Bayesian modelling of imperfect ascertainment methods in cancer studies

Sasha Bernatsky[1,2], Lawrence Joseph[1,2,*,†], Patrick Bélisle[1], Jean-François Boivin[2,3], Raghu Rajan[1,4], Andrew Moore[5] and Ann Clarke[1,5]

[1] *Division of Clinical Epidemiology, Department of Medicine, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Que., Canada H3G 1A4*
[2] *Department of Epidemiology and Biostatistics, 1020 Pine Avenue West, McGill University, Montreal, Que., Canada H3A 1A2*
[3] *Centre for Clinical Epidemiology and Community Medicine, Jewish General Hospital, Montreal, Que., Canada*
[4] *Department of Oncology, McGill University, Montreal, Que., Canada*
[5] *Division of Clinical Immunology and Allergy, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Que., Canada*

SUMMARY

Tumour registry linkage, chart review and patient self-report are all commonly used ascertainment methods in cancer epidemiology. These methods are used for estimating the incidence or prevalence of different cancer types in a population, and for investigating the effects of possible risk factors for cancer. Tumour registry linkage is often treated as a gold standard, but in fact none of these methods is error free, and failure to adjust for imperfect ascertainment can lead to biased estimates. This is true both if the goal of the study is to estimate the properties of each ascertainment type, or if it is to estimate cancer incidence or prevalence from one or more of these methods. Although rarely applied in the literature to date, when cancer is ascertained by three or more methods, standard latent class models can be used to estimate cancer incidence or prevalence while adjusting for the estimated imperfect sensitivities and specificities of each ascertainment method. These models, however, do not account for variations in these properties across different cancer sites. To address this problem, we extend latent class methodology to include a hierarchical component, which accommodates different ascertainment properties across cancer sites. We apply our model to a data set of 169 lupus patients with three ascertainment methods and eight cancer types. This allows us to estimate the properties of each ascertainment method without assuming any to be a gold standard, and to calculate a standardized incidence ratio for cancer for lupus patients compared to the general population. As our data set is small, we also illustrate the effects as more data become available. We show that our model produces

*Correspondence to: Lawrence Joseph, Division of Clinical Epidemiology, Department of Medicine, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Que., Canada H3G 1A4.
†E-mail: lawrence.joseph@mcgill.ca

parameter estimates that are substantially different from the currently most popular method of ascertainment, which uses tumour registry data alone. Copyright © 2005 John Wiley & Sons, Ltd.

1. INTRODUCTION

Methods of cancer ascertainment commonly used in epidemiological research include self-report (or report by a proxy), medical chart review, and linkage with cancer registries. Each method has advantages and limitations.

One benefit of using self-report is that information can be obtained not only on outcomes, but also on exposures that may not be captured by other means. However, this method of cancer ascertainment may introduce inaccuracies, because of under- or over-reporting by patients or proxies [1–3]. For example, errors in self-report may arise from lack of knowledge of tumour type, or because pre-malignant lesions may be confused with actual cancers, such as in cervical disease and skin lesions [2]. In addition, use of self-report will miss cases of cancer in those deceased or lost to follow-up, potentially causing substantial under-reporting. Similarly, ascertainment by chart review poses a problem for patients deceased or lost to follow-up, or when not all care occurs at a single hospital. Registry linkage is often considered to be the best option, but it is well known that tumour registries may not include all cases, because of under-reporting, incomplete coverage, or human error [4–8].

In the past 15 years, numerous authors have reported on the variable accuracy of regional cancer registries in North America and the U.K. [4–7], and a few assessments of other methods of cancer ascertainment, including chart review and self-report, have been published [1, 2]. There is little data, however, on the comparison of chart review, self-report, and tumour registry linkage within the same population, and few have commented on the consequences of imperfect ascertainment inherent in all of these methods. In most studies, a single method of ascertainment is chosen, and results such as cancer incidence are reported ignoring the possible misclassification error. The usual procedure when comparing one method of cancer ascertainment to another is to select one as the gold standard; generally, this standard is linkage with a registry. Of course, this ignores the possibility that even tumour registries have imperfect sensitivities and specificities [8]. A further problem is that the sensitivity and specificity of each method of cancer ascertainment may vary with cancer type, even though general tendencies across cancer types may be expected within each ascertainment method. In addition, many studies are based on low numbers of cases of each cancer type; thus, simple estimates, aside from being biased by misclassification through use of an imperfect reference standard, may also be inaccurate due to small sample sizes.

Previous methods for parameter estimation in studies using imperfect tests have included latent class models from both frequentist [9, 10] and Bayesian [11] perspectives. Although in theory these methods are applicable to cancer ascertainment studies, they have rarely been applied in this setting. Conditional independence occurs when test results are statistically independent of each other, given the true disease status of each subject. When three or more conditionally independent tests or ascertainment methods are available, maximum likelihood methods can provide asymptotically unbiased estimates of the sensitivity and specificity of all

tests or methods, as well as the overall incidence or prevalence of the condition, without imposing any restrictions on the parameter space [9]. Restrictions are necessary when data from only one or two tests are available, or when tests are not conditionally independent. In such cases Bayesian methods can replace the need for restrictions with the use of an informative prior distribution over a subset of the unknown parameters. Bayesian [12] and frequentist [13] methods are also available when the tests or ascertainment methods are correlated with each other (i.e. not conditionally independent).

Whether frequentist or Bayesian, previous methods do not allow for the combining of information on several methods of disease ascertainment across two or more cancer types. Without this, one either must assume a common sensitivity and specificity across all cancer types, which is unrealistic, or treat each cancer type as a completely distinct problem, ignoring the general tendencies of each ascertainment method across cancer types, and resulting in very wide interval estimates when sample sizes are small or moderate. Here, we propose a hierarchical extension to the Bayesian methods previously developed [11], which has two main advantages. First, it allows for simultaneous estimation of distinct sensitivity and specificity parameters for all three methods and for each cancer type. Second, it allows for the usual 'borrowing of strength' across cancer types, in order to obtain reasonable parameter estimates for all types, even in small data sets [14]. This represents a compromise between the pooling of data across cancer types at one extreme, and independent estimates for each cancer type at the other, with the degree of pooling determined by the data. Typically, a small amount of bias in each individual parameter estimate due to hierarchical pooling is traded for lower overall mean square error, averaged across all parameter estimates [14]. If three or more tests are available, diffuse or non-informative prior distributions can be used, so that the final statistical inferences are driven almost entirely by the data. Of course, if reliable prior information is available, it can serve to further sharpen the inferences.

Adjusting for imperfect ascertainment data is important for several reasons. First, the sensitivities and specificities of each ascertainment method are of interest by themselves, as these properties need to be well understood by researchers planning cancer epidemiology studies or interpreting results from such studies. Second, the incidence or prevalence of cancer arising from studies using imperfect ascertainment methods needs to be correctly adjusted in order to obtain consistent estimates. Third, biased estimation of primary parameters carries over to quantities based on these estimates, such as measures of effect. For example, large errors can occur in calculating SIRs, owing to a possibly different mix of ascertainment methods used to form the numerator and denominator.

In this article, we first provide the details of our latent class hierarchical model for cancer ascertainment data. Next, for a small data set of 169 lupus patients, we use our model to estimate sensitivity and specificity of self-report, chart review and tumour registry methods for each of eight cancer types, along with the incidence of each type. Our model also provides summary estimates of overall sensitivity and specificity parameters for each method across all cancer types. We compare our estimates to those from standard methods which assume registry data to be a perfect gold standard, and also to a non-hierarchical Bayesian latent class model, which accounts for imperfect registry data, but does not pool data or 'borrow strength' across cancer types. To illustrate the effect of increasing sample size, we compare parameter estimates from all models assuming our data set was ten times its actual size. We examine how imperfect ascertainment biases SIR estimates from our lupus data, and present an adjusted estimate. To provide context to this work, we also include

a brief survey of methods of cancer ascertainment in current use in three major medical journals.

2. METHODS

2.1. Literature review

We surveyed the frequency of methods of cancer ascertainment used in cohort studies in three prominent medical journals (the New England Journal of Medicine, the Journal of the American Medical Association, and the Journal of the National Cancer Institute) during 1999–2001. A search was done for articles whose title or abstract included the keywords 'cancer' or 'malignancy' together with 'cohort' or 'prospective'. All articles describing original cohort studies with a primary outcome of cancer incidence (i.e. not of cancer mortality) were included, and information on which cancer ascertainment methods were used was extracted. We tabulated the number of times each method or combination of methods was used, as well as whether adjustments for imperfect ascertainment were carried out.

2.2. A Bayesian hierarchical model for cancer ascertainment

The statistical methods described below were applied to data from a cohort consisting of 207 patients followed at the Montreal General Hospital Lupus Clinic during the period from 1984 to 1998. Each patient attending the clinic was invited to complete a postal survey on malignancy occurrence and to give their permission for a chart review and for linkage of their name with the Quebec Tumour Registry. Approval was obtained from the local ethics review board, and written informed consent was obtained from each participant. All but 16 agreed to participate, although a further 22 were unable or unwilling to complete the survey and thus were not included in the study. The survey asked, 'Have you ever had a cancer or malignant tumour? If so, what type of cancer was it?' and when the cancer occurred. The patients seen in the Lupus Clinic are followed regularly, with an annual clinic visit. For our study, the clinic charts were reviewed by a physician for documentation of cancer occurrence. The participating patients were also linked to the Quebec Tumour Registry, a provincial database that records cancer occurrence primarily from hospital admissions and day-hospital procedures [15]. The registry also receives data from other provinces when Quebec cancer patients are treated there. We thus had data on cancer occurrence using all three methods for 169 subjects.

While the aetiology is still unknown, it has long been suspected that lupus patients have cancer rates well above that of the general population [16]. Therefore, although our data set is small, high rates of cancers in lupus patients provided a total of 16 cancers among these patients, or 9.5 per cent. The cancers were of eight different types: haematological, breast, endometrial, cervical, colorectal, central nervous system, melanoma, and ovarian.

Our three level latent class hierarchical model is described as follows: Label the subjects $i = 1, 2, \ldots, 169$, the three cancer ascertainment methods, $j = 1$ (self-report), 2 (chart review) or 3 (tumour registry), and let the cancer types be labelled $k = 1, 2, \ldots, 8$, in the same order as in the above list. The probability that a randomly selected subject truly has developed cancer of type k is given by the prevalence, π_k. Given cancer of type k, the probability that cancer ascertainment method j detects the cancer is given by the sensitivity S_{jk}, while the specificity, or the probability of ascertainment method j not detecting cancer of type k in subjects truly free from

that cancer type, is denoted by C_{jk}. Let $x_{ijk} = 1$ (0) denote that subject i is detected as having (not having) cancer of type k by ascertainment method j. Given this notation, the probability that ascertainment method j detects cancer of type k in subject i, including both true and false positives, is given by $\Pr\{x_{ijk} = 1\} = \pi_k S_{jk} + (1 - \pi_k)(1 - C_{jk})$. Similarly, the probability that ascertainment method j does not detect cancer of type k in subject i, including both true and false negatives, is given by $\Pr\{x_{ijk} = 0\} = \pi_k(1 - S_{jk}) + (1 - \pi_k)C_{jk}$. If we assume conditional independence, which is reasonable in our situation since the methods of cancer ascertainment operate very differently (see Section 4), expressions for results across all three ascertainment methods are available. For example, $\Pr\{x_{i1k} = 1 \text{ and } x_{i2k} = 1 \text{ and } x_{i3k} = 1\} = \pi_k S_{1k} S_{2k} S_{3k} + (1 - \pi_k)(1 - C_{1k})(1 - C_{2k})(1 - C_{3k})$, and so on.

Let n_{111k} denote the total number of subjects in our data set who are ascertained as having cancer of type k by all three methods, n_{110k} denote the number of subjects positive on the first two tests but not in the tumour registry, n_{101k} denote those positive by tumour registry and self-report but not by chart review, and so on, down to n_{000k}, which represents the number of subjects who are ascertained as negative by all three methods. Of course, in our data set, $n_{111k} + n_{011k} + n_{101k} + n_{110k} + n_{001k} + n_{010k} + n_{100k} + n_{000k} = 169$. For each cancer type k, the independent likelihood function contribution for our data, across all patients and ascertainment methods is

$$[\pi_k S_{1k} S_{2k} S_{3k} + (1 - \pi_k)(1 - C_{1k})(1 - C_{2k})(1 - C_{3k})]^{n_{111k}}$$

$$\times [\pi_k S_{1k} S_{2k}(1 - S_{3k}) + (1 - k)(1 - C_{1k})(1 - C_{2k})C_{3k}]^{n_{110k}}$$

$$\times [\pi_k S_{1k}(1 - S_{2k})S_{3k} + (1 - \pi_k)(1 - C_{1k})C_{2k}(1 - C_{3k})]^{n_{101k}}$$

$$\times [\pi_k(1 - S_{1k})S_{2k}S_{3k} + (1 - \pi_k)C_{1k}(1 - C_{2k})(1 - C_{3k})]^{n_{011k}}$$

$$\times [\pi_k S_{1k}(1 - S_{2k})(1 - S_{3k}) + (1 - \pi_k)(1 - C_{1k})C_{2k}C_{3k}]^{n_{100k}}$$

$$\times [\pi_k(1 - S_{1k})S_{2k}(1 - S_{3k}) + (1 - \pi_k)C_{1k}(1 - C_{2k})C_{3k}]^{n_{010k}}$$

$$\times [\pi_k(1 - S_{1k})(1 - S_{2k})S_{3k} + (1 - \pi_k)C_{1k}C_{2k}(1 - C_{3k})]^{n_{001k}}$$

$$\times [\pi_k(1 - S_{1k})(1 - S_{2k})(1 - S_{3k}) + (1 - \pi_k)C_{1k}C_{2k}C_{3k}]^{n_{000k}}$$

This forms the first level of our hierarchical model, which takes into account individual sampling variability, including variability due to whether an individual truly has cancer of type k, and whether this cancer is detected by one or more of the methods. At the second level, the logit of each of the eight sensitivities and specificities (across cancer types) from each test are assumed to follow normal distributions respectively, which reflects the fact that each ascertainment method can operate differently within each cancer type. At this stage, we allow variability of the properties of each ascertainment method across different cancer types, but within a common model. We used $\mathrm{logit}(S_{jk}) \sim \mathrm{normal}(\mu_{1j}, \sigma_{1j}^2)$, and $\mathrm{logit}(C_{jk}) \sim \mathrm{normal}(\mu_{2j}, \sigma_{2j}^2)$, $k = 1, 2, \ldots, 8$, $j = 1, 2, 3$. This stage allows for the 'borrowing of strength' across cancer types when estimating the properties of each ascertainment method, while still allowing distinct parameters for each method. In the third level of our hierarchical model, we placed prior distributions on these second level hierarchical parameters, with $\mu_{hj} \sim \mathrm{normal}(0, 4)$, and $\sigma_{hj} \sim \mathrm{uniform}[0.2, 2]$, $h = 1, 2$, $j = 1, 2, 3$. We bounded the

uniform distribution for the σ_{hj} parameters away from zero since it is highly implausible for these to be very close to zero. Finally, our model specification is completed by independent uniform prior distributions over the interval $[0,1]$ for the incidences, π_k, $k=1,2,\ldots,8$. While usually considered as a non-informative prior, in very low prevalence situations, this prior may result in slightly increased prevalence estimates, as it is equivalent to adding one negative and one positive subject to the data.

The choice of hierarchical variance parameters were based on substantive considerations. As probabilities range from 0.01, to 0.99, the logistic model covers a range of approximately -4.6 to $+4.6$, and the range is about -6.9 to $+6.9$ for probabilities in the range 0.001 to 0.999. Therefore, standard deviations (SDs) of 2–3 easily cover the entire range of feasible values. The SD of 2 for the normal hierarchical parameters is relatively non-informative, since within each ascertainment type, the sensitivities and specificities should cover a much smaller range than those given above, so the normal distribution should be relatively flat in the area of posterior concentration. Similarly, the uniform range for the standard deviations covers the most plausible range. Nevertheless, we checked the robustness of the chosen range by also running models with $\sigma_{hj} \sim \text{uniform}[0.05,4]$ and $\sigma_{hj} \sim \text{uniform}[0.1,3]$, $h=1,2$, $j=1,2,3$.

Note that inferences about the sensitivities and specificities of the cancer ascertainment methods are available at two different levels. First, marginal posterior densities can be estimated for each of the 24 sensitivity and specificity parameters, each providing information about how a particular ascertainment method operated within a given cancer type. Second, overall summaries of these properties across cancer types within each ascertainment method are available from the second level of the hierarchical model, by taking the inverse logit of the normal hierarchical distributions.

In order to further assess the effect of the hierarchical component of our model, we also fit a model [11] with no hierarchical component. While this model accounts for the absence of a gold standard, it does not pool data across cancer types. This involves the trade-off of fewer modelling assumptions for wider interval estimates within each cancer type. It is also difficult to summarize overall properties of each ascertainment methods across cancer types using this non-hierarchical model.

Estimating functions of the above parameters, such as standardized incidence ratios (SIRs) that are adjusted for imperfect ascertainment methods, is straightforward. For example, consider the situation of estimating the SIR for lupus patients, defined as the ratio of cancer incidence in lupus patients divided by the incidence in the general population. The incidence of cancer in our lupus cohort is defined as $\sum_{i=1}^{8} \pi_k$, which is already adjusted for the imperfect reference standards if the methods described above are used. The denominator may be estimated from the Quebec Tumour Registry, by applying incidence rates in the registry to the age–sex matched distribution of person years in the lupus cohort. While this number is not adjusted for the imperfect registry data, an adjusted estimate can be derived from the formula $(\pi + C - 1)/(S + C - 1)$, where S and C are the appropriate sensitivity and specificity estimates for the tumour registry, and π is the unadjusted estimate for the incidence of cancer from the registry [11]. Note that while incidences are usually defined in terms of events per person-year rather than raw counts of cancers, the person-years are the same in the numerator and denominator in the above calculation, and so cancel out. Misclassification errors, however, do not cancel out, since different ascertainment methods leading to different magnitudes of errors are used in the numerator and denominator of the SIR. The methods described above rectify this problem.

No analytic solution is available for estimating the parameters of these complex models. We therefore used the Gibbs sampler, implemented via the WinBUGS program [17], to generate random samples from our target posterior densities. These samples are then used for inferences about the marginal posterior density of each parameter, which we summarized by posterior means and 95 per cent highest posterior density (HPD) credible intervals. The Gibbs sampler is an iterative method, where random samples are drawn from the full conditional distribution of each parameter in turn. The full conditional distribution of each parameter is defined as the distribution of that parameter conditional on all other parameters. In our case, the set of full conditional distributions is made simpler by the addition of latent variables, represented by the true but unobserved cancer status for each subject. See Reference [11] for further details about using the Gibbs sampler in a similar diagnostic test setting with latent variables, Gelman *et al.* [14] or Gilks *et al.* [18] for general information about the Gibbs sampler, and Spiegelhalter *et al.* [17] for information on the BUGS programming language. We used the method of Gilks *et al.* [18] to estimate the number of iterations required for accurate estimation, leading to a choice of 20 000 iterations following a 'burn-in' (number of initial samples discarded before convergence is reached) of 4500. All analyses were run from a variety of starting points to further ensure convergence [14]. The WinBUGS code used is available from the authors.

Throughout, we use confidence intervals when reporting results from frequentist methods, and HPD intervals for summarizing results from Bayesian methods. Confidence intervals are only used for comparison purposes in the upper half of Table II.

3. RESULTS

3.1. Literature review

We retrieved 42 articles [19–61], with the vast majority using a single method of cancer ascertainment, without assessing its accuracy. In articles relying on tumour registry data, all authors implicitly assumed that accuracy of the registry data was perfect. Cancer outcome was assessed using a cancer registry alone in 18 of the 42 articles (43 per cent) [44–61], while 16 (38 per cent) used self-report (telephone or postal survey) alone [19–34] and 1 (2 per cent) used a review of medical records alone [35]. Two studies (4 per cent) used two methods, including chart review plus self report [39] or plus physical exam [40]. Another two studies [41, 42] used three methods (self-report, cancer registry and chart review). Finally, three studies used other methods, including endoscopy [36, 37] and prostate specific antigen assays [38]. Of the 16 studies that relied on self-report alone, a handful mentioned an attempt to confirm the cases with medical documentation such as a pathology report [22, 23, 30, 34, 43], but the vast majority assumed their ascertainment methods to be perfect, even when the main method used was not linkage with a tumour registry. None of the papers we reviewed mentioned latent class modelling or made any attempt to adjust results for an imperfect gold standard.

3.2. Data analysis

The average age of the 169 participating Lupus Clinic patients was 47.2 (standard deviation 13.5), similar to that of the 38 cohort members who did not participate (42.0, standard deviation 15.0). As systemic lupus erythematosus is a disease primarily affecting women, the

majority (91 per cent) of participating subjects were female as were the majority of those who declined (89 per cent).

Table I presents the data from our study. Sixteen cancers were reported by one or more methods; eight of these were found by all three methods, and the other eight were reported by only one or two methods. Therefore, there were disagreements among the methods for half of all cancers found.

Table II presents overall estimates of the sensitivity and specificity of postal survey and chart review methods, first assuming the tumour registry is a perfect gold standard, and then using our three level Bayesian hierarchical model, which does not assume any gold standard. The estimates in the upper section of the table are found by pooling the data over all cancer types, treating tumour registry as a perfect gold standard reference, and calculating the sensitivities and specificities for the postal survey and chart review separately. Confidence intervals were then found by an exact procedure [62]. The Bayesian estimates in the lower half of

Table I. Number of subjects in the Montreal General Hospital Lupus cohort with a cancer occurrence (by type) as determined by the three methods of cancer ascertainment.

Method			Cancer type							
SR	CR	TR	Haematologic	Breast	Endometrial	Cervical	Colorectal	CNS	Melanoma	Ovarian
+	+	+	1	4	1	1	1	0	0	0
+	+	−	0	0	0	0	0	0	0	0
+	−	+	0	0	0	0	0	0	0	0
−	+	+	0	0	0	2	0	0	0	0
+	−	−	0	0	0	2	0	0	2	1
−	+	−	0	0	0	0	0	0	0	0
−	−	+	0	0	0	0	0	1	0	0
−	−	−	168	165	168	164	168	168	167	168

SR = self-report, CR = chart review, TR = tumour registry linkage, CNS = central nervous system.

Table II. Sensitivity and specificity estimates for three methods of cancer ascertainment in the Montreal General Hospital Lupus Cohort.

	Sensitivity (95 per cent CI)	Specificity (95 per cent CI)
Considering tumour registry data as the reference standard		
Postal survey	0.727 (0.390, 0.940)	0.968 (0.928, 0.990)
Chart review	0.909 (0.587, 0.998)	1.000 (0.977, 1.000)
Tumour registry	1.000*	1.000*
	Sensitivity (95 per cent HPD)	Specificity (95 per cent HPD)
Bayesian estimates (no single method is considered the reference standard)		
Postal survey	0.736 (0.236, 0.999)	0.992 (0.975, 1.000)
Chart review	0.876 (0.511, 1.000)	0.998 (0.993, 1.000)
Tumour registry	0.885 (0.547, 1.000)	0.996 (0.990, 1.000)

*Assumed to be perfectly accurate by the methodology. CI indicates a standard confidence interval, while HPD represents Bayesian highest posterior density intervals.

Table II were found by sampling from the normal(μ_{hj}, σ_{hj}^2) distributions, $h = 1, 2$, and applying the inverse logit, representing, respectively, the sensitivities and specificities for a 'typical' cancer type. The index j varied across the three ascertainment methods, producing the results for each.

Of 11 cancers identified by tumour registry linkage, eight were identified on postal survey, for a sensitivity (compared to tumour registry) of 0.727 (95 per cent CI 0.390, 0.940); the specificity was 0.968 (95 per cent CI 0.928, 0.990). Chart review identified 10 of 11 malignancies recorded in the tumour registry, for a sensitivity of 0.909 (95 per cent CI 0.587, 0.998). All 10 malignancies found on chart review were recorded in the tumour registry. Using our Bayesian hierarchical model, the intervals for the sensitivities are generally wider, since the possibility of errors in the tumour registry is allowed, leading to a higher degree of uncertainty. The estimates for the specificities changed little. This methodology allows us to estimate the sensitivity and specificity of the tumour registry for cancer occurrence; these values were 0.885 (95 per cent HPD 0.547, 1.000) and 0.996 (95 per cent HPD 0.990, 1.000), respectively. While the point estimates are high, sensitivities as low as 55 per cent are not ruled out for at least some cancer types.

It is important to understand the different interpretations of the two types of inferences presented in Table II. The sensitivity and specificity estimates based on the tumour registry standard are found by a simple pooling of data across cancer types. Therefore, for example, five missed cancers of any single type count the same as five missed cancers, each of a different type. It is very difficult to provide estimates within each cancer type, because the small sample sizes lead to wide confidence intervals. In other words, pooling the data hides differences in the properties of the ascertainment methods across different cancer types, while individual estimates (by cancer type) are imprecise. Finally, no estimates of the true tumour registry sensitivity and specificity are provided, as this method is (unrealistically) assumed to be perfect.

In contrast, the hierarchical model provides estimates of the properties of all three methods, and provides a compromise between pooled and individual estimates. Each cancer type is given its own sensitivity and specificity parameters for each type of ascertainment method. The overall summary is of these parameters, and changes depending on how the detected cancers are distributed across cancer types. In addition, individual parameter estimates can be estimated more precisely, because (depending on the observed data) a certain degree of pooling takes place. If there is no strong evidence of differences in properties across cancer types (as is the case with our data) then a larger degree of pooling will automatically take place. The interpretation of the overall summary measures given in Table II, therefore, is that for a 'typical cancer type'. The 95 per cent HPD intervals reflect both random sampling variability and the variability due to differences in sensitivity and specificity across cancer types. The later term is omitted when estimates are based on tumour registry data alone.

If one assumes that tumour registry provides a perfect reference standard, then the incidence of all cancers combined is simply estimated by the 11 cancers found by this method in 169 subjects, giving 6.5 per cent (95 per cent CI 3.3 per cent, 11.3 per cent). However, the hierarchical model gives more weight to the five cancers ascertained by other methods and the possibility that there are cancers detected by none of the methods, thus providing a much higher incidence estimate of 10.8 per cent (95 per cent HPD 6.2 per cent, 16.2 per cent). Individual incidences ranged from lows of less than 1 per cent for CNS, melanoma and ovarian cancers to a high of 2.9 per cent for breast cancer.

Table III. Bayesian estimates for the sensitivity and specificity of the three methods of cancer ascertainment in the Montreal General Hospital Lupus Cohort, according to type of cancer.

Cancer Type (N)	Self-report	Chart review	Tumour registry
Sensitivity			
Haematologic (1)	0.789 (0.403, 0.999)	0.890 (0.626, 1.000)	0.903 (0.646, 1.000)
Breast (4)	0.856 (0.601, 0.999)	0.923 (0.744, 1.000)	0.927 (0.754, 1.000)
Endometrial (1)	0.788 (0.407, 0.999)	0.895 (0.623, 1.000)	0.903 (0.647, 1.000)
Cervical (5)	0.592 (0.184, 0.946)	0.913 (0.701, 1.000)	0.918 (0.717, 1.000)
Colorectal (1)	0.791 (0.401, 0.999)	0.897 (0.623, 1.000)	0.902 (0.640, 1.000)
CNS (1)	0.723 (0.213, 0.999)	0.859 (0.430, 1.000)	0.886 (0.554, 1.000)
Melanoma (2)	0.737 (0.229, 0.999)	0.868 (0.465, 1.000)	0.880 (0.515, 1.000)
Ovarian (1)	0.736 (0.233, 0.999)	0.872 (0.478, 1.000)	0.882 (0.530, 1.000)
Specificity			
Haematologic (1)	0.997 (0.992, 1.000)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)
Breast (4)	0.997 (0.991, 1.000)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)
Endometrial (1)	0.997 (0.991, 1.000)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)
Cervical (5)	0.992 (0.981, 0.999)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)
Colorectal (1)	0.997 (0.991, 1.000)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)
CNS (1)	0.997 (0.991, 1.000)	0.999 (0.996, 1.000)	0.997 (0.991, 1.000)
Melanoma (2)	0.992 (0.981, 0.999)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)
Ovarian (1)	0.995 (0.987, 0.999)	0.999 (0.996, 1.000)	0.998 (0.995, 1.000)

N = total number of cases reported by any method. CNS = central nervous system. Numbers in parentheses represent Bayesian 95 per cent HPD intervals.

Table III presents estimates for the sensitivity and specificity of the three methods of cancer ascertainment according to type of cancer, from our three level hierarchical model. Across all cancer types, the sensitivity of tumour registry was higher than that of self-report, although tumour registry and chart review appeared similar in terms of both sensitivity and specificity. With our small data set, little can be said about the differences in the properties of each ascertainment method across cancer types.

In larger data sets, HPD intervals for the sensitivities and specificities for each method within each cancer type will be narrower, and firmer conclusions will be available. To illustrate this, and the impact of different models, Figure 1 compares four possible models for estimating the sensitivity of self-report for haematologic cancer using a small data set (Table I) and a larger data set (Table I data with each cell multiplied by 10). Model 1 is the simplest possible model, where sensitivity is assumed to be the same across all cancer types, so that data from all types can be pooled. Model 2 does not pool, using data from the column for haematologic cancer only. In both Models 1 and 2, registry is considered as the gold standard. Models 3 and 4 are the two models described in Section 2.2. While neither assume tumour registry to be a perfect gold standard, Model 4 pools across cancer types, while Model 3 does not. In the small data set, there are large differences between models, with non-pooled models providing virtually no useful information due to very wide confidence intervals. Comparing the models which do pool data, the hierarchical model provides a slightly wider interval, as uncertainty is introduced by acknowledging the imperfections in the tumour registry. Compared to the smaller data set, the larger data set provides narrower interval estimates, and the different assumptions behind the models in some cases provide quite different estimates. Overall, the tradeoffs are clear: assuming that the distinct sensitivity and specificity parameters are drawn
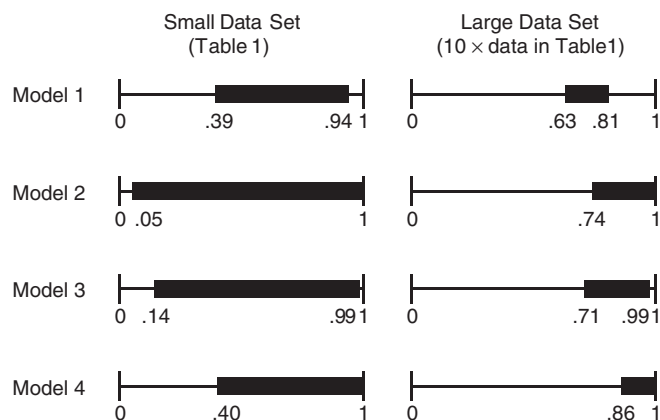
Figure 1. Ninety-five per cent HPD interval estimates for the sensitivity of self-report haematologic cancer across four different models, in two different data sets. Models 1 and 2 assume the tumour registry to be a perfect gold standard, while models 3 and 4 do not. Models 2 and 3 do not pool data across cancer types, while model 1 pools the data from all types. Model 4 is a hierarchical model that allows the data to automatically select the degree of pooling which takes place.

from the same distribution (hierarchical model) results in narrower intervals compared to the unpooled models that do not impose this assumption. Pooled models that assume all cancer types have the same sensitivity may be unrealistic, as sensitivites may vary across cancer types.

The bias inherent in falsely assuming tumour registry data to be a gold standard, in turn, has consequences for other estimates based on these numbers. For example, there is growing evidence that lupus patients have higher rates of cancer, which is possibly due to either intrinsic pathogenic pathways, or external exposures such as immunosuppressive medications [16]. Using the Quebec Tumour Registry, we calculated 6.54 expected cases cancers in our cohort of 169 lupus patients, assuming the patients are at the same risk as the general Quebec population. We observed 11 tumour registry cases in our subjects, resulting in an estimated SIR of 1.68 (95 per cent HPD 0.81, 2.68). However, adjusting both the numerator and denominator for possible biases due to imperfections in the tumour registry results in an estimate of 2.56 (95 per cent HPD 1.32, 3.90), a substantially different result. Adjusting the numerator alone leads to an estimate of 2.81 (95 per cent HPD 1.62, 4.19).

Almost all results remained stable as we changed the range of the uniform prior distribution of the standard deviation from [0.2, 2] to [0.1, 3] to [0.05, 4]. In particular, all prevalences, specificities and SIR estimates remained virtually unchanged. The sensitivities, however, were affected by the choice of prior distribution. This is not surprising, given that the low prevalences across all cancer types meant that very little data were available for estimating the sensitivity parameters. This effect was diminished in a similar robustness study using 10 times our amount of data.

4. DISCUSSION

Tumour registry linkage is a common method of cancer ascertainment in epidemiological studies, and is often assumed to be error-free. We describe a method for parameter estimation

in this setting without assuming a gold standard. This in turn leads to adjusted incidence, prevalence and SIR estimates that account for imperfect ascertainment methods. The methodology is applicable when data from one or more methods are available across two or more cancer types, or more generally, whenever diagnostic tests are applied to two or more related populations. We focus on latent data models; other methods have been suggested, such as using a composite reference standard [63], but these methods can also lead to bias as again an imperfect standard is used.

Our methodology depends on the model being a reasonable approximation to reality, including the forms of the hierarchical distributions, conditional independence, independence of cancer incidence within each individual, and the simplifying assumption that cancer incidence is constant across the lupus population. The latter assumption can easily be removed by using a logistic regression model for incidence that allows different patients to assume different probabilities of developing each cancer type, depending, for example, on characteristics such as age and length of follow-up. With only 16 events across eight cancer types, however, we did not attempt to add the many additional parameters that would be required by this model. Although our model assumed that the logits of individual sensitivity and specificity values across cancer types follow normal distributions, other hierarchical forms can easily be substituted. Independence of cancer incidence within each individual is not a strong assumption in practice, since low incidences within each cancer type means that the probability of contracting two or more types is very low, regardless of how it is modelled.

Our chart data come from outpatient charts not used by the cancer registry, and patient self-report was by mail and not tied to any physician visit. While conditional independence thus seems reasonable for our data, this may not carry over to other settings. Conditional dependence can also be handled [12, 13], although further constraints must be imposed [9]. Our model assumes *a priori* exchangeability between ascertainment methods, but if prior information suggests differences between methods, this can be accommodated by adding a fourth level to the model. For example, one might allow the sensitivities to depend on a regression term such as cancer type.

In our data set, the specificities of all three methods of cancer ascertainment were similarly high, but our sensitivity estimates suggest that malignancies can be missed by each of the different methods. This in turn led to higher prevalence estimates than using any single method as a gold standard. Self-report did not seem as sensitive as chart review or tumour registry linkage, and this appeared to hold across cancer types, although our small data set does not allow for strong conclusions.

While we used data on eight cancer types, we in fact scanned the database for 25 difference cancer types, only eight of which were found by one or more ascertainment methods. Thus, our data set included zero cases for another 17 cancer types. In theory, we could have included all 25 cancer types, with the addition of 119 extra parameters (17 additional prevalence parameters, and 6×17 additional sensitivity and specificity parameters across the three ascertainment parameters). However, with no observed cases, there would be little information upon which to estimate the sensitivities, and specificities would again be very high, matching the distributions already estimated. Prevalence estimates would be low, since no cases were observed. Therefore, the extra information to be gained was judged to be not worth the much larger model required.

Our population may have some unique characteristics (a group of persons with a serious disease, under regular medical follow-up) that may affect generalizability of the results,

although lower sensitivity of self-report compared to other methods has been documented else-where [2]. Many factors affect completeness of cancer case ascertainment; for chart review, this includes the inability to find charts, which may vary from centre to centre [64, 65].

Our calculations of the sensitivity and specificity of tumour registry data and of the SIR for lupus patients illustrate that estimates can be highly inaccurate if a 'gold standard' which itself is imperfect is used without adjustment. Although our estimate of the sensitivity of the Quebec tumour registry linkage for cancer ascertainment was relatively high, and other registries may be able to do even better [65, 66], even small imperfections can have large effects on important parameters estimated in studies [67]. Therefore, whenever possible, future studies should consider using a combination of different methods of cancer ascertainment, and present not only estimates of the sensitivity and specificity of the methods, but also adjusted estimates of any quantities which depend on these parameters.

REFERENCES

1. Bergmann MM, Calle EE, Mervis CA *et al*. Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *American Journal of Epidemiology* 1998; **147**:556–562.
2. Schrijvers C, Stronks K, van de MD *et al*. Validation of cancer reports from postal survey with cancer registry record. *American Journal of Epidemiology* 1994; **139**:408–414.
3. Berthier F, Grosclaude P, Bocquet H *et al*. Prevalence of cancer in the elderly: discrepancies between self-reported and registry data. *British Journal of Cancer* 1997; **75**:445–447.
4. Swerdlow AJ, Douglas AJ, Vaughan HG *et al*. Completeness of cancer registration in England and Wales. *British Journal of Cancer* 1993; **67**:326–329.
5. Rushton LR. Comparison of the diagnosis of leukemia from death certificates, cancer registration and histological reports—implications for occupational case–control studies. *British Journal of Cancer* 1997; **75**(11):1694–1698.
6. Bowie C. The validity of a cancer register in leukemia epidemiology. *Community Medicine* 1987; **9**:152–159.
7. Alexander FE, McClaren EA, Cartwright RA. Cancer registration of leukaemias and lymphomas. *Community Medicine* 1989; **11**:81–89.
8. Wilson S, Prior P, Woodman CBJ. Use of cancer surveillance data for comparative analyses. *Journal of Public Health Medicine* 1992; **14**(2):152–156.
9. Walter S, Irwig L. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *Journal of Clinical Epidemiology* 1988; **41**:923–937.
10. Formann A, Kohlmann T. Latent class analysis in medical research. *Statistical Methods in Medical Research* 1996; **5**:179–211.
11. Joseph L, Gyorkos T, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 1995; **141**:263–272.
12. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; **57**:158–167.
13. Qu Y, Tan M, Kutner M. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996; **52**:797–810.
14. Gelman A, Carlin J, Stern H, Rubin H. *Bayesian Data Analysis*. Chapman & Hall: New York, 1995.
15. Online Fichier des Tumeurs du Quebec Sante et Services sociaux. Government of Quebec, 2002 (http://www.msss.gouv.qc.ca/f/statistiques/tumeurs.htm).
16. Bernatsky S, Clarke A, Ramsey-Goldman R. Malignancy and systemic lupus erythematosus. *Current Rheumatology Reports* 2002; **4**(4):351–358.
17. Spiegelhalter D, Thomas A, Best N. *WinBUGS Version 1.2 User Manual*. MRC Biostatistics Unit: Cambridge, U.K., 1999.

18. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, U.K., 1996.

19. Zhang S, Hunter DJ, Hankinson SE *et al.* A prospective study of folate intake and the risk of breast cancer. *Journal of the American Medical Association* 1999; **281**:1632–1637.

20. Michaud DS, Giovannucci E, Willett WC *et al.* Physical activity, obesity, height, and the risk of pancreatic cancer. *Journal of the American Medical Association* 2001; **286**:921–929.

21. Michaud DS, Spiegelman D, Clinton SK *et al.* Fruit and vegetable intake and incidence of bladder cancer in a male prospective cohort. *Journal of the National Cancer Institute* 1999; **91**:605–609.

22. Platz EA, Rimm EB, Willett WC *et al.* Racial variation in prostate cancer incidence and in hormonal system markers among male health professionals. *Journal of the National Cancer Institute* 2000; **92**:2009–2017.

23. Feskanich D, Ziegler RG, Michaud DS *et al.* Prospective study of fruit and vegetable consumption and risk of lung cancer among men and women. *Journal of the National Cancer Institute* 2000; **92**:1812–1823.

24. Velie E, Kulldorff M, Schairer C *et al.* Dietary fat subtypes and breast cancer in postmenopausal women: a prospective cohort study. *Journal of the National Cancer Institute* 2000; **92**:833–837.

25. Neglia JP, Friedman DL, Yasui Y *et al.* Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study. *Journal of the National Cancer Institute* 2001; **93**:618–629.

26. Fuchs C, Edward LG, Graham AC *et al.* Dietary fiber and the risk of colorectal cancer and adenoma in women. *New England Journal of Medicine* 1999; **340**:169–176.

27. Sturmer T, Glynn RJ, Lee M *et al.* Lifetime cigarette smoking and colorectal cancer incidence in the physicians' health study—I. *Journal of the National Cancer Institute* 2000; **92**:1178–1182.

28. Bertone ER, Willett WC, Rosner BA *et al.* Prospective study of recreational physical activity and ovarian cancer. *Journal of the National Cancer Institute* 2001; **93**:942–948.

29. The Breast Cancer Linkage Consortium. Cancer risks in Brca2 mutation carriers. *Journal of the National Cancer Institute* 1999; **91**:1310–1316.

30. Schairer C, Lubin J, Troisi R *et al.* Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk. *Journal of the American Medical Association* 2000; **283**:485–491.

31. Grabrick DM, Hartmann LC, Cerhan JR *et al.* Risk of breast cancer with oral contraceptive use in women with a family history of breast cancer. *Journal of the American Medical Association* 2000; **284**:1791–1798.

32. Lowenfels A, Maisonneuve P, Whitcomb D *et al.* Cigarette smoking as a risk factor for pancreatic cancer in patients with hereditary pancreatitis. *Journal of the American Medical Association* 2001; **286**(2):169–170.

33. Holmes M, Hunter DJ, Colditz G *et al.* Association of dietary intake of fat and fatty acids with risk of breast cancer. *Journal of the American Medical Association* 1999; **281**:914–920.

34. Gertig DM, Hunter DJ, Cramer DW *et al.* Prospective study of talc use and ovarian cancer. *Journal of the National Cancer Institute* 2000; **92**:249–252.

35. Schoen R, Tangen C, Kuller L *et al.* Increased blood glucose and insulin, body size, and incident colorectal cancer. *Journal of the National Cancer Institute* 1999; **91**:1147–1154.

36. Bani-Hani K, Martin IG, Hardie LJ *et al.* Prospective study of cyclin D1 overexpression in Barrett's esophagus: association with increased risk of adenocarcinoma. *Journal of the National Cancer Institute* 2001; **92**:1316–1321.

37. Limburg PJ, Qiao Y-L, Mark SD *et al.* Helicobacter pylori seropositivity and subsite-specific gastric cancer risks in linxian china. *Journal of the National Cancer Institute* 2001; **93**:226–228.

38. Carter HB, Landis PK, Metter EJ *et al.* Prostate-specific antigen testing of older men. *Journal of the National Cancer Institute* 1999; **91**:1733–1735.

39. Hemstreet GP, Yin S, Ma Z *et al.* Biomarker risk assessment and bladder cancer detection in a cohort exposed to benzidine. *Journal of the National Cancer Institute* 2001; **93**:427–436.

40. Mark SD, Qiao Y-L, Dawsey SM *et al.* Prospective study of serum selenium levels and incident esophageal and gastric cancers. *Journal of the National Cancer Institute* 2000; **92**:1753–1763.

41. Schulman S, Lindmarker P. Incidence of cancer after prophylaxis with warfarin against recurrent venous thromboembolism. *New England Journal of Medicine* 2000; **342**:1953–1958.

42. Yu M, Chang HC, Liaw Y-F *et al.* Familial risk of hepatocellular carcinoma among chronic hepatitis B carriers and their relatives. *Journal of the National Cancer Institute* 2000; **92**:1159–1164.

43. Michaud DS, Spiegelman D, Clinton SK *et al.* Fluid intake and the risk of bladder cancer in men. *New England Journal of Medicine* 1999; **340**:1390–1397.

44. Holowaty P, Miller A, Rohan T *et al.* Natural history of dysplasia of the uterine cervix. *Journal of the National Cancer Institute* 1999; **91**:252–258.

45. Stolzenberg-Solomon R, Albanes D, Javier N *et al.* Pancreatic cancer risk and nutrition-related methyl-group availability indicators in male smokers. *Journal of the National Cancer Institute* 1999; **91**:535–541.

46. Travis LB, Holowaty EJ, Bergfeldt K *et al.* Risk of leukemia after platinum-based chemotherapy for ovarian cancer. *New England Journal of Medicine* 1999; **340**:351–357.

47. Woodson K, Tangrea JA, Barrett MJ *et al.* Serum a-tocopherol and subsequent risk of lung cancer among male smokers. *Journal of the National Cancer Institute* 1999; **91**:1738–1743.

48. Chow WH, Gridley G, Fraumeni Jr JF *et al.* Obesity, hypertension, and the risk of kidney cancer in men. *New England Journal of Medicine* 2000; **343**:1305–1311.
49. Melbye M, Wohlfahrt J, Lei U *et al.* A-fetoprotein levels in maternal serum during pregnancy and maternal breast cancer incidence. *Journal of the National Cancer Institute* 2000; **92**:111–115.
50. Frisch M, Biggar R, Engels E *et al.* Association of cancer with aids-related immunosuppression in adults. *Journal of the American Medical Association* 2001; **285**:1736–1745.
51. Cernan JR, Kushi LH, Olson JE *et al.* Twinship and risk of postmenopausal breast cancer. *Journal of the National Cancer Institute* 2001; **92**:261–266.
52. Hjalgrim H, Askling J, Srensen P *et al.* Risk of Hodgkin's disease and other cancers after infectious mononucleosis. *Journal of the National Cancer Institute* 2000; **92**:1522–1528.
53. Johansen C, Boice Jr JD, McLaughlin JK *et al.* Cellular telephones and cancer—a nationwide cohort study in Denmark. *Journal of the National Cancer Institute* 2001; **93**:203–205.
54. Lichtenstein P, Holm NV, Verkasalo PK *et al.* Environmental and heritable factors in the causation of cancer— analyses of cohorts of twins from Sweden, Denmark, & Finland. *New England Journal of Medicine* 2000; **343**:78–85.
55. Mork J, Lie AK, Glattre E *et al.* Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *New England Journal of Medicine* 2001; **344**:1125–1131.
56. Pala V, Krogh V, Muti P *et al.* Erythrocyte membrane fatty acids and subsequent breast cancer: a prospective Italian study. *Journal of the National Cancer Institute* 2001; **93**:1088–1095.
57. Signorello LB, Ye W, Fryzek J *et al.* Nationwide study of cancer risk among hip replacement patients in Sweden. *Journal of the National Cancer Institute* 2001; **93**:1405–1410.
58. Terry P, Giovannucci E, Michels KB *et al.* Fruit vegetables dietary fiber and risk of colorectal cancer. *Journal of the National Cancer Institute* 2001; **93**:525–529.
59. Tsubono Y, Nishino Y, Komatsu S *et al.* Green tea and the risk of gastric cancer in Japan. *New England Journal of Medicine* 2001; **344**:632–636.
60. Anttila T, Saikku P, Koskela P *et al.* Serotypes of *Chlamydia trachomatis* and risk for development of cervical squamous cell carcinoma. *Journal of the American Medical Association* 2001; **285**:47–51.
61. Biggar R, Frisch M, Goedert J. Risk of cancer in children with AIDS. AIDS-cancer match registry study group. *Journal of the American Medical Association* 2000; **284**:205–209.
62. Clopper C, Pearson E. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; **26**:404–413.
63. Alonzo T, Pope M. Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine* 1999; **18**:2987–3003.
64. Brewster, Crichton D, Muir J. How accurate are Scottish cancer registration data? *British Journal of Cancer* 1994; **70**:954–959.
65. Rawson NS, Robson DL. Concordance on the recording of cancer in the Saskatchewan Cancer Agency Registry, hospital charts & death registrations. *Canadian Journal of Public Health* 2000; **91**:390–393.
66. Cibere J, Sibley J, Haga M. Systemic lupus erythematosus and the risk of malignancy, Lupus. *Journal of Rheumatology* 2001; **10**:394–400.
67. Joseph L, Gyorkos T. Inferences for likelihood ratios in the absence of a gold standard. *Medical Decision Making* 1996; **16**:412–417.