasonable variations upon them do not cause substantial changes in the inference.

In each analysis (with and without the age covariate), 100,000 iterations of a Gibbs sampler over M , θ , and Z were run to obtain estimates of the posterior model probabilities and features of the full posterior. Features of individual models were determined by running subsequent Gibbs samplers for 20,000 iterations each with M fixed at the model of interest.

The data were extracted from a dynamic database, in which records are updated and corrected as more information become available. Since the data we present here were extracted at a later date than the simpler data of Lie *et al.* (1994), which were not tabulated by maternal age, there are slight differences between these two data sets. A parallel analysis was conducted for data that were adjusted so that the margins agreed with the data presented in Lie *et al.* (1994). This second analysis gave results that were scarcely different from those presented herein; thus, our results are still directly comparable to those given previously.

3.1 Analysis without Age Covariate

We begin by considering the data in Table 1, which do not include information on the age of the mothers, in order to compare our results to those in Lie *et al.* (1994).

Graphical representations of the models, and features of the posteriors, are given in Table 2. Model I is that of independence between registries conditional upon whether Down’s syndrome is present; this was considered in Nedelman (1988). Model II assumes that, for the doubly sampled data, the MIA registry R_2 will find all cases missed by the national registry R_1 ; this special kind of dependence is denoted graphically by $R_1 \xrightarrow{*} R_2$. Lie *et al.* (1994) proposed this model. Model II is particularly relevant if the personnel working with the presumably superior diagnostic procedures underlying MIA are aware of previous notifications to the MBR and perhaps consider additional reporting unnecessary if a particular birth defect has already been reported. Note that this model does not assume that registry R_2 is infallible; there can be cases which were reported under R_1 which will be missed by R_2 . In Model III the registries are again dependent, but, unlike Model II, the dependence is not constrained and could be positive or negative. Lastly, features of the overall posterior (1) are given.

Table 2: about here

Each of the three models was assumed to be equally likely *a priori*. The parameters for the prior on $\text{pr}(S)$, the prevalence of Down’s syndrome, were based upon raw data from the MBR registry for 1979–1984 without any adjustment for misclassification (Lie *et al.*, 1991). During that period, there were 0.97 observed cases per 1,000 births; the $\text{beta}(0.0097, 9.9903)$ prior was chosen to have that rate as its expected value and so that our prior knowledge is equivalent to having observed 10 hypothetical births. Choosing the prior to represent 100 hypothetical cases does not have a substantial impact upon the results. The priors for the probability of a case being identified by the registries reflect our belief that the registries are more likely to find cases of the syndrome than not. For the national MBR registry, we place a $\text{beta}(4,2)$ prior on $\text{pr}(R_1 | S)$. In Model I the probability of a case being correctly

identified in the MIA registry, $\text{pr}(R_2 | S)$, is also $\text{beta}(4,2)$. In Model III, the probabilities $\text{pr}(R_2 | S, R_1)$ and $\text{pr}(R_2 | S, \bar{R}_1)$ are each given $\text{beta}(2.22, 1.11)$ distributions, so that the mean and variance of the marginal quantity $\text{pr}(R_2 | S)$ is the same as in Model I. For Model II, a symmetric $\text{beta}(1.5, 1.5)$ prior for $\text{pr}(R_2 | S, R_1)$ was used; the precision was chosen for reasons to be described in the next section. Of course, $\text{pr}(R_2 | S, \bar{R}_1) = 1$ in Model II. Thus priors for both Models II and III are balanced between the possibilities of positive and negative association between the two registries.

The posterior distribution for the prevalence of Down's syndrome averaged over models is given in Figure 2. Posteriors for each individual model are also shown, scaled so that the area under the posterior for model M_k is $\text{pr}(M_k | Y)$, the posterior probability of that model; thus, the overall posterior is the sum of all the curves for individual models.

Figure 2: about here

The frequentist analysis of Lie *et al.* (1994) has been reproduced for the version of the data given here; maximum likelihood estimates are given in Table 3 for comparison. The difference in the estimates between Models I and II for the Bayesian approach is less extreme than under the frequentist approach, but is still considerable. Since the data do not obviously favour one model over another, the model averaging approach provides a reasonable compromise between these disparate estimates, with a larger posterior variance than would be obtained by just choosing the model which predicts the data the best.

Table 3: about here

3.2 Analysis with Age Covariate

We now expand the class of models to include those where age of the mother has an impact upon S , R_1 , and/or R_2 . We disallow any model where maternal age affects the registries but not the presence of Down's syndrome. This gives us five ways to choose a relationship between maternal age and the other variables; together with the three assumptions about

the relationship between the two registries, this gives us 15 models. Once again, we assume that they are all equally likely *a priori*. The age of the mother was extracted from the registry databases and tabulated in a categorical manner, and is given in Table 4. The 6 age categories (A_1, \dots, A_6) represent mothers of ages ≤ 24 , 25—29, 30—34, 35—39, 40—44, and ≥ 45 .

Table 4: about here

In order to compare models that do and do not consider an interaction between age of mother and other variables, we choose the priors so that they roughly agree between the models. If maternal age does not affect the prevalence of Down’s syndrome, the prior for $\text{pr}(S)$ is given as in the previous section. If the prevalence does vary with age, we will have 6 parameters $\text{pr}(S | A_i)$ instead of the single $\text{pr}(S)$. In that case, we choose prior distributions for for the six parameters so that $\text{Var}(\text{pr}(S))$ is the same for all models. Specifically, we choose the variances for $\text{pr}(S | A_i)$ to be identical for all i , subject to the constraint that

$$\text{Var}(\text{pr}(S)) = \sum_{i=1}^6 \pi_i^2 \text{Var}(\text{pr}(S | A_i)), \quad (5)$$

where π_i is the proportion of the births with mother in age class A_i . We choose a prior for each $\text{pr}(S | A_i)$ with expectation given by historical data for 1979–1984, which are 0.59, 0.59, 0.97, 3.04, 6.88, and 18.50 cases per 1000 for age groups A_1, \dots, A_6 , as presented in Lie *et al.* (1991). For models in which an ascertainment probability for the registries varies by age, a single parameter such as $\text{pr}(R_1 | S)$ is replaced by six age-specific parameters, $\text{pr}(R_1 | S, A_i)$. These are chosen to be independent and identically distributed (for different values of i) so that the priors agree with that of the corresponding smaller model in the same manner as (5). The reason for choosing the beta(1.5, 1.5) prior for $\text{pr}(R_2 | S, R_1)$ in Model II in the previous section is that it is impossible to decompose this parameter by age and keep the same overall variance if the variance of the original quantity is too large. The variance of the beta(1.5, 1.5) distribution is just small enough to ensure that the variance of $\text{pr}(R_2 | S, R_1)$ is the same across all relevant models.

Table 5: about here

Summaries of the five best models, and the results for model averaging, are given in Table 5. Note that the model averaging is over all 15 models. The five models shown in Table 5 account for 98% of the total posterior probability. Modelling prevalence as a function of age substantially improves predictive ability — the models of the previous subsection each have posterior model probabilities of less than 10^{-12} in this analysis. The results indicate the possibility of an interaction between mothers age and the ascertainment probability of the MIA registry; however, the individual ascertainment probabilities do not change in any systematic way across the age categories. The posterior for the prevalence rate for the entire population along with the posteriors for the models with highest posterior probability (scaled according to posterior probabilities) are given in Figure 3. The 5th percentile, mode, and 95th percentile for the prevalence in each age category are given in Figure 4.

Figure 3: about here

Figure 4: about here

3.3 Sensitivity

There are two potential sources of instability in these results : variability due to Monte Carlo approximation methods, and sensitivity to the choice of the prior parameters. Informal explorations of Monte Carlo variability leave us confident that means, modes, and standard deviations presented here are accurate up to two significant figures. The estimates of the posterior model probabilities are somewhat more variable, and can vary in the second significant figure from simulation to simulation.

Estimates of the prevalence rate are relatively insensitive to reasonable changes in the prior, such as adjusting the sum of the parameters for one or more of the beta priors (estimates range from 1.63 to 1.71 per 1,000 for the prevalence under a variety of such priors)

or using a prior for $\text{pr}(S)$ that is not based on historical data (using a $\text{beta}(1,9)$ prior for the prevalence rate can give a slightly higher estimate, in the range of 1.7 to 1.8 per 1,000 depending upon the other priors). Estimates of the error rates of the registries are more unstable (the error probability of the national MBR registry has been observed to vary between 0.28 and 0.37, and the error probability of the MIA registry can range between 0.49 and 0.58 depending upon the choice of prior). Posterior model probabilities are somewhat sensitive to the relationships between the priors for all the different models considered, so it is important that these priors be chosen to have similar precisions.

4 Discussion

This paper represents an attempt to analyze data for which the frequentist approach has few refinements to offer. The data set considered is sparse in the sense that some configurations of the variables are represented by very few observations, and a frequentist analysis would be hampered by the relatively large number of parameters of interest. Moreover, it is far from obvious what kind of relationship should be hypothesised between the model components. In this situation we feel that a Bayesian approach, taking into account the uncertainty with regard to model description, offers a general solution that allows us to fully utilise the information available. In surveillance of birth defects carried out over a long period, considering yearly or quarterly registry data, it will sometimes be necessary to make important decisions on the basis of new observations from very limited time spans. Ultimately, such practical decisions must take into account prior information dealing with previous time periods as well. The Bayesian approach can easily incorporate such information through suitable prior distributions.

A major advantage of this approach is the possibility of adjusting for a covariate such as maternal age. However, even in the simple situation summarised in Table 2, without any adjustment, our procedure provides a unified picture of the plausibility of the assumptions embodied in the models considered. Model II, which assumes that all true cases are reported

from at least one registry, has the highest posterior probability; yet there is no strong evidence against either Model I or III. The complete data from the two registries also deal with several other birth defects, and the general instructions given to the personnel involved in the registration were meant to indicate that all cases assumed to belong to any particular defect should be reported. In retrospect, however, it does not seem unreasonable that the MIA registration may have been regarded more as a supplementary source of data by the practitioners involved, a practice which would be closer to Model II.

Our analysis has confirmed that there is a substantial amount of underreporting in the standard MBR registry as well as in MIA, regardless of choice of models. Hence MIA might be considered a poor source of additional information to supplement the MBR. Nonetheless, it is this extra, rather imprecise information which makes it possible at all to estimate without bias both the general prevalence of Down syndrome and the error rate in the MBR. For a defect such as Down's syndrome, an error rate of this magnitude should be unacceptable in the long term. Thus, a nationwide registration system based on more accurate information at birth should be warranted in Norway, despite the problems associated with routine collection of more sensitive information.

Lie *et al.* (1991) considered the possibility that ascertainment of Down's syndrome in the MBR was lower among infants of young mothers. If so, the observed twofold increase in the occurrence of Down's syndrome in MBR among young mothers in 1985–86 could be explained as a temporarily increased ascertainment in this group. However, in the data grouped across 1985–1988, no strong evidence of a maternal age dependent ascertainment in MBR was found.

In the data considered here, the overall estimates are rather similar in the analyses with and without age categorisation. This should not give the impression that age is a covariate that can be omitted. An important aspect of our analysis is the evaluation of the plausibility of models with direct links between age and case ascertainment in either registry. In particular, if data of this kind are analysed continually in a surveillance context, it is also essential to monitor the registry ascertainment, and look for changes which might

be indicated by interactions with age. Even in the simpler models in Table 5 where age has a direct influence only on the true state, changes over time in the general maternal age distribution may require a study of parameters specific to age categories. Figure 3 gives an example of more detailed results that can be produced in this way.

The prior knowledge about the relationship between risk of Down’s syndrome and maternal age was incorporated into our models by selecting suitable expectations in the prior distributions for the terms $\text{pr}(S | A_i)$. An alternative natural approach is to restrict the prior distributions so that the terms $\text{pr}(S | A_1), \text{pr}(S | A_2), \dots, \text{pr}(S | A_6)$ form an increasing sequence with probability one. The resulting posterior is analytically intractable but quite easy to sample from; see York (1992b) for details. The resulting posterior for the age-specific prevalences is scarcely different from what is presented here, and so we focus upon the simpler approach.

If the presence of false positives is suspected in doubly sampled data with no infallible classifier it may be difficult to extend the methodology presented here. There will typically be no empirical evidence available on the number of false positives, and so results would be very sensitive to the choice of priors for the false positive probabilities. This would also substantially increase the dimensionality of the unobservable data, which could create problems for the Monte Carlo schemes used here.

Perhaps the key failing of existing approaches to the analysis of doubly sampled data is that they fail to account for model uncertainty. The individual models we considered provide sharply different estimates for the prevalence of Down’s syndrome, yet the data does not clearly identify any one of the models as the “best”. In this context, basing inference on a single model is misleading. Raftery *et al.* (1994) and Draper (1994) demonstrate that the same problem arises in many other contexts such as linear regression and survival analysis. Model averaging, as described in this paper, provides a solution. The most sensitive point in the analysis is choosing priors for the different models so that they have roughly the same implications for shared quantities. This may not always be possible. Nonetheless, Raftery *et al.* (1994) and the references therein, show that model averaging consistently provides

superior out-of- sample predictive performance.

In summary, the advantages of the approach to double sampling presented in this paper are that it deals with covariates in spite of the errors in the registries and the large number of individuals, it facilitates incorporation of prior knowledge, and crucially , it accounts for model uncertainty.

References

- Besag, J., York, J.C. and Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics (with discussion.) *Ann. Inst. Statist. Math.*, **43**,1–53.
- Chen, T.T. (1979) Log-linear models for categorical data with misclassification and double sampling. *J. Am. Statist. Ass.*, **74**,481–488.
- Chen, T.T. (1989) A review of methods for misclassified categorical data in epidemiology. *Statist. Med.*, **8**,1095–1106.
- Cuckle, H.S., Wald, N.J. and Thompson, S.G. (1987) Estimating a woman’s risk of having a pregnancy associated with Down’s syndrome using her age and serum alpha-fetoprotein level. *Br. J. Obstet. Gynaecol.*, **94**,387–402.
- Draper, D. (1994) Assessment and propagation of model uncertainty. *J. R. Statist. Soc B*, to appear.
- Ekholm, A. (1991) Algorithms versus models for analyzing data that contain misclassification errors. *Biometrics*, **47**,1171–1182.
- Ekholm, A. and Palmgren, J. (1987) Correction for misclassification using doubly sampled data. *J. Official Statist.*, **3**,419–429.
- Espeland, M.A. and Odoroff, C.L. (1985) Log-linear models for doubly sampled categorical data fitted by the EM algorithm. *J. Am. Statist. Ass.*, **80**,663–670.
- Espeland, M.A. and Hui, S.L. (1987) A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics*, **43**,1001–1012.
- Gelfand, A. and Smith, A. F. M. (1991) Gibbs sampling for marginal posterior expectations. *Commun. Statist. Theory Meth.*, **20**, 1747–1766.
- Geng, Z. and Asano, C. (1989) Bayesian estimation methods for categorical data with misclassifications. *Commun. Statist. Theory and Meth.*, **18**,2935–2954.

- Hochberg, Y. (1977) On the use of double sampling schemes in analyzing categorical data with misclassification errors. *J. Am. Statist. Ass.*, **72**,914–921.
- Kass, R.E. and Raftery, A.E. (1994) Bayes factors and model uncertainty. *J. Am. Statist. Assoc.*, to appear.
- Lie, R.T., Heuch, I. and Irgens, L.M. (1991) A temporary increase of Down syndrome among births of young mothers in Norway: An effect of risk unrelated to maternal age? *Genet. Epidemiol.*, **8**, 217–230.
- Lie, R.T., Heuch, I. and Irgens, L.M. (1994) Maximum likelihood estimation of the proportion of congenital malformations using double registration systems. *Biometrics*, to appear.
- Madigan, D. and Raftery, A.E. (1994) Model selection and accounting for model uncertainty in graphical models using Occam’s window. *J. Am. Statist. Ass.*, to appear.
- Madigan, D. and York, J.C. (1993) Bayesian graphical models. Tech. Rep. 259, Department of Statistics, University of Washington.
- Nedelman, J. (1988) The prevalence of malaria in Garki, Nigeria: double sampling with a fallible expert. *Biometrics*, **44**,635–655.
- Raftery, A.E., Madigan, D., and Volinsky, C.T. (1994) Accounting for model uncertainty in survival analysis improves predictive performance. In : J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (Eds.) *Bayesian Statistics V*, Oxford University Press, to appear.
- Regal, R.R. and Hook, E.B. (1991) The effects of model selection on confidence intervals for the size of a closed population. *Statist. Med.* **10**, 717–721.
- Smith, A.F.M. and Roberts, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc B*, **55**, 3–23.
- Spiegelhalter, D.J. and Lauritzen, S.L. (1990) Sequential updating of conditional probabilities on directed graphical structures. *Networks*, **20**,579–605.
- Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation (with discussion). *J. Am. Statist. Ass.*, **81**, 82–86.
- Tenenbein, A. (1970) A double sampling scheme for estimating from binomial data with misclassification. *J. Am. Statist. Assoc.*, **65**,1350–1361.
- Tenenbein, A. (1972) A double sampling scheme for estimating from misclassified multinomial data with application to sampling inspection. *Technometrics*, **14**,187–202.

- Wei, G.C.G. and Tanner, M. A. (1990) Calculating the content and the boundary of the highest posterior density region via data augmentation, *Biometrika*, **77**, 649–652.
- Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.
- York, J.C. (1992a) Use of the Gibbs sampler in expert systems. *Artif. Intel.*, **56**, 112–130.
- York, J.C. (1992b) Bayesian methods for the analysis of misclassified or incomplete multivariate discrete data. PhD Thesis, Department of Statistics, University of Washington.