Biometrics 66, 603–612
June 2010
DOI: 10.1111/j.1541-0420.2009.01279.x

Bayesian Estimation of the Probability of Asbestos Exposure from Lung Fiber Counts

Scott Weichenthal,[1] Lawrence Joseph,[2,3,*] Patrick Bélisle,[2] and André Dufresne[4]

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec, H3A 1A2, Canada
[2]Division of Clinical Epidemiology, McGill University Health Center, 687 Pine Avenue West, V Building, Room V2.10, Montreal, Quebec, H3A 1A1, Canada
[3]Department of Epidemiology and Biostatistics, McGill University, 1020 Pine Avenue West, Montreal, Quebec, H3A 1A2, Canada
[4]Department of Occupational Health, McGill University, 1130 Pine Avenue West, Montreal, Quebec, H3A 1A3, Canada
[*]*email:* Lawrence.Joseph@mcgill.ca

Summary. Asbestos exposure is a well-known risk factor for various lung diseases, and when they occur, workmen's compensation boards need to make decisions concerning the probability the cause is work related. In the absence of a definitive work history, measures of short and long asbestos fibers as well as counts of asbestos bodies in the lung can be used as diagnostic tests for asbestos exposure. Typically, data from one or more lung samples are available to estimate the probability of asbestos exposure, often by comparing the values with those from a reference nonexposed population. As there is no gold standard measure, we explore a variety of latent class models that take into account the mixed discrete/continuous nature of the data, that each subject may provide data from more than one lung sample, and that the within-subject results across different samples may be correlated. Our methods can be useful to compensation boards in providing individual level probabilities of exposure based on available data, to researchers who are studying the test properties for the various measures used in this area, and more generally, to other test situations with similar data structure.

Key words: Asbestos exposure; Bayesian analysis; Diagnostic test; Hierarchical model; Latent class model; ROC curve; Sensitivity; Specificity.

1. Introduction

Occupational asbestos exposure is a well-recognized risk factor for the development of serious lung diseases including malignant mesothelioma. Reaching a conclusive decision on the association between asbestos exposure and lung diseases for the purpose of occupation related compensation is a process that may become arduous when a reliable work history is unavailable. Nevertheless, a decision must be made, and the impacts of such decisions can be substantial, affecting the financial welfare of families.

When limited information is available concerning the likelihood of asbestos exposure, compensating agencies rely on lung fiber counting and analysis as an exposure assessment tool. This practice is supported by a number of studies which have related lung fiber burden to asbestos exposure (Gylseth et al., 1981; Mowé et al., 1985; Dufresne et al., 1996). Several benchmarks of lung fiber retention have been proposed. For example, a lung fiber concentration of more than one million per gram of dried lung tissue is thought to be indicative of occupational exposure to asbestos (Gylseth et al., 1981).

Short asbestos lung fibers are less than 5 micrometers (μm) in length, whereas long fibers are greater than or equal to 5 μm in length. Long fibers are better predictors of diseases such as mesothelioma compared to short fibers, but short fibers are usually found in greater numbers. Although the association between short fibers and disease is less strong, short fibers may also expose an occupational history, as, like long fibers, they are found in higher concentrations in exposed individuals compared to a reference population. Asbestos bodies are asbestos fibers that have been coated with ferritin by macrophages.

Although data on all three types of lung fibers can provide useful information regarding the likelihood of occupational asbestos exposure, the numbers of short or long lung fibers or asbestos bodies counted can vary substantially between tissue samples both within and between individuals. This is an important concern for inferences about exposure because decisions based on lung fiber analysis based on a single block of lung tissue may lead to misclassification of a truly exposed or unexposed individual. In the past, evaluation boards relied on the analysis of only a single block of lung tissue in classifying occupational asbestos exposures. In recent years, however, it has become more common to examine three or four blocks of lung tissue to decrease the probability of exposure misclassification.

Whether using one or more tissue blocks, very little has been published about the properties of lung fiber counts as diagnostic tests for asbestos exposure history. Here we develop Bayesian latent class models to evaluate and compare the test properties of asbestos exposure classification from all three types of fiber counts discussed above. We consider each fiber type used singly, as well as when information from all three types of data are used in combination. We compare the test properties when only a single block of lung tissue is analyzed relative to when data from two to four blocks of lung tissue are available per subject. The problem is rendered especially difficult because, in the absence of a strong work exposure history, there is no gold standard method by which to classify each subject into truly exposed versus truly nonexposed categories. An additional complication is that the data provided by the three types of fiber counts do not follow standard distributions, in large part because there is always a value below which tests are not sensitive, leaving a mass of probability at that value. Further, the data within the different types of fiber counts may be correlated within individuals, indicating lack of conditional independence between blocks within each of these three tests. Although past literature has been divided about the importance of accounting for such dependence among diagnostic tests (Dendukuri and Joseph, 2001; Gustafson, 2005), we derive inferences from dependent as well as independent models, and compare results. Our model thus accommodates all important features of the data, and summarizes the inferences in several useful ways. In addition to standard diagnostic testing properties, such as sensitivity and specificity leading to receiver operating characteristic (ROC) curves, we provide methods for calculating the individual level probabilities of exposure given the test values on one or more tests and using data from one or more tissue blocks.

Bayesian latent class models for dichotomous diagnostic data arising from laboratory tests in the absence of a gold standard have been discussed by many, including Gastwirth, Johnson, and Reneau (1991); Joseph, Gyorkos, and Coupal (1995); Demissie et al. (1998); Johnson, Gastwirth, and Pearson (2001); and Gustafson (2005). Similar models for continuous data and ROC curves have been investigated by Zou and O'Malley (2005), although they assumed availability of a gold standard test result for each patient. Scott et al. (2008)considered normally distributed diagnostic test data in the absence of a gold standard, but considered only a single continuous test result from each subject, and did not account for the possibility of discrete probability masses in the test result distribution. Choi, Johnson, and Thurmond (2006)estimated disease prevalence and predictive probabilities for individual level continuous test results in the absence of a gold standard, but only considered results from a single test, and assumed the availability of a training sample of known positive and negative test results to estimate the distributions of truly disease positive and truly disease negative subjects. Erkanli et al. (2006) proposed a nonparametric analysis using mixtures of Dirichlet processes to model distributions within diseased and nondiseased subjects, leading to nonparametric ROC curve estimation, but again assumed a gold standard test. Because it is often the case in practice that results are available on more than one test per subject, including both repeated measures of the same test and results from different tests, the models developed extend the methodology available to date in important directions.

Section 2 describes the study setting, whereas Section 3 presents our models, discusses estimation of the test properties, and calculation of the probability of asbestos exposure. The results of applying our models to our fiber count data are in Section 4, and we conclude with a discussion.

2. Study Setting and the Source of Data

Occupational exposure to asbestos typically involves a mixture of mineral fibers of amphibole (banned since the early 1980s) or serpentine origin. Between 1996 and 2000, lung fiber retention analyses were conducted for 78 Quebec workers who had died of lung diseases potentially caused by occupational asbestos exposure. As previously described (Dufresne et al., 1996), lung retention data for the numbers of long and short fibers and asbestos bodies per milligram of dry lung tissue were collected. Seventy-five of these cases were men, and in total, 35 workers had three and 43 workers had four blocks of lung tissue examined. We also were able to obtain data on a limited number ($N = 41$) of controls. These data were from lung tissue of either accidental death or death caused by acute myocardial infarction in men autopsied between 1990 and 1992. These results will help estimate distributional parameters for our tests for nonexposed individuals.

3. Statistical Methods

Examining the available data for our three types of fiber counts revealed clear patterns. For short fibers, data ranged from 35 fibers per milligram (f/mg) to over 164,000 f/mg. The smallest values among our data set of controls was 70 f/mg, but in the data set of possible cases the lower limit was 35 f/mg, the next lowest value being over 100 f/mg. In each case, the lower limits (i.e., 35 or 70 f/mg) were reported for many subjects, making it clear that this was the minimum possible value. We therefore modeled these lowest values as a probability mass, and considered a lump sum probability as representing all values $\leq$70 f/mg. Once these lowest values were removed, a histogram of the logarithms of the remaining values looked approximately normal, both within the data sets of controls and among possible cases, the latter being similar to a binormal mixture of cases and controls. Near identical patterns were seen for long fibers, where the data ranged from 70 f/mg to almost 80,000 f/mg. Using a lowest cutoff value of 70 f/mg and taking logarithms again revealed approximate normality. For asbestos bodies, the lower cutoff limit was 40 bodies per gram (b/g), and the range of data was 40 b/g to almost 83,000 b/g.

All of our models assume that none of our tests provide perfectly accurate results, i.e., that no test is a gold standard for asbestos exposure. Our basic Bayesian latent class model consisted of a latent indicator of exposure status, and, conditional on this latent variable, the logarithms of the fiber count data were assumed to follow normal distributions with additional point masses representing all values equal to or below the relevant minimum possible value. Conditional on the latent exposure status, we first assumed that the counts from each type of fiber were independent from each other both within and between individuals. In a second similar model, we allow for within-individual correlations across tissue blocks for

each test. We also fit a hierarchical model that assumed conditional independence both between and within subjects, but only after further conditioning on a distinct mean value for each fiber type within each subject, these means assumed to be drawn from a common distribution between subjects. This model is a Bayesian version of a random effects model. We now provide the details of our models, starting with a model for each test used singly, and then show how the model can be extended when more than one type of fiber count is used, and/or there are several repeated values from eachsubject.

3.1 *One Mixed Discrete/Continuous Diagnostic Test*

Let θ be the true asbestos exposure rate in the population of interest. Let Z_i represent the latent true exposure history for individual $i, i = 1, 2, \ldots, n$. That is, $Z_i = 1$ if individual i was truly exposed to asbestos, and $Z_i = 0$ otherwise. Let $X_i, i = 1, 2, \ldots, n$ represent the logarithms of the fiber counts across the n individuals in the study, and let C represent the cutoff value representing the minimum value for the test. For the remaining values above this cutoff, let μ_E and σ_E represent the mean and standard deviation (SD) of the fiber count distribution conditional on a positive true (latent) asbestos exposure status, and let μ_{NE} and σ_{NE} represent the same quantities among the truly not asbestos exposed. Finally, let p_E and p_{NE} denote the probabilities of obtaining the minimum possible value in the exposed and nonexposed groups, respectively.

Let $\underset{\sim}{X}$ and $\underset{\sim}{Z}$ be vectors of fiber counts and latent exposure history on n subjects, respectively. Assuming for the moment that just a single observation is available per subject, the likelihood function of the data $\underset{\sim}{X}$ and augmented data $\underset{\sim}{Z}$ for a model using results from any one fiber test is

$$\begin{aligned} f\big(\underset{\sim}{X}, \underset{\sim}{Z} | \theta, \mu_E, \sigma_E^2, \mu_{NE}, \sigma_{NE}^2, p_E, p_{NE}\big) &= \prod_{i=1}^{n} \Bigg(\theta(1-p_E) \frac{1}{\sqrt{2\pi}\sigma_E} \\ &\quad \times \exp\left\{ -\frac{1}{2}(x_i - \mu_E)^2/\sigma_E^2 \right\} \Bigg)^{z_i \times I\{x_i > C\}} \\ &\quad \times (\theta p_E)^{z_i \times I\{x_i \le C\}} \\ &\quad \times \Bigg((1-\theta)(1-p_{NE}) \frac{1}{\sqrt{2\pi}\sigma_{NE}} \\ &\quad \times \exp\left\{ -\frac{1}{2}(x_i - \mu_{NE})^2/\sigma_{NE}^2 \right\} \Bigg)^{(1-z_i) \times I\{x_i > C\}} \\ &\quad \times ((1-\theta)p_{NE})^{(1-z_i) \times I\{x_i \le C\}}, \end{aligned} \tag{1}$$

where $I\{\cdot\}$ is the indicator function.

The estimation problem is similar to that of fitting a normal mixture model, with the extra complication of the point masses, which adds two parameters to be estimated, p_E and p_{NE}. We will use Bayesian methods, and thus require prior distributions over the parameter space (θ, μ_E, σ_E^2, μ_{NE}, σ_{NE}^2, p_E, p_{NE}). We will use beta prior distributions for θ, p_E and p_{NE}, which includes the beta(1,1) or uniform distribution as a special case, useful when little prior information will be incorporated into the analysis. Normal prior distributions are used for μ_E and μ_{NE}, and uniform distributions over a suitably chosen finite range for σ_E and σ_{NE}. As there is no closed form solution for the posterior distributions, we will use the Gibbs sampler as implemented in `WinBUGS version 1.4.1` (Lunn et al., 2000). We used 5000 iterations for burn-in and ran a further 20,000 iterations for use in inferences. Convergence of the Gibbs sampler algorithm was assessed by examining history plots across all iterations and by running the sampler several times from different starting values.

Of course, once all of the above parameters are estimated, one can estimate any function of these parameters, including the sensitivity and specificity of the fiber count test for any given cutpoint for positivity, which in turn leads to ROC curves (Hanley, 1996), although these ROC curves are not quite of the usual form because of the point masses. ROC curves plot the false positive rate (1 − specificity) versus the true positive rate (sensitivity) across the range of possible cutoff values for the test. In addition, it is of great clinical importance to know the probability of being exposed given any fiber count value, x_i. Once (θ, μ_E, σ_E^2, μ_{NE}, σ_{NE}^2, p_E, p_{NE}) have been estimated, the posterior probabilities P(exposed | x_i, θ, μ_E, σ_E^2, μ_{NE}, σ_{NE}^2, p_E, p_{NE}) are obtainable via the formula

$$P\big(\text{exposed} \,|\, x_i, \theta, \mu_E, \sigma_E^2, \mu_{NE}, \sigma_{NE}^2, p_E, p_{NE}\big) = \frac{\theta \frac{1}{\sqrt{2\pi}\sigma_E} \exp\left\{ -\frac{1}{2}(x_i - \mu_E)^2/\sigma_E^2 \right\}}{\theta \frac{1}{\sqrt{2\pi}\sigma_E} \exp\left\{ -\frac{1}{2}(x_i - \mu_E)^2/\sigma_E^2 \right\} + (1-\theta) \frac{1}{\sqrt{2\pi}\sigma_{NE}} \exp\left\{ -\frac{1}{2}(x_i - \mu_{NE})^2/\sigma_{NE}^2 \right\}},$$

when $x_i > C$, and P(exposed | x_i, θ, μ_E, σ_E^2, μ_{NE}, σ_{NE}^2, p_E, p_{NE}) = $\frac{p_E \theta}{p_E \theta + p_{NE}(1-\theta)}$ when $x_i \le C$. These formulae arise from the probability of obtaining the given fiber count value conditional on true exposure status, using the prevalence as a prior probability of being exposed, and converting these to probabilities of being truly exposed through Bayes' theorem.

To obtain unconditional posterior probabilities of exposure, P(exposed | x_i), one integrates over the posterior densities of all unknown parameters. This is easily accomplished by averaging P(exposed | x_i, θ, μ_E, σ_E^2, μ_{NE}, σ_{NE}^2, p_E, p_{NE}) over iterations in the Gibbs sampler.

3.2 *More than One Diagnostic Test*

Similar methods apply when there is more than one form of fiber counts used in the analysis. For example, where data from all three tests (long and short fibers and asbestos bodies) are available, the likelihood function of the data and augmented data (1) extends to

$$f(\underset{\sim}{X}_1, \underset{\sim}{X}_2, \underset{\sim}{X}_3, \underset{\sim}{Z} | \Theta) = \prod_{i=1}^{n} \prod_{j=1}^{3} \Bigg(\theta(1-p_{Ej}) \frac{1}{\sqrt{2\pi}\sigma_{Ej}}$$

$$\times \exp\left\{-\frac{1}{2}(x_{ij}-\mu_{Ej})^2/\sigma^2_{Ej}\right\}\Bigg)^{z_i \times I\{x_{ij}>C\}}$$
$$\times (\theta p_{Ej})^{z_i \times I\{x_{ij}\le C\}}$$
$$\times\Bigg((1-\theta)(1-p_{NEj})\frac{1}{\sqrt{2\pi}\sigma_{NEj}}$$
$$\times \exp\left\{-\frac{1}{2}(x_{ij}-\mu_{NEj})^2/\sigma^2_{NEj}\right\}\Bigg)^{(1-z_i)\times I\{x_{ij}>C\}}$$
$$\times((1-\theta)p_{NEj})^{(1-z_i)\times I\{x_{ij}\le C\}}, \quad (2)$$

where $\Theta = (\theta,\ \mu_{E1},\ \sigma^2_{E1},\ \mu_{NE1},\ \sigma^2_{NE1},\ \mu_{E2},\ \sigma^2_{E2},\ \mu_{NE2},\ \sigma^2_{NE2},\ \mu_{E3},\ \sigma^2_{E3},\ \mu_{NE3},\ \sigma^2_{NE3}, p_{E1},\ p_{E2},\ p_{E3},\ p_{NE1},\ p_{NE2},\ p_{NE3})$, and j indicates the jth test, $j = 1, 2, 3$. This model assumes conditional independence of results across three tests within each individual. Conditional independence is a much weaker condition compared to unconditional independence, in that results are assumed independent conditional on the (latent) true disease state for each individual. In Section 3.3 below, we describe two different models that account for correlations between repeated measurements of the same test within individuals, but we chose not to also account for any possible correlations between different tests, for two reasons: First, correlations between different tests seemed much lower than those within the same test within each individual. Second, to account for both types of correlations within the same model requires inverting a nine by nine (or larger) matrix, where each entry of the matrix is a very complex function of parameters of the model. Thus, the model becomes very unwieldy, making it difficult to estimate parameters, even using Monte Carlo techniques.

Although there are now many more parameters, the forms of the prior distributions can be chosen to be identical to the case of a single test. Of course, similar models can be created when any two tests are used rather than all three tests.

3.3 *More than One Measurement Per Test for Each Individual*

The above models assume just a single observation from each of the three tests is analyzed from each individual. Assuming independence of the observations both between and within individuals, conditional on the latent true state for each subject and given the values for all unknown parameters relating to the normal distributions and probabilities of values below the cutoffs, within subject observations for repeated observations on a single test can easily be accommodated. The likelihood function given in equation (2) can be used, the only addition being a further product term over the numbers of observations from each subject.

In our data set, some subjects had three observations for each fiber count type (i.e., three different tissue samples were taken), whereas others had four. As discussed in the introduction, it is of great interest to compensation boards to compare test properties as larger numbers of tissue samples are taken. In Section 4, we report on the evolving probabilities of exposure as more tissue samples are available to be analyzed.

All of the models discussed so far assume that the data across different blocks within each subject are conditionally independent. That is, given the true exposure status, there is no information about the results from one block given the results of the other blocks. However, this assumption likely does not hold for these data, as we found within subject across-block correlations above 0.3 within all three types of fiber counts. These approximate correlations were calculated after dividing the subjects into exposure classes "by eye," except for the 41 subjects who were a priori known to be nonexposed.

Some have argued that correlations may not always have a strong effect on final inferences (Dendukuri and Joseph, 2001; Gustafson, 2005), and sometimes a simpler model not including correlations may perform better. To investigate this issue, we created two distinct models that explicitly accommodate such correlations, and compared inferences from these models to those assuming independence. One model added specific correlation parameters within a multivariate normal distribution, whereas the other model handled correlations implicitly, via addition of a hierarchical component. As discussed in the next section, results were very similar between the two types of models accommodating correlations, but there were substantial differences between correlated and noncorrelated inferences. Given the observed correlations, the multivariate normal and hierarchical models are more plausible. Therefore, although we report results from both independent and correlated models, those from the independent models are included mostly for comparison purposes. Given the large number of different models run, and because results from our multivariate normal and hierarchical models are so similar, in Section 4 we present detailed results only from the hierarchical model.

The likelihood function for our correlated model using data from each of the three tests is similar to that given by equation (1) of Section 3.1, modified by considering m observations per subject from the test via an m-dimensional multivariate normal distribution, with a distinct correlation parameter for each test. Within each test, however, we assumed the same correlation parameter between each pair of measures, because these represent different samples from the same lung. Thus, the multivariate normal distribution used a single mean parameter across blocks, with variance-covariance matrix having just a single covariance parameter for all off-diagonal elements (compound symmetry). For the ith subject, and for data across blocks $j, j = 1, \ldots, m$, and test k we have

$$x_{i\cdot k} \sim MVN\left(\mu_k \underbrace{\begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}}_{m\times 1},\ \Sigma_k\right),$$

where the "$\cdot$" represents $j = 1, \ldots, m$, and where Σ_k is an $m \times m$ variance-covariance matrix with σ^2_k on the diagonal entries, and all off diagonal entries equal to $\rho_k\sigma^2_k$, where ρ_k is the between-block correlation parameter for test k.

In our hierarchical model, all observations, both within and between subjects were again considered as independent, but conditioned on distinct individual level mean parameters for the logarithms of each type of fiber count. These parameters, in turn, are tied together through a hierarchical normal distribution across individuals for each fiber type. This allows for the correlations that occur within individuals if only a single overall mean is used. Therefore, at the first level of our hierarchical model, the test results within each block for each test type are considered as independently normally distributed, but now each subject has their own unique mean value within each test. In turn, at the second level of the

hierarchical model, the mean values belonging to each subject within each test type are assumed to follow a normal distribution, with common mean and variance across all subjects. Thus, we have $x_{ijk} \sim N(\mu_{ik}, \sigma^2_{k,\text{within}})$, and $\mu_{ik} \sim N(\mu_k, \sigma^2_{k,\text{between}})$, where $\sigma^2_{k,\text{within}}$ is the within-subject variance across blocks and $\sigma^2_{k,\text{between}}$ is the between-subject variance.

In Section 4, we compare results from a wide variety of models, including those that use data from only one test at a time, and then from models that accommodate all three tests together. When more than one tissue block per subject is considered, we discuss models with both independent and correlated observations, the correlations modeled either through a multivariate normal model, or through a hierarchical structure. In addition, we compare results using $m = 1$, 2, 3, or 4 tissue samples per subject.

3.4 *Prior Distributions*

The prior distributions we used were roughly flat over the range of plausible values for each parameter. Throughout all models, we used uniform prior distributions on the range [0, 1] for the prevalence of asbestos exposure, equivalent to a beta(1,1) density. For the means of the three types of asbestos fiber counts on the log scale, we used independent normal prior distributions. For long fibers, we used $N(5,1)$ and $N(8,1)$ prior distributions for the unexposed and exposed populations, respectively. Similarly, for both short fibers and asbestos bodies, we used $N(6,1)$ and $N(8,1)$ distributions for the unexposed and exposed log means, respectively. Although these distributions help to focus the normal curves for the observed data within a reasonable range, avoiding possible convergence problems when using the Gibbs sampler, they are low in information on the log scale. For example, 95% of the range for a $N(8,1)$ distribution is approximately 400 to 22,000 f/mg when transformed back on the original scale, and the mean must be well within this range. Again on the log scale, we used uniform distributions on the range [0, 2] and [0, 6] for the SDs of long fibers in the unexposed and exposed populations, respectively, and similar uniform densities were used for both short fibers and asbestos bodies, with ranges [0, 2] and [0, 3] for unexposed and exposed groups, respectively. We also used uniform prior distributions over the range [0, 1] for the probabilities of being at the lowest possible value. In the correlated models, we used uniform prior distributions for the correlation parameters between all observations on the range [0, 0.95]. Correlations higher than 0.95 were not only implausible, but sometimes created convergence problems in running the Gibbs sampler, and so were eliminated a priori.

For our hierarchical models, we used normal prior distributions to represent the fiber count means across individuals. Again on a log scale, for short fibers and asbestos bodies, we used $N(6, 1)$ and $N(8,1)$ distributions within the nonexposed and exposed groups, whereas we used $N(5, 1)$ and $N(8,1)$ distributions for nonexposed and exposed means for long fibers. For within- and between-subject SDs for all three tests, we used uniform prior distributions covering the range [0.1, 3] and [0.01, 3], respectively. Lower SDs were not only implausible, but again occasionally caused convergence problems.

Although we did not always use the familiar flat prior distributions over a very wide range, our choice of prior parameter values are very low in information compared to the information in our data set. Varying the prior distributions produced no noticeable changes in results, so we do not report further on these robustness checks here.

4. Results

4.1 *Descriptive Statistics and Fiber Count Distributions*

The data collected included 78 persons with unknown asbestos exposure status that are to be classified, and 41 control subjects assumed to be asbestos free. Although the control subjects each contributed one block of tissue, among the 78 possible cases, 35 contributed three tissue blocks, whereas 43 contributed four blocks. The mean age was 64.5 years, with SD equal to 11.2 years. There were over 25 different primary diagnoses, the most common being adenocarcinoma (21 cases or 27%) and epidermoid (20 cases or 26%).

In order to appreciate the contributions from increasing amounts of data, we compare results from the situation with the least data, that is, a single observation from a single test, to cases with increasing amounts of data, up to results using three tests with four observations per subject per test. We refrain from discussing the prevalence of exposure, because with our specially selected sample, this parameter has little clinical meaning. Rather, we first draw inferences about distributions of test results in exposed and unexposed populations, and then closely examine how the test properties (sensitivity and specificity at various test cutpoints leading to ROC curves) and individual probabilities of exposure change using different tests and changing numbers of samples taken from each subject for each test.

Table 1 contains fiber count distributional results from our hierarchical latent class model including all of the data (i.e., three or four samples for subjects having more than one observation). For long fibers, there is a clear separation of the distributions in exposed versus unexposed subjects. Although under 4% ($1 - 0.966 = 0.034$) of exposed subjects have values at the lower limit of detection, almost 60% of unexposed subjects are at this value. For subjects above this threshold, the means on the log scale are also well separated ($7.72 - 5.44 = 2.28$, see also Figure 4). In addition, exposed subjects have larger variation compared to unexposed subjects (SD = 1.42 versus 0.3). The area under the ROC curve (AUC) for long fibers is 0.92, although with a relatively wide credible interval, because unlike the other estimates in this table, this value is based only on one block of data. ROC curves are not well defined when two or more blocks of data are used, because the choice of cutoff to use is not unique for two or more dimensional data.

Short fibers may be slightly less diagnostic than long fibers, with only 32% of unexposed subjects estimated to have values at the lower limit, compared to almost 60% in the case of long fibers. However, those above the lower threshold are widely separated, with a larger mean difference compared to long fibers. On the other hand, asbestos bodies have larger estimated SDs, and so the distributions of the exposed and unexposed will have larger overlap compared to long fibers. Therefore, we might expect the sensitivities and specificities

Table 1

Posterior medians and 95% credible intervals for the means and SDs of the three continuous tests for asbestos exposure, within both exposed and unexposed groups. P (above lowest value) represents the probability that a score for that test will not be at the lowest possible detectable value. Results are from the hierarchical model using all available data, except for estimating the AUCs, which are only well defined when one sample is used from each subject.

Variable	Posterior median	95% credible interval
Long fibers (on a log scale)		
Mean for exposed subjects	7.72	(7.23, 8.15)
SD for exposed subjects	1.42	(1.14, 1.79)
P (above lowest value) in exposed subjects	0.966	(0.928, 0.991)
Mean for unexposed subjects	5.44	(5.18, 5.85)
SD for unexposed subjects	0.30	(0.03, 0.76)
P (above lowest value) in unexposed subjects	0.423	(0.324, 0.517)
Area under the ROC curve	0.92	(0.76, 0.98)
Short fibers (on a log scale)		
Mean for exposed subjects	8.78	(8.38, 9.17)
SD for exposed subjects	1.33	(1.07, 1.68)
P (above lowest value) in exposed subjects	0.973	(0.937, 0.995)
Mean for unexposed subjects	6.30	(5.98, 6.65)
SD for unexposed subjects	0.43	(0.02, 0.90)
P (above lowest value) in unexposed subjects	0.680	(0.595, 0.756)
Area under the ROC curve	0.86	(0.74, 0.96)
Asbestos bodies (on a log scale)		
Mean for exposed subjects	8.44	(7.95, 8.95)
SD for exposed subjects	1.73	(1.40, 2.18)
P (above lowest value) in exposed subjects	0.958	(0.910, 0.995)
Mean for unexposed subjects	6.23	(5.68, 6.72)
SD for unexposed subjects	1.24	(0.87, 1.74)
P (above lowest value) in unexposed subjects	0.512	(0.425, 0.600)
Area under the ROC curve	0.88	(0.76, 0.97)

for short fibers and asbestos bodies to be slightly lower compared to long fibers.

4.2 *ROC Curves*

Figure 1 displays ROC curves for short fibers, long fibers, and asbestos bodies, based on data from block 1 only. ROC curves are not well defined when two or more blocks of data are used, because the cutoff to choose for each subject when they provide two or more values is not unique. The small jumps arise from the masses that represent the probability of being at the lower limits of detection for each test.

The AUC represents the probability that a randomly selected truly positive subject and a randomly selected truly negative subject will be correctly ordered by the continuous test. For long fibers the AUC is 0.92, although with a relatively wide credible interval (see Table 1), in part because this estimate is based only on one block of data. The AUC for short fibers is 0.86, slightly less than the point estimate for long fibers; however, the very wide credible intervals leave much uncertainty. The AUC for asbestos bodies is 0.88, intermediate to the AUCs from the other two tests.

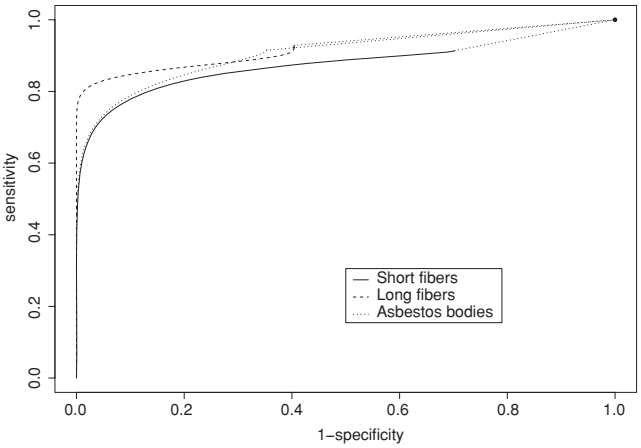

Figure 1. ROC curves for short fibers, long fibers, and asbestos bodies, based on data from block 1 only.

4.3 *Individual Probabilities of Asbestos Exposure*

Of most interest to compensation boards is the probability of exposure, given data from any subject. A byproduct of running any of the models described in this article are estimates of the probabilities of exposure across all subjects. We will provide examples of how these probabilities vary between subjects by selecting three prototypic subjects and examining their exposure probabilities across the full range of models we have developed.

Subject #1 has a mixof values both at and above the lower limits of detection. Subject #2 has high values for each test, with no observations at the lower limit. Subject #3 has moderate values across the three tests, again with none at the lower limit. The data (on a log scale) from all three tests and

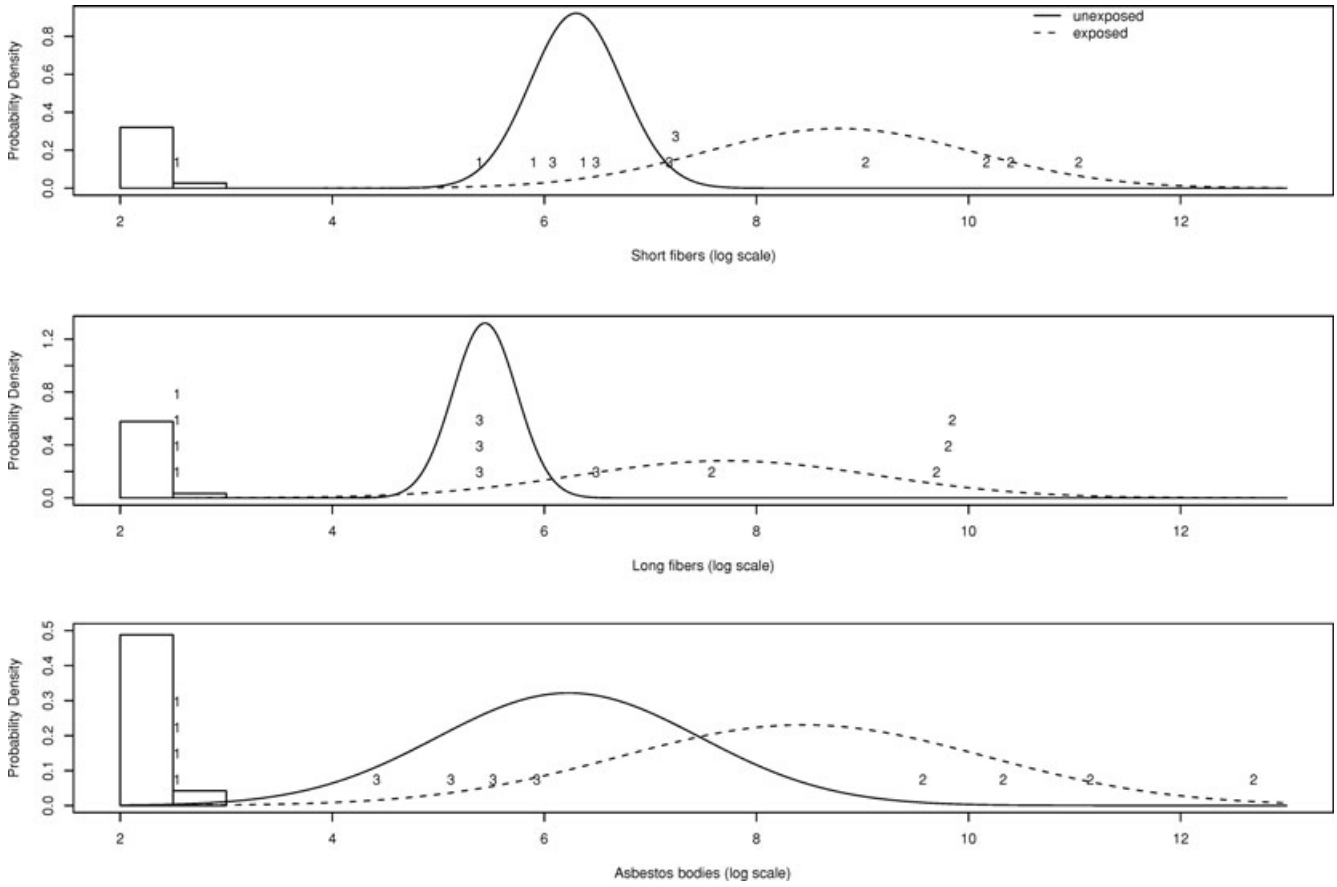

Figure 2. Posterior median estimates of the distributions for nonexposed and exposed subjects for short fibers, long fibers, and asbestos bodies. Superimposed on each graph are the data values for each of our three subjects, as discussed in Section 4.3.

all four time points from these three subjects are depicted in Figure 2. The figure also displays the population distributions for all three tests for unexposed and exposed groups, based on the posterior medians for the mean and SD parameters for each curve, as given in Table 1.

Subject #1 had all results for both long fibers and asbestos bodies at the lowest possible value C, but four different values for short fibers, one at the lowest possible value C, and three above this threshold. Depending on which model is used for inferences and which subset of the data is included in the analysis, one can obtain different estimates of the probability that subject #1 has been exposed to asbestos. We have three different tests, each having up to four blocks of tissue data to analyze using each of the models we ran. Of course, one would usually want to use all available data in any real case, but here it is also of great interest to examine how our estimates change as more data accumulate, and as more tests are included. We display results only for the hierarchical model that accounts for within-subject correlations, because the results for the multivariate normal model with correlation parameters were virtually identical to those from the hierarchical model across all parameters of interest.

Figure 3 presents the probability of exposure for all three subjects across all models and using increasing numbers of tissue blocks. If one only looks at the results from subject #1 short fibers (labeled as SF in the figure), one can say very little about the probability of exposure, as even with all four tissue blocks included in the analysis, the credible interval for this probability ranges from close to zero to almost 0.9. Similarly, little can be concluded if only a single block of data is used from long fibers or asbestos bodies for this subject. However, as soon as data from two or more blocks are included, the probability of exposure is concentrated very near to 0, indicating no exposure. Only one block of data is required if one combined data from all three tests, again indicating a probability of exposure close to zero. For subject #1, there is little difference between a model that assumed independent observations within subjects, versus the hierarchical model.

Subject #2 had generally high values across all three tests, with no values hitting the lower threshold, C. Here, even a single observation from any test is sufficient to classify this subject as exposed, and these high probabilities remain near one with very high probability regardless of which model is run or how many data points are used. Note that the lower limit of the y axis of this graph is at 0.992, so that even a single short fibers observation provides very strong evidence of exposure.

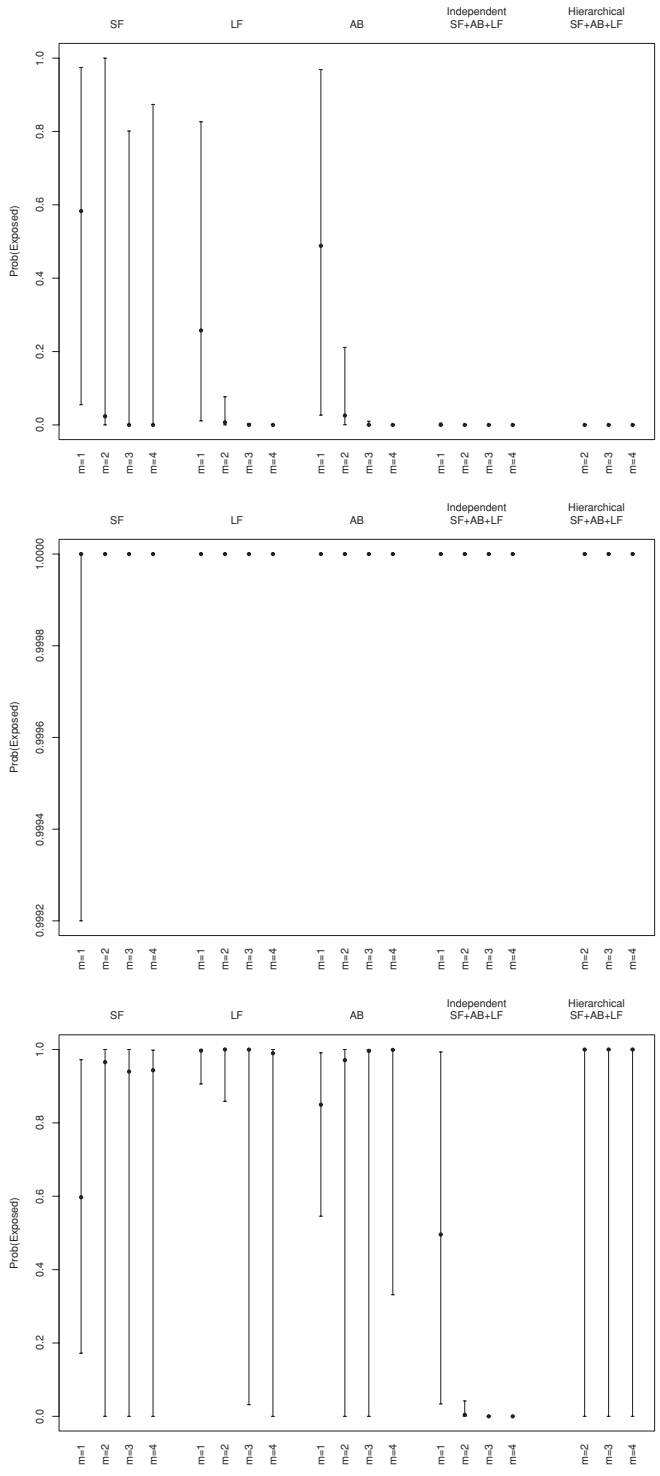

Figure 3. From top to bottom, probability of asbestos exposure for subjects #1, #2, and #3, respectively, from a wide variety of models, and for changing values of tissue blocks used in the analyses, ranging from $m = 1$ to $m = 4$ blocks. From left to right, we have models using data from short fibers alone, long fibers alone, asbestos bodies alone, all three tests with independence assumed between observations within subjects, and a hierarchical model using all three tests that accounts for correlations within subjects. The dots represent posterior medians.

In contrast to subjects #1 and #2, where clear decisions are possible, the data from subject #3 shows that this is not always the case, even when all available data are used. Results range from near certainty of exposure if one or two long fiber observations are examined, to certainty of nonexposure if all data are used, and the model assumes independence between all observations. However, all other models indicate great uncertainty about the probability of exposure, and because the independence model is probably not valid, one must admit that the data are not sufficient to make a strong recommendation. This example also clearly shows the danger of relying on only a single type of fiber count rather than all the data, and illustrates that a model that assumes independence can provide very different estimates from our correlated data models, even though exactly the same data are input into both models.

The contrasting results between correlated and noncorrelated models from subject #3 raises the issue of goodness-of-fit or model selection procedures. Indeed, some authors (for example Black and Craig, 2002) have performed formal model selection procedures or averaged the results over several models. Here, however, we have very strong reasons to doubt models that do not incorporate dependence between blocks within subjects, and both models that account for these correlations provide virtually identical estimates. This is not surprising, as correlations arise because data for each test within subjects are more similar compared to data between subjects, and both of our models account for this.

Figure 4 displays the mean exposed minus nonexposed mean differences, $\mu_E - \mu_{NE}$ across all models, which is interesting for several reasons. First, this parameter is of importance by itself, because the distance between exposed and nonexposed distributions is a marker for the usefulness of a continuous diagnostic test. Second, we can clearly see how accuracy for this parameter is affected by the number of data blocks used. For asbestos bodies, there is not much increase in accuracy, as judged by the length of the credible interval, as more data blocks are added. For both short and long fibers, there is an increase in accuracy going from $m = 1$ to $m = 2$ tests, but not much improvement after that. We can also see smaller mean differences in the hierarchical model as compared to a model that assumes independence of data blocks within subjects, perhaps an indicator that the amount of information is somewhat exaggerated in the independence model. We can also see similar sized credible intervals for $m =$ 2, 3, or 4 in the hierarchical model, indicating that there is extra variability accounted for here which does not substantially decrease with increasing m.

5. Discussion

We have developed a series of Bayesian latent class models for mixed continuous/discrete diagnostic test data, and applied these models to determine the probability of asbestos exposure from lung fiber count data. We have shown that incorrect inferences may be made if only a single block of data is analyzed, and that for many subjects a clear decision is possible using a model that uses all possible data. However, for some subjects, even 12 data points are not sufficient for a definitive assessment.

There is no substitute for a detailed work history in determining the likelihood of occupational asbestos exposure. This

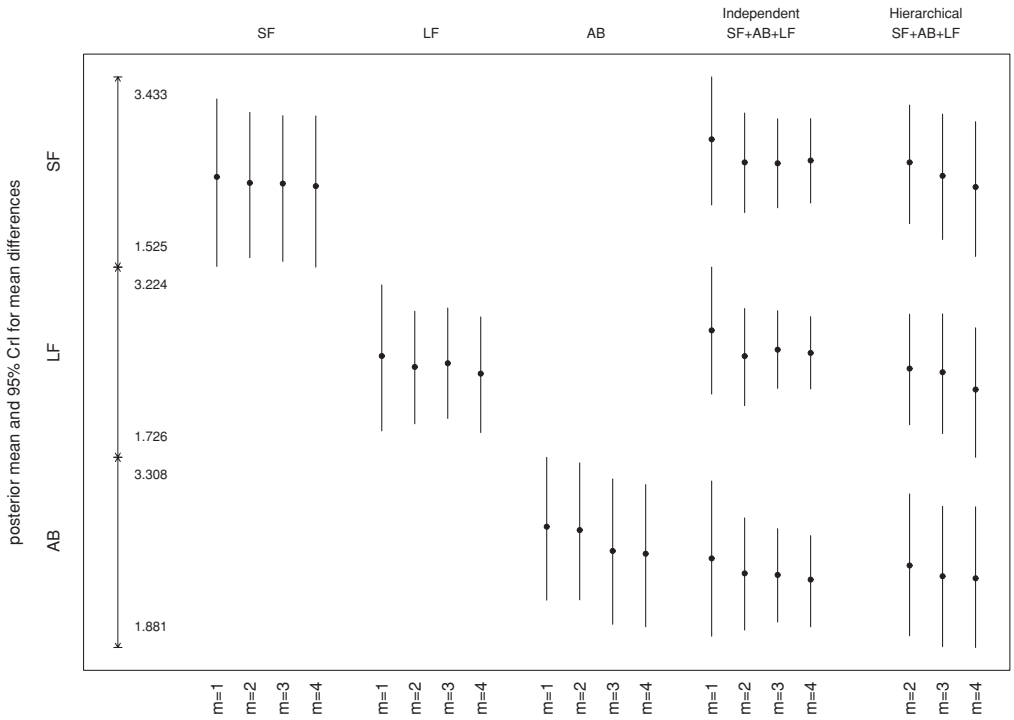

Figure 4. Posterior median and 95% credible intervals for the difference between means in the exposed minus nonexposed populations from a wide variety of models, and for changing values of tissue blocks used in the analyses, ranging from $m = 1$ to $m = 4$ blocks. From left to right, we have models using data from short fibers alone, long fibers alone, asbestos bodies alone, all three tests with independence assumed between observations within subjects, and a hierarchical model using all three tests that accounts for correlations within subjects.

is particularly true for occupational exposure to chrysotile fibers, which do not accumulate as readily in the lungs, and for which the number of fiber-years of exposure is the best indicator of lung fiber burden. Unfortunately, detailed occupational exposure histories are not available for all suspected cases of asbestos-related lung disease, and in such circumstances, lung fiber retention analysis is one alternative method. Our models provide direct estimates of the probability of exposure, given all data collected, and have shown that collecting more than one data block per subject improves these estimates, resulting in better decisions concerning compensation.

References

Black, M. A. and Craig, B. A. (2002). Estimating disease prevalence in the absence of a gold standard. *Statistics in Medicine* **21,** 2653–2669.

Choi, Y. K., Johnson, W. O., and Thurmond, M. C. (2006). Diagnosis using predictive probabilities without cut-offs. *Statistics in Medicine* **25,** 699–717.

Demissie, K., White, N., Joseph, L., and Ernst, P. (1998). Bayesian estimation of asthma prevalence, and comparison of exercise and questionnaire diagnostics in the absence of a gold standard. *Annals of Epidemiology* **8,** 201–208.

Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* **57,** 208–217.

Dufresne, A., Begin, R., Churg, A., and Masse, S. (1996). Mineral fiber content of lungs in mesothelioma cases seeking compensation in Quebec. *American Journal of Respiratory and Critical Care Medicine* **153,** 711–718.

Erkanli, A., Sung, M., Costello, E. J., and Angold, A. (2006). Bayesian semi-parametric ROC analysis. *Statistics in Medicine* **25,** 3905–3928.

Gastwirth, J. L., Johnson, W. O., and Reneau, D. M. (1991). Bayesian analysis of screening data: Application to AIDS in blood donors. *Canadian Journal of Statistics* **19,** 135–150.

Gustafson, P. (2005). On model expansion, model contraction, identifiability, and prior information: Two illustrative scenarios involving mismeasured variables (with discussion). *Statistical Science* **20,** 111–129.

Gylseth, B., Mowé, G., Skaug, V., and Wannag, A. (1981). Inorganic fibers in lung tissue from patients with pleural plaques or malignant mesothelioma. *Scandinavian Journal of Work, Environment, and Health* **7,** 109–113.

Hanley, J. A. (1996). The use of the "binormal" model for parametric ROC analysis of quantitative diagnostic tests. *Statistics in Medicine* **15,** 1575–1585.

Johnson, W. O., Gastwirth, J. L., and Pearson, L. M. (2001). Screening without a "gold standard": The Hui-Walter paradigm revisited. *American Journal of Epidemiology* **153,** 921–924.

Joseph, L., Gyorkos, T. W., and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* **141,** 263–272.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). `WinBUGS`—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10,** 325–337.

Mowé, G., Bjorn, G., Hartveit, F., and Skaug, V. (1985). Fiber concentration in lung tissue of patients with malignant mesothelioma. *Cancer* **56,** 1089–1093.

Scott, A., Joseph, L., Bélisle, P., Behr, M., and Schwartzman, K. (2008). Bayesian estimation of tuberculosis clustering rates from DNA sequence data. *Statistics in Medicine* **27,** 140–156.

Zou, K. H. and O'Malley, A. J. (2005). A Bayesian hierarchical nonlinear regression model in receiver operating characteristic analysis of clustered continuous diagnostic data. *Biometrical Journal* **47,** 417–427.

Received June 2008. *Revised February* 2009. *Accepted February* 2009.