Liu et al. BMC Medical Research Methodology 2014, 14:110
http://www.biomedcentral.com/1471-2288/14/110

BMC
Medical Research Methodology

RESEARCH ARTICLE Open Access

A two-stage Bayesian method for estimating accuracy and disease prevalence for two dependent dichotomous screening tests when the status of individuals who are negative on both tests is unverified

Jin Liu, Feng Chen*, Hao Yu, Ping Zeng and Liya Liu

Abstract

Background: Estimating the disease prevalence and test accuracy (sensitivity and specificity) for two dependent screening tests when the status of individuals who are negative on both tests is unverified represents a considerable challenge, as the disease rates for individuals negative on both tests are not identifiable without additional assumptions.

Methods: This article presents a unified framework for handling this non-identifiability problem using two-step hierarchical informative prior on the sensitivities by two-stage Bayesian modeling with the characterized by joint testing strategies based on the inherent attribute of screening/diagnostic tests. We assign a diffuse and less risky two-step hierarchical informative uniform prior to the sensitivities while assigning a uniform (0,1) prior distribution to the specificities and prevalence. Strategies for model evaluation, general global evaluations, and individual cell checking are presented. Simulations are conducted under various scenarios to evaluate the performance of the proposed method. Applications to real data are also presented to illustrate the potential impact and benefit of the proposed method.

Results: Our results indicate that when the priors of sensitivities are assigned as appropriate two-step hierarchical informative priors, or even in the absence of the priors for the specificities and prevalence, the parameters involved in this study can still be estimated well. The advantages and limitations of this method in solving such problems are discussed and compared with other two-stage methods.

Conclusions: We developed a two-stage Bayesian method for two dependent dichotomous screening tests with unverified individuals who are negative on both tests, and addressed the ad hoc model evaluation and checking procedures. The method can be understood easily and used conveniently by non-statisticians.

Keywords: Bayesian modeling, Screening test, Test accuracy, Disease prevalence, Model checking

Background

Screening programs for a specific disease or condition of interest are typically divided into two stages. In the first stage, a population of known size n is screened by simple and rapid but imperfect tests to classify people as likely or unlikely to have the disease or condition of interest. In the second stage, individuals who appear likely to have the disease or condition are examined further using a gold standard test to confirm the disease or condition. To improve the sensitivity of these screenings, two simple and rapid tests are often used in the first stage [1]. If either of the tests is positive, then a full evaluation of the correct disease status using a gold standard classification is undertaken in the second stage [2]. This strategy is widely used in screening for chronic diseases, infectious

* Correspondence: dr.chenfeng@163.com
Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, PR China

BioMed Central

© 2014 Liu et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

diseases, and animal diseases [2-4]. One characteristic of this strategy is that negatives in both tests are not verified by a gold standard test because the disease probability in the sub-population is so low that further investigations are costly, unacceptable, and possibly unethical. Another characteristic is that the two tests in the first stage are often conditionally dependent on disease status and/or non-disease status [5]. If the two screening tests have a similar biological basis, as is often the case, the conditional dependence assumption is suitable. For example, when we use two fecal occult blood tests (occult blood tests and immunohistochemical methods) for colon cancer screening, a positive occult blood test is often accompanied by a positive immunochemical result; thus, the results of the two tests are dependent [2].

Estimating the disease prevalence and test accuracy (sensitivity and specificity) when the two tests are dependent and when individuals negative on both tests are unverified is a considerable challenge, as the disease rate for individuals who are negative on both tests is not identifiable without additional assumptions [5-9]. Li et al. reviewed several methods for resolving the non-identifiability problem in the framework of classical statistics [8]. They classified these methods into two types. The methods of the first type are intended to solve the non-identifiability problem by introducing additional assumptions about the association of the two tests. Walter et al. assumed that the two test errors were independent [9], while van der Merwe also assumed that the two tests were independent under the disease status but dependent under non-disease status [10]. The methods in the second case resolve the non-identifiability problem directly by adding an assumption that is more likely to hold than the independence assumption [7]. Bohning and Patilea proposed a capture-recapture approach for screening using two diagnostic tests with information on disease status available for the test positives only [11]. Chu et al. proposed latent class models for screening studies using two screening tests, with a categorical disease status verified in positives test only [7]. However, these studies did not provide sufficient insight and details to solve the problem [8]. Li et al. presented a unified framework for solving the non-identifiability problem by saturating the model using certain homogeneous association assumptions [8]. In their paper, although the five distinct models provided different estimators, they were all saturated models and provided the same maximum likelihood. Therefore, they are not differentiable in terms of goodness-of-fit [8].

Recently, Bayesian analysis has been used increasingly to solve the non-identifiability problem using prior information [12-17]. Briefly, the use of this analysis is appropriate for three scenarios. The first involves the use of Bayesian model averaging (BMA) to address this challenging estimation problem, as a different homogeneous-dependent model can provide the same goodness-of-fit for the data but with substantively different estimates [7]. The second scenario involves the application of BMA to inferences over conditional independence and dependence models for the estimation of disease prevalence in a situation involving two imperfect tests in the absence of a gold standard [18]. The third scenario involves the administration of a moderately effective but relatively cheap screening test to all subjects in the first stage, after which a gold standard test is performed on a subset of the high screen and low screen groups in the second stage [19].

This article presents a unified framework to handle the non-identifiability problem by a two-stage Bayesian model using two-step hierarchical informative uniform prior on the sensitivities. We provide detailed Bayesian modeling for stages one and two, the specification of the prior distribution, and the calculation of the posteriors of the parameter distributions. The strategies for model evaluation and checking are proposed. We illustrate our approach using an example and simulation and conclude this article with a discussion and some remarks.

Methods

Two-stage Bayesian modeling

Data structure and estimated parameters

Table 1 presents the outcomes of two screening tests (T_1, T_2) from a total of n subjects, with disease statuses (D_+, D_-) verified only for individuals with at least one positive test. D_+ represents a true diseased condition, whereas D_- represents the non-diseased, as determined by a gold standard test. $T_{i(+,-)}$ ($i = 1,2$) represents the results of the test i. Let vector $x = (x_{11}, x_{10}, x_{01}, x_{00})$ denote the observed number for each combination of two test outcomes (1 or 0 for test positive or test negative). Let $SI_{(+,-)}$ (abbreviation of *Si*multaneous testing) represent the outcomes of the simultaneous testing in stage one, and let a_{11}, a_{10}, a_{01}, a_{00} denote D_+ SI_+, D_- SI_+, D_+ *SI-*, *D- SI-*, respectively. Because the full evaluation of the disease status is not performed if neither of the two tests is positive in the first stage, the frequencies of $[a_{01}]$ and $[a_{00}]$ in the bracket are unobserved, whereas the total frequency of both tests being negative $a_{0.} = [a_{01}] + [a_{00}]$ is known.

Table 1 Data structure of two-stage Bayesian modeling for two dependent dichotomous screening tests with individuals who are negative in both tests unverified

	Stage one				Stage two		
	T_{2+}	T_{2-}	Total		D+	D-	Total
T_{1+}	x_{11}	x_{10}	$x_{1.}$	SI_+	a_{11}	a_{10}	$a_{1.}$
T_{1-}	x_{01}	x_{00}	$x_{0.}$	SI_-	$[a_{01}]$	$[a_{00}]$	$a_{0.}$
Total	$x_{.1}$	$x_{.0}$	n	Total	$[a_{.1}]$	$[a_{.0}]$	n

Note: SI, simultaneous testing; [] indicates that the frequencies are unavailable.

The prevalence is defined as $\pi = P(D+)$. The sensitivity and specificity for the ith test are defined as $Se_i = P(T_{i+}|D_+)$ and $Sp_i = P(T_{i-}|D_-)$, respectively. Furthermore, let $CovDp$ and $CovDn$ be the covariances between the two tests for diseases ($CovDp$) and for non-diseases ($CovDn$). Let ρ_{D+} and ρ_{D-} be the correlation coefficients under the conditions of disease and non-disease, and let PPV_{JE} be the positive predictive value of the simultaneous testing. These parameters are defined as follows [12,20,21]:

$$PPV_{JE} = \frac{Se_{JE}.\pi}{Se_{JE}.\pi + \left(1 - Sp_{JE}\right)(1-\pi)} \tag{1}$$

where Se_{JE} and Sp_{JE} denote the joint sensitivity and specificity of the simultaneous testing (described in equation (6))

$$CovDp = Se_{11} - Se_1 Se_{2,},\ CovDn = Sp_{22} - Sp_1 Sp_2, \tag{2}$$

where $Se_{11} = P(T_{1+}, T_{2+}|D_+)$, $Sp_{22} = P(T_{1-}, T_{2-}|D_-)$

$$\rho_{D+} = \frac{CovDp}{\sqrt{Se_1(1-Se_1)Se_2(1-Se_2)}}, \quad \rho_{D-} = \frac{CovDn}{\sqrt{Sp_1(1-Sp_1)Sp_2(1-Sp_2)}} \tag{3}$$

Bayesian modeling for stage one

According to the screening programs, in the first stage (see Table 1, left column), two simple and rapid but potentially imperfect screening tests, denoted as T_1 and T_2, are applied simultaneously to each of the units sampled. If either individual test is positive, this condition indicates a positive result of the joint test. The strategy is often called simultaneous testing and seeks to obtain a joint testing strategy with increased sensitivity [22]. Let vector $Px = (Px_{11}, Px_{10}, Px_{01}, Px_{00})$ be the probabilities, respectively, from the vector $x = (x_{11}, x_{10}, x_{01}, x_{00})$. Supposing that n subjects are sampled randomly with prevalence π, the vector x has the following multinomial sampling distribution:

$$x \sim multinomial\ ((Px_{11}, Px_{10}, Px_{01}, Px_{00}), n) \tag{4}$$

Considering that the two tests are conditionally dependent under disease status and non-disease status, the multinomial cell probabilities are given by

$$\begin{aligned} Px_{11} &= P(T_{1+}, T_{2+}) = \pi(Se_1 Se_2 + CovDp) + (1-\pi)((1-Sp_1)(1-Sp_2) + CovDn), \\ Px_{10} &= P(T_{1+}, T_{2-}) = \pi(Se_1(1-Se_2) - CovDp) + (1-\pi)((1-Sp_1)Sp_2 - CovDn), \\ Px_{01} &= P(T_{1-}, T_{2+}) = \pi((1-Se_1)Se_2 - CovDp) + (1-\pi)(Sp_1(1-Sp_2) - CovDn), \\ Px_{00} &= P(T_{1-}, T_{2-}) = \pi((1-Se_1)(1-Se_2) + CovDp) + (1-\pi)(Sp_1 Sp_2 + CovDn). \end{aligned} \tag{5}$$

In stage one, the individuals are classified as likely to have the disease or condition of interest only if either of the two test outcomes is positive. The joint sensitivity and specificity of the simultaneous testing (denoted by the subscript JE, for 'Joint Either') are as follows:

$$\begin{aligned} Se_{\mathrm{JE}} &= P\left(T_1^+ \cup T_2^+ | D^+\right) = Se_1 + Se_2 - Se_1 Se_2 - CovDp \\ Sp_{\mathrm{JE}} &= P\left(T_1^- \cap T_2^- | D^-\right) = Sp_1 Sp_2 + CovDn. \end{aligned} \tag{6}$$

Bayesian modeling for stage two

In the second stage, a gold standard test is used only for positive samples from the simultaneous testing of the first stage, whereas the negative samples are not verified by the gold standard test (see the right-hand column in Table 1). Obviously, in this case, the simultaneous testing of the first stage and the gold standard testing of the second stage constitute sequential testing [22]. This sequential testing is used mainly to increase feasibility because applying the gold standard test to the negatives is often costly, unacceptable, and unethical. Given that the gold standard test and the simultaneous testing are independent and that the sensitivity and specificity of the gold standard test are equal to 100%, the joint sensitivity and specificity of the sequential testing (denoted by the subscript JB, for 'Joint Both') are as follows:

$$\begin{aligned} Se_{JB} &= P\left(T_{JE}^+ \cap T_{GS}^+ | D^+\right) = Se_{JE} Se_{GS} = Se_{JE} \\ Sp_{JB} &= P\left(T_{JE}^- \cup T_{GS}^- | D^-\right) = Sp_{JE} + Sp_{GS} - Sp_{JE} Sp_{GS} = 1 \end{aligned} \tag{7}$$

where Se_{GS} and Sp_{GS} are the sensitivity and specificity of the gold standard (denoted by the subscript GS, for '*G*old *S*tandard'), respectively.

Let vector $a = (a_{11}, a_{10}, a_{0.})$ denote the observed number for each combination in stage two, where $a_{0.}$ is the sum of the test negatives from the first stage. Let $Pa = (Pa_{11}, Pa_{10}, Pa_{0.})$ represent the respective probabilities from vector a. Consider that elements from the proportion vector of multinomial a are required the probabilities to sum to 1. The multinomial sampling distribution is given by

$$a \sim multinomial\ ((Pa_{11}, Pa_{10}, Pa_{0.}), n) \tag{8}$$

where the multinomial cell probabilities are given by

$$\begin{aligned} Pa_{11} &= \pi Se_{JE}Se_{GS} + (1-\pi)(1-Sp_{JE})(1-Sp_{GS}) = \pi Se_{JE} \\ Pa_{10} &= \pi Se_{JE}(1-Se_{GS}) + (1-\pi)(1-Sp_{JE})Sp_{GS} = (1-\pi)(1-Sp_{JE}) \\ Pa_{0\cdot\cdot} &= 1-Pa_{11}-Pa_{10} \end{aligned} \tag{9}$$

Prior distributions

For Bayesian statistics, we wish to use all available information at the design stage but might prefer a more vague, less risky prior at the data analysis stage [23]. To reduce the influence of subjective opinions on the data analysis as much as possible, we placed an informative prior distribution on a minimum number of parameters [24].

In principle, a uniform prior distribution or Beta prior distribution can be used over the set of unknown parameters. We chose the uniform non-Beta prior distribution because a uniform prior distribution is more easily accessible for non-statisticians and less risky in this study. For example, to determine the value of α and β for a Beta prior distribution, we need to know the corresponding mean and standard deviation or the 2.5 and 97.5 percentiles [25], whereas for a uniform prior distribution, we need only to determine the range of estimated parameters. Especially for the main parameters (Se_1, Se_2, Sp_1, Sp_2) in this study, lower bound values for the uniform distribution are often estimated securely based on expert opinion, published papers, or even test kit instructions. For example, an epidemiologist can easily be assured that the sensitivity of the enzyme-linked immunosorbent assay (ELISA) for HIV antibody screening is higher than 80% [4]; i.e., the lower bound of the uniform distribution is 0.8. However, a reasonable choice for the upper bound of a uniform prior distribution is sometimes difficult to determine because the accuracy of the test kit might differ under various practical conditions. We use the method of two-step hierarchical priors to set the upper bound of the uniform distribution. For an estimated parameter, such as sensitivity (Se_i), we first provide a uniform prior distribution (a_{Sei}, b_{Sei}) and then give b_{Sei} another uniform distribution (b_{1Sei}, b_{2Sei}). The method of two-step hierarchical priors is a good strategy if it is difficult to determine the prior distribution using a one-step method; even though an incorrect prior is set in step two, the risk of this action resulting in a mistaken result is smaller than in a one-step prior [26].

For sensitivity, we set two steps for the hierarchical uniform priors:

$$Se_i \sim uniform\ (a_{Sei};\ b_{Sei}), b_{Sei} \sim uniform\ (b_{1Sei};\ b_{2Sei}). \tag{10}$$

For specificities and prevalence, we might set a uniform prior distribution based on the characteristics of this type of study (see the Discussion section):

$$Sp_i \sim uniform\ \left(a_{Spi}; b_{Spi}\right), \pi \sim uniform\ (a_\pi; b_\pi). \tag{11}$$

For covariances *CovDp and CovDn,* the feasible range is determined by the sensitivities among the diseased subjects and the specificities among the non-diseased subjects, where $0 \le CovDp \le \min\ (Se_1,Se_2)Se_1Se_2$ for the diseased subjects and $0 \le CovDn \le \min(Sp_1,Sp_2)Sp_1Sp_2$ for the non-diseased subjects [20]. Because prior information regarding the two covariances is typically unavailable, uniform prior distributions over these ranges can be used for *CovDp* and *CovDn*:

$$\begin{aligned} &CovDp \sim uniform\ (0,\ min\ (Se_1, Se_2)Se_1Se_2), \\ &CovDn \sim uniform\ (0, min(Sp_1, Sp_2)Sp_1Sp_2. \end{aligned} \tag{12}$$

Calculation of the posteriors of parameter distributions

The posteriors of parameter distributions were calculated using Markov-chain Monte Carlo techniques, in particular the Gibbs sampler in WinBUGS (MRC Biostatistics Unit, Cambridge, UK) [27]. For the analyses presented in this paper, inferences were based on 105,000 iterations after discarding an initial burn-in of 5,000 iterations, with convergence assessed by running multiple chains from various starting values [28]. The WinBUGS code used in this paper is available in the Additional files 1 and 2 and can be altered easily for use with different data.

Model evaluation and checking

Model evaluation and checking are highly active areas of Bayesian statistics research. Researchers can use various statistics to determine the plausibility of an assumption of interest in light of the observed data [29]. In this study, we divide the model evaluation and the checking of the assumptions into individual and overall diagnostics. Individual checks are based on the cell in Table 1, and overall diagnostics aim to check the more general assumptions of the model by *DIC* [21], *pD* [5], and the Chi-squared goodness-of-fit test. The technical details were given in Additional file 3. Given the special circumstance that the negatives for both tests are unavailable, none of the methods could be used alone in the preceding model checks because they diagnosed the models from a different perspective. For example, a local Chi-squared goodness-of-fit test is based only on known cells. For this reason, it is recommended that the above methods be combined for model checking.

Below, we present an ad hoc preliminary, exploratory model that serves to check the criteria, based on empirical information for this type of study. If (1) the values of *DIC* and *pD* are reasonable [5]; (2) the *p*-value of the local χ^2 test is close to 0.5 or far from zero or one [29]; (3) the 95% *BCI* of the estimated *PPV* (model key point)

includes its actual value; and (4) most frequencies of the cells in Table 1 fall within their corresponding Bayesian credible intervals, we may conclude that the model fits the data well based on the current information and at a specified probability level.

Results and discussion

Results

Screening study for colorectal cancers

The data in Table 2 are based on the data of Castiglione et al. [30]. That study compared rehydrated guaiac testing (Hemoccult) on three consecutive bowel movements with immunochemical testing using reversed passive haemagglutination (RPHA—Hemselect) on the first bowel movement only to detect colorectal cancers or adenomas ≥ 10 mm in 5727 individuals aged 40–59 years. Subjects with a positive hemoccult and/or a positive/borderline hemeselect test were invited to undergo pancolonoscopy. A double-contrast barium enema was performed when pancolonoscopy was not possible. Further details of the methods used in the study were provided by Castiglione et al. [2], and the data were analyzed by Geoffrey Berry et al [6].

According to the description in of *Prior distribution* section of this paper, we placed hierarchical uniform priors on Se_1 and on Se_2. These hierarchical priors were elicited based on the opinions of a co-author (Dr. Hao Yu) and on published papers [2,30]. The expert opinion and previous information showed that the sensitivity of test 1 (rehydrated hemoccult) was less than the sensitivity of test 2 (RPHA-Hemeselect) and that the values of their lower bounds values were at least 0.5 and 0.6, respectively. The values of the upper bounds for Se_1 and Se_2 were assigned as much diffuse, less risky two-step hierarchical priors, as follows:

$$Se_1 \sim uniform\ (0.5, b_{Se1}), b_{Se1} \sim uniform\ (0.55,\ 0.99),$$
$$Se_2 \sim uniform\ (0.6, b_{Se2}), b_{Se2} \sim uniform\ (0.65,\ 0.99).$$

For the specificities and prevalence, we set the following uniform prior:

$$Sp_1 \sim uniform\ (0,1), Sp_2 \sim uniform\ (0,1), \pi \sim uniform\ (0,1).$$

Table 2 Observed frequencies in the screening study for colorectal cancers in 5727 individuals aged 40–59 years [6]

	Stage one				Stage two		
	T_{1+}	T_{1-}	Total		D+	D-	Total
T_{2+}	39	91	130	SI_+	29	338	367
T_{2-}	237	5360	5597	SI_-	[a_{01}]	[a_{00}]	5360
Total	276	5451	5727	Total	[$a_{.1}$]	[$a_{.0}$]	5727

Note: T_1, rehydrated Hemoccult; T_2, RPHA-Hemeselect; SI, simultaneous testing; [] indicates that the frequencies are unavailable.

For the covariances *CovDp* and *CovDn*, the uniform priors were assigned according to equation (12).

It is clear from Table 3 that the results from the two-stage Bayesian model with two-step hierarchical prior on Se_1 and Se_2 fit the data well in accordance with the criteria presented in the *Model evaluation and checking* section. All of the estimated values were close to their true values, and all fell within 95% Bayesian credible intervals over their corresponding true values. The effective number of parameters, a *pD* of approximately 3.9, indicates that the two-stage Bayesian strategy substantially improves the identifiability of the models. The $\chi^2 = 0.002$ and $P = 0.61$ values indicate that the distributions of the replicated and actual data are similar.

Table 3 shows the sensitivity and specificity of joint testing to be approximately 0.81 and 0.94, respectively, with the strategy of joint-simultaneous testing adding approximately 0.13 (Se_{JE} (0.8103) minus Max (Se_1, Se_2) (0.6825)) to the sensitivity at the expense of only an approximate loss of 0.04 (Sp_{JE} (0.9405) minus $Max(Sp_1, Sp_2)$ (0.9812)) in specificity for the colorectal cancer data. This result conformed to the theory of joint-simultaneous testing. The correlation coefficients under disease and non-disease status, which were approximately 0.4 and 0.1, respectively, suggested moderate and weak dependence under the conditions of disease and non-disease, respectively.

To analyze the influence of the prior distribution of sensitivities (Se_i, ($i = 1,2$)) on posterior distributions, the non-informative prior, pessimistic and optimistic two-step hierarchical informative priors were assigned, respectively, as below:

$$Se_i \sim uniform\ (0,\ 1);$$
$$Se_i \sim uniform\ (0.1,\ uniform\ (0.2,\ 0.99));$$
$$Se_i \sim uniform\ (0.8,\ uniform\ (0.9,\ 0.99)).$$

Table 3 Posterior parameters and models checking for the application of two-stage Bayesian models to the colorectal cancers data

Posterior parameters		Model evaluation and checking		
Median (95% BCI)		True value Median (95% BCI)		
π	0.0065 (0.0043, 0.0095)	PPV_{JE}	0.08	0.0807 (0.0559, 0.1107)
Se_1	0.6040 (0.5048, 0.8834)	x_{11}	39	39 (24, 58)
Se_2	0.6825 (0.6039, 0.9036)	x_{10}	91	91 (67, 117)
Sp_1	0.9812 (0.9772, 0.9849)	x_{01}	237	237 (200, 277)
Sp_2	0.9559 (0.9513, 0.9603)	x_{00}	5360	5359 (5314, 5402)
Se_{JE}	0.8103 (0.6503, 0.9557)	a_{11}	29	30 (16, 46)
Sp_{JE}	0.9405 (0.9360, 0.9450)	a_{10}	338	338 (296, 382)
			pD = 3.9	DIC = 39.7
			$\chi^2 = 0.0002$	P = 0.6101

For the specificities and prevalence, the following uniform priors were assigned:

$$Sp_i \sim uniform\ (0,1), \pi \sim uniform\ (0,1).$$

Table 4 shows that prior information of the sensitivities produced substantive influence on the posterior distributions. The hierarchical informative priors on the sensitivities must be assigned reasonably in these types of studies.

Berry et al. analyzed the data using maximum likelihood by fitting the four versions of the models: (1) the independence model, (2) independence in the non-diseased group, (3) independence in the diseased group, and (4) the dependence model [6]. Based on the χ^2 test statistics, our models provide a better fit to the data than those in the Berry et al. paper. In addition, our method can estimate parameter intervals that were not reported in the Berry et al. paper. In fact, it is difficult for the method of Berry et al. to calculate the confidence intervals of certain complex statistics involved in joint simultaneous testing, such as the joint positive predictive value (PPV_{JE}), joint sensitivity (Se_{JE}), and joint specificity (Sp_{JE}), which are important in evaluating model fitting in this type of study.

Simulation studies

To further illustrate the performance of the two-stage Bayesian models, we applied them to a series of simulated data sets with a conditional correlation coefficient of $\rho_{D+} = 0.5$ for the diseased and $\rho_{D-} = 0.4$ for the non-diseased, as would typically be found in practice. Because our models contain a large number of parameters and because the study design has been associated with a wide range of prevalence values, sample sizes, and test properties, it is impossible to investigate the performance of these models across all possible scenarios. We therefore selected a range related to the actual screening and diagnostic tests, based on the following parameters:

(1) low prevalence and large sample size, such as cancer screening for community population [30] and HIV-antibody screening for blood donors [31].
  (1.1) high test accuracy, $(\pi, n, (Se_1, Sp_1), (Se_2, Sp_2)) = (0.01, 20000, (0.90, 0.95), (0.95, 0.90))$;
  (1.2) low test accuracy, $(\pi, n, (Se_1, Sp_1), (Se_2, Sp_2)) = (0.01, 20000, (0.60, 0.70), (0.70, 0.60))$.
(2) high prevalence and small sample size, such as the screening of suspicious patients in the hospital or clinic.
  (2.1) high test accuracy, $(\pi, n, (Se_1, Sp_1), (Se_2, Sp_2)) = (0.40, 200, (0.90, 0.95), (0.95, 0.90))$;
  (2.2) low test accuracy, $(\pi, n, (Se_1, Sp_1), (Se_2, Sp_2)) = (0.40, 200, (0.60, 0.70), (0.70, 0.60))$

We also examined the consequences of specifying different prior information, which were divided into the four scenarios (Table 5) given below:

(1) informative priors for model: the prior density of the estimated model parameters (Se_1, Se_2, Sp_1, Sp_2, and π) was centered at their true values.
(2) informative priors for sensitivities alone: the prior density of the sensitivities alone was centered at their true values, whereas the specificities and prevalence were assigned as non-informative priors.
(3) two-step informative priors for sensitivities alone: the priors of the sensitivities were assigned according to a two-step hierarchical prior with their upper bounds assigned as much vaguer and less risky priors, as described in the *Prior distribution* section.
(4) informative priors for specificities alone: the prior density of the specificities alone was centered at their true values, whereas the sensitivities and prevalence were assigned as non-informative priors.

The results of applying our models to the simulated data sets are given in Tables 6 and 7.

Table 4 Influence of the prior distributions of sensitivities on posterior distributions

	$Se_i \sim U(0, 1)$	$Se_i \sim U(0.1, U(0.2, 0.99))$	$Se_i \sim U(0.8, U(0.9, 0.99))$
	Median (95% BCI)	**Median (95% BCI)**	**Median (95% BCI)**
π	0.0102 (0.0044, 0.1793)	0.0130 (0.0056, 0.0323)	0.0055 (0.0038, 0.0075)
Se_1	0.3053 (0.0075, 0.9666)	0.2461 (0.1065, 0.1065)	0.8691 (0.803, 0.9612)
Se_2	0.2658 (0.0055, 0.9397)	0.2414 (0.1057, 0.7268)	0.8675 (0.8032, 0.9596)
Sp_1	0.9805 (0.9740, 0.9851)	0.9806 (0.9762, 0.9848)	0.9818 (0.9781, 0.9853)
Sp_2	0.9540 (0.9434, 0.9595)	0.9545 (0.9494, 0.9594)	0.9561 (0.9516, 0.9604)
Se_{JE}	0.5162 (0.0293, 0.9861)	0.3996 (0.1703, 0.8295)	0.9466 (0.8561, 0.9899)
Sp_{JE}	0.9398 (0.9276, 0.9445)	0.9401 (0.9354, 0.9446)	0.9405 (0.9360, 0.9449)

Note: U, uniform distribution; BCI, Bayesian credible interval.

Table 5 Priors of two-stage Bayesian modeling for simulation studies

True value	Informative priors for model	Informative priors for Se alone	Two-step informative priors for Se alone	Informative priors for Sp alone
Se = 0.6	U(0.5,0.7)	U(0.5,0.7)	U(0.5,U(0.55,0.99))	U(0,1)
Se = 0.7	U(0.6,0.8)	U(0.6,0.8)	U(0.6,U(0.65,0.99))	U(0,1)
Se = 0.8	U(0.7,0.9)	U(0.7,0.9)	U(0.7,U(0.75,0.99))	U(0,1)
Se = 0.9	U(0.85,0.95)	U(0.85,0.95)	U(0.85,U(0.9,0.99))	U(0,1)
Se = 0.95	U(0.90,0.99)	U(0.90,0.99)	U(0.9,U(0.95,0.99))	U(0,1)
Sp = 0.6	U(0.5,0.7)	U(0,1)	U(0,1)	U(0.5,0.7)
Sp = 0.7	U(0.6,0.8)	U(0,1)	U(0,1)	U(0.6,0.8)
Sp = 0.8	U(0.7,0.9)	U(0,1)	U(0,1)	U(0.7,0.9)
Sp = 0.9	U(0.85,0.95)	U(0,1)	U(0,1)	U(0.85,0.95)
Sp = 0.95	U(0.90,0.99)	U(0,1)	U(0,1)	U(0.90,0.99)
$\pi = 0.4$	U(0.3,0.5)	U(0,1)	U(0,1)	U(0,1)
$\pi = 0.1$	U(0.05,0.15)	U(0,1)	U(0,1)	U(0,1)
$\pi = 0.01$	U(0.005,0.015)	U(0,1)	U(0,1)	U(0,1)

Note: π, prevalence; U, uniform distribution.

Here, we would like to make the following general observations:

(1) It is clear that, as long as the priors of the sensitivities were correctly set, the models with informative priors for the model, informative priors for the sensitivities alone, and two- step informative priors for the sensitivities alone produced similar results, all of which were centered over the true values. This finding suggests that the prior distributions of the prevalence and specificities could be reasonably assigned as uniform (0, 1) in these types of studies.

(2) For models with informative priors for the specificities alone, the estimates of the prevalence and sensitivities were biased, and the corresponding credible intervals were very wide, suggesting that the precision of the estimated parameters was low. The negative *pD*-values meant that informative priors for the specificities alone are insufficient to estimate the parameters. Thus, the Bayesian models are non-identifiable from the perspective of probabilistic constraints [5]. The *DIC* was decreased because of the negative *pD*.

(3) The parameter specificities were always estimated with greater precision than the sensitivities when

Table 6 Posterior medians and 95 per cent posterior credible intervals for the application of the models to simulated data, with low prevalence (π = 0.01) and large sample size (n = 20000)

True value	Informative priors for model	Informative priors for Se alone	Two-step informative priors for Se alone	Informative priors for Sp alone
$\pi = 0.01$	0.0094 (0.0079,0.0120)	0.0095 (0.0080,0.0119)	0.0093 (0.0075,0.0118)	0.0115 (0.0072,0.0590)
$Se_1 = 0.6$	0.6013 (0.5035,0.6944)	0.5971 (0.5063,0.6936)	0.5972 (0.5045,0.8695)	0.4181 (0.0195,0.9603)
$Se_2 = 0.7$	0.6998 (0.6047,0.7949)	0.6904 (0.6036,0.7940)	0.6816 (0.6040,0.9019)	0.3656 (0.0177,0.9636)
$Sp_1 = 0.7$	0.7000 (0.6941,0.7052)	0.6998 (0.6947,0.7054)	0.700 (0.6945,0.7054)	0.6982 (0.6824,0.7057)
$Sp_2 = 0.6$	0.5996 (0.5951,0.6045)	0.5999 (0.5948,0.6049)	0.5998 (0.5948,0.6048)	0.5964 (0.5762,0.6044)
pD	3.99	3.93	3.87	-14.72
DIC	53.84	53.74	53.63	35.09
$\pi = 0.01$	0.0099 (0.0089, 0.0110)	0.0099 (0.0089, 0.0110)	0.0099 (0.0090, 0.0111)	0.0133 (0.0093, 0.0377)
$Se_1 = 0.90$	0.8996 (0.8526, 0.9479)	0.9019 (0.8545, 0.9479)	0.8940 (0.8528, 0.9678)	0.5017 (0.0321, 0.9563)
$Se_2 = 0.95$	0.9445 (0.9020, 0.9876)	0.9429 (0.9031, 0.9875)	0.9341 (0.9015, 0.9759)	0.5130 (0.0720, 0.9736)
$Sp_1 = 0.95$	0.9500 (0.9472, 0.9528)	0.9499 (0.9469, 0.9528)	0.9499 (0.9469, 0.9527)	0.9480 (0.9404, 0.9522)
$Sp_2 = 0.90$	0.8998 (0.8969, 0.9027)	0.8998 (0.8969, 0.9028)	0.8997 (0.8967, 0.9027)	0.8975 (0.8906, 0.9018)
pD	3.82	4.01	4.17	-2.41
DIC	65.59	65.95	66.28	59.47

Table 7 Posterior medians and 95 per cent posterior credible intervals for the application of the models to simulated data, with high prevalence (π = 0.4) and small sample size (n = 200)

True value	Informative priors for model	Informative priors for Se alone	Two-step informative priors for Se alone	Informative priors for Sp alone
$\pi = 0.40$	0.3916 (0.3186, 0.4831)	0.3972 (0.3160, 0.5015)	0.3905 (0.3018, 0.4997)	0.3627 (0.2813, 0.5415)
$Se_1 = 0.6$	0.6029 (0.5051, 0.6946)	0.6054 (0.5059, 0.6953)	0.6004 (0.5077, 0.8582)	0.6220 (0.3960, 0.9331)
$Se_2 = 0.7$	0.6943 (0.6052, 0.7920)	0.6928 (0.6041, 0.7944)	0.6833 (0.6041, 0.9111)	0.7441 (0.5105, 0.9790)
$Sp_1 = 0.7$	0.7009 (0.6106, 0.7915)	0.7073 (0.5872, 0.8308)	0.7093 (0.5799, 0.8478)	0.7087 (0.6046, 0.7958)
$Sp_2 = 0.6$	0.5993 (0.5167, 0.6826)	0.5933 (0.5134, 0.6900)	0.5970 (0.5033, 0.7004)	0.6085 (0.5092, 0.6928)
pD	3.40	3.59	3.50	3.05
DIC	44.18	44.68	44.66	44.07
$\pi = 0.40$	0.3969 (0.3556,0.4395)	0.3971 (0.3499,0.4428)	0.3999 (0.3561,0.4462)	0.4209 (0.3588,0.6559)
$Se_1 = 0.90$	0.9039 (0.8531,0.9477)	0.9011 (0.8525,0.9479)	0.8933 (0.8517,0.9581)	0.8443 (0.5371,0.9874)
$Se_2 = 0.95$	0.9515 (0.9041,0.9879)	0.9531 (0.9035,0.9884)	0.9380 (0.9022,0.9808)	0.9022 (0.5775,0.9947)
$Sp_1 = 0.95$	0.9549 (0.9069,0.9887)	0.9548 (0.8822,0.9968)	0.9570 (0.8863,0.9978)	0.9537 (0.9038,0.9886)
$Sp_2 = 0.90$	0.9039 (0.8610,0.9415)	0.9024 (0.8459,0.9486)	0.9021(0.8496,0.9478)	0.8977 (0.8548,0.9376)
pD	2.69	3.28	3.13	0.81
DIC	29.26	30.41	30.25	27.99

the prevalence was low. Because the specificities relate to the negative subjects, when the prevalence is low, there are more truly negative subjects who provide more experimental information for specificities than for sensitivities; thus, the more precise posteriori estimates of specificities are obtained.

All of the above results indicate that prior information on the sensitivities plays a key role in these types of studies. We consider one of the reasons might be that the prevalence was low, because in absence of enough positive individuals, the prior distributions of sensitivities produced substantive influence on the posterior distribution. The results from both practical case (Tables 3 and 4) and simulation study (Tables 6 and 7) supported the consideration.

As described in the Background section, one of characteristics of two stage screening tests is that the two tests in the first stage are often conditionally dependent on disease status and/or non-disease status [5]. To further study how the prevalence and test accuracy estimates vary with the model assumptions (dependent or independent) and to evaluate the impact of the misspecification of different models on the estimation of the prevalence and test accuracy, we performed four sets of simulations under the assumptions of the independent model for independent data, the independent model for dependent data, the dependent model for independent data, and the dependent model for dependent data. For ease of presentation and interpretation, we considered moderate parameter values for the prevalence, sample size, and test accuracy, as shown below:

$$(\pi, n, (Se_1, Sp_1), (Se_2, Sp_2)) = (0.10, 2000, (0.70, 0.80), (0.80, 0.70)).$$

Table 8 shows that the independent models might generate incorrect statistical inferences for the conditional dependent data and larger *DIC* (323.08) values, indicating that the model was not appropriate for the data. For conditional independent data, the independent models and dependent models had similar results closer to their true values. Thus, we suggest that, when the researcher is uncertain as to whether the data are dependent or independent, the less risky dependent model should be used despite the slight loss of precision among the estimated parameters (widened credible interval).

Discussion

This article presents additional methods for the estimation problem for screening studies using two screening tests with the disease statuses being verified for test positives only. The core of the methods is that Bayesian modeling is divided into two stages characterized by a joint testing strategy based on the inherent attributes of this type of screening. The first stage is a joint simultaneous testing consisting of two dependent binary tests. The second stage is a special joint sequential testing that consists of simultaneous testing in the first stage and a gold standard test in the second.

Another method presented in this study involves assigning a diffuse and less risky two-step hierarchical

Table 8 Posterior medians and 95 per cent posterior credible intervals for the application of the dependent and independent models to simulated dependent and independent data (two-step informative priors for sensitivities alone were assigned (see Table 5))

	Dependent data		Independent data	
	Dependent model	Independent model	Dependent model	Independent model
$\pi = 0.10$	0.0963 (0.0830,0.1151)	0.1184 (0.1088,0.1277)*	0.1030 (0.0919,0.1162)	0.1000 (0.0900,0.1107)
$Se_1 = 0.70$	0.6985 (0.6048,0.7960)	0.7971 (0.7841,0.7999)*	0.6558 (0.6028,0.7847)	0.6979 (0.6068,0.7940)
$Se_2 = 0.80$	0.7935 (0.7044,0.8950)	0.8965 (0.8818,0.8999)*	0.7864 (0.7030,0.8942)	0.7994 (0.7057,0.8942)
$Sp_1 = 0.80$	0.7978 (0.7775,0.8196)	0.8866 (0.8700,0.9028)*	0.7948 (0.7740,0.8156)	0.8002 (0.7801,0.8205)
$Sp_2 = 0.70$	0.6981 (0.6792,0.7164)	0.7677 (0.7511,0.7846)*	0.6966 (0.6764,0.7170)	0.6992 (0.6797,0.7191)
pD	3.87	3.08	3.87	3.38
DIC	58.63	323.08	58.63	58.76

*: incorrect statistical inference.

prior structure for sensitivities. Because reasonably informative prior information is necessary for resolving the non-identifiability problem in this type of study, we emphasized the setting of the sensitivity prior distribution based on the following three points. First, the priors of the sensitivity are more important than the priors of the specificity or prevalence when the prevalence is lower than 50%, as is often found in this type of study. Second, we placed the prior information of the sensitivity on the least possible quantity, i.e., diffuse prior or less risky prior information, such as two-step hierarchical prior structure, to reduce the influence of subjective opinion as much as possible on the results. Third, we set the prior structure to be as easy as possible for use by non-statisticians so that uniform prior distributions could be used in this study.

We found that when the priors of the sensitivities were assigned as appropriate priors, such as two-step hierarchical uniform priors, in the absence of priors for the specificities and prevalence (e.g., assigning a uniform distribution on (0,1)), the parameters involved in this study could still be estimated well. However, changing the prior structure of the sensitivities has a substantive effect on the estimators (Tables 4, 6 and 7). The reasons for these results are presented in the *Simulation studies* section. Here, we would like to state that this finding has important practical significance because the sensitivities (related to the diseased population) can often, but not always, be obtained more easily than the specificities (related to the healthy population) in medical practice. This characteristic is partially based on the increased feasibility of applying the gold standard test to the diseased population compared with the healthy population, owing to the frequent invasiveness, expense, and sophistication of many gold standard tests, particularly ones with further ethical implications. Therefore, we suggest that the prior structure of the sensitivities should be thoroughly investigated for scenarios in low-prevalence populations.

To our knowledge, this study is the first to estimate test accuracy and disease prevalence for two dependent dichotomous screening tests with unverified negative individuals on both tests using two-stage Bayesian modeling. Recently, Bohning and Patilea, Chu et al., and Li et al. proposed two-stage methods to address the non-identifiability problem [7,8,11]. Their two-stage estimation methods differ from ours. In the first stage, Bohning and Patilea [11], Chu et al., and Li et al. [7,8] estimated the parameters of known cells; in the second stage, they estimated the parameters of unknown cells by making assumptions regarding the cell probabilities of various homogeneous association models. They circumvented the non-identifiability problem using a capture–recapture approach or constrained maximum likelihood estimation. Because all of the models are saturated, homogeneous association assumptions are not testable. Therefore, the dependence structure modeling considered can only be viewed as a sensitivity analysis [7]. They did suggest that Bayesian methods that incorporate prior information could be a reasonable alternative to obtain improved estimates, although further research is needed [8]. Their suggestion was an important factor in initiating our study.

Our two-stage approach has certain advantages over the previous methods. First, our method can be understood more easily and used with greater convenience by non-statisticians because the procedure of constructing the model fully agrees with the practical screening procedure. Additionally, the modeling often only involves the prior distribution for the sensitivities, because the prior distributions of the specificities and prevalence are assigned a non-informative uniform distribution of (0, 1), respectively. Moreover, the method of model checking is simple, intuitive, and convenient, as *pD* and *DIC* can be obtained directly using WinBUGS. Second, this method improves non-identifiability more reasonably, as probabilistic constraints not deterministic constraints (simplifying the model) are used. Third, the method could obtain more

information to evaluate the screening strategy than the capture–recapture approach or constrained maximum likelihood estimation [7,8,11]. For example, the method could conveniently calculate the joint sensitivities and joint specificities and make corresponding statistical inferences, such as 95% credible intervals (Table 3). In the framework of frequency statistics, such as capture–recapture approaches or constrained maximum likelihood estimations, it is often difficult to obtain statistical inferences about the joint sensitivities and joint specificities in these types of studies.

Conclusions

We developed a two-stage Bayesian method for two dependent dichotomous screening tests with unverified individuals who are negative on both tests. We also addressed the ad hoc model evaluation and checking procedures based on empirical information for these types of studies. Although the practical example and simulation studies considering various practical situations showed that these models fit well, three points should be clarified. (1) Due to the lack of full data, it is impossible to check the model completely without further information. (2) Although our Bayesian two-stage modeling improves non-identifiability based on probabilistic constraints, the informative prior information on the sensitivities must be assigned reasonably, even though the prior could be diffuse, as in the two-step hierarchical uniform prior. (3) Using our real example and simulated scenarios related to real-life situations, the priors of the sensitivities are more important than the priors of the specificities. However, when the prevalence is large, especially if it is greater than 50% (rare in medical screening tests), the same attention should be paid to the priors of the specificities.

Additional files

Additional file 1: The WinBUGS code, data and results.

Additional file 2: Figure S1. The convergence diagnosis.

Additional file 3: The technical details of model evaluation and checking.

Abbreviations

BCI: Bayesian credible interval; DIC: Deviance Information Criterion; pD: Effective number of parameters; PPV: Positive predictive value; SI: Simultaneous testing.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LJ and CF conceived and conducted the study, and drafted the manuscript. YH participated in the design, data generation and interpretation of the analyses. ZP and LL participated in the simulation and interpretation of the analyses. All the authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [grant number 81273184].

Received: 7 July 2014 Accepted: 11 September 2014
Published: 23 September 2014

References

1. Macaskill P, Walter SD, Irwig L, Franco EL: **Assessing the gain in diagnostic performance when combining two diagnostic tests.** *Stat Med* 2002, **21**(17):2527–2546.
2. Castiglione G, Zappa M, Grazzini G, Mazzotta A, Biagini M, Salvadori P, Ciatto S: **Immunochemical vs guaiac faecal occult blood tests in a population-based screening programme for colorectal cancer.** *Brit J Cancer* 1996, **74**(1):141–144.
3. Branscum AJ, Gardner IA, Johnson WO: **Bayesian modeling of animal- and herd-level prevalences.** *Prev Vet Med* 2004, **66**(1–4):101–112.
4. Liu P, Xiao S, Shi ZX, Bi XX, Yang HT, Jin H: **Bayesian evaluation of the human immunodeficiency virus antibody screening strategy of duplicate enzyme-linked immunosorbent assay in Xuzhou Blood Center, China.** *Transfusion* 2011, **51**(4):793–798.
5. Berkvens D, Speybroeck N, Praet N, Adel A, Lesaffre E: **Estimating disease prevalence in a Bayesian framework using probabilistic constraints.** *Epidemiology* 2006, **17**(2):145–153.
6. Berry G, Smith CL, Macaskill P, Irwig L: **Analytic methods for comparing two dichotomous screening or diagnostic tests applied to two populations of differing disease prevalence when individuals negative on both tests are unverified.** *Stat Med* 2002, **21**(6):853–862.
7. Chu H, Zhou Y, Cole SR, Ibrahim JG: **On the estimation of disease prevalence by latent class models for screening studies using two screening tests with categorical disease status verified in test positives only.** *Stat Med* 2010, **29**(11):1206–1218.
8. Li F, Chu H, Nie L: **A two-stage estimation for screening studies using two diagnostic tests with binary disease status verified in test positives only.** *Stat Met Med Res* 2011. Epub ahead of print.
9. Walter SD: **Estimation of test sensitivity and specificity when disease confirmation is limited to positive results.** *Epidemiology* 1999, **10**(1):67–72.
10. van der Merwe L, Maritz JS: **Estimating the conditional false-positive rate for semi-latent data.** *Epidemiology* 2002, **13**(4):424–430.
11. Bohning D, Patilea V: **A capture-recapture approach for screening using two diagnostic tests with availability of disease status for the test positives only.** *J Am Stat Assoc* 2008, **103**(481):212–221.
12. Branscum AJ, Gardner IA, Johnson WO: **Estimation of diagnostic test sensitivity and specificity through Bayesian modeling.** *Prev Vet Med* 2005, **68**(2–4):145–163.
13. Cheng D, Stamey JD, Branscum AJ: **Bayesian approach to average power calculations for binary regression models with misclassified outcomes.** *Stat Med* 2009, **28**(5):848–863.
14. Hanson TE, Johnson WO, Gastwirth JL: **Bayesian inference for prevalence and diagnostic test accuracy based on dual-pooled screening.** *Biostatistics* 2006, **7**(1):41–57.
15. Stamey JD, Young DM, Seaman JW Jr: **A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression.** *Stat Med* 2008, **27**(13):2440–2452.
16. Suess EA, Gardner IA, Johnson WO: **Hierarchical Bayesian model for prevalence inferences and determination of a country's status for an animal pathogen.** *Prev Vet Med* 2002, **55**(3):155–171.
17. Xia M, Gustafson P: **A Bayesian method for estimating prevalence in the presence of a hidden sub-population.** *Stat Med* 2012, **31**(21):2386–2398.
18. Black MA, Craig BA: **Estimating disease prevalence in the absence of a gold standard.** *Stat Med* 2002, **21**(18):2653–2669.
19. Erkanli A, Soyer R, Stangl D: **Bayesian inference in two-phase prevalence studies.** *Stat Med* 1997, **16**(10):1121–1133.
20. Dendukuri N, Joseph L: **Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests.** *Biometrics* 2001, **57**(1):158–167.
21. Spiegelhalter DJ, Best NG, Carlin BP: **Bayesian measures of model complexity and fit (with discussion).** *J R Stat Soc Ser B* 2002, **64**:583–640.
22. Su CL, Gardner IA, Johnson WO: **Diagnostic test accuracy and prevalence inferences based on joint and sequential testing with finite population sampling.** *Stat Med* 2004, **23**(14):2237–2255.

23. Berry SM, Carlin BP, Lee JJ, Muller P: **Bayesian adaptive methods for clinical trials.** CRC press; 2010: 72.
24. Lu Y, Dendukuri N, Schiller I, Joseph L: **A Bayesian approach to simultaneously adjusting for verification and reference standard bias in diagnostic test studies.** *Stat Med* 2010, **29**(24):2532–2543.
25. Joseph L, Gyorkos TW, Coupal L: **Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard.** *Am J Epidemiol* 1995, **141**(3):263–272.
26. Mao SS: *Bayesian Statistics.* Beijing: China Statistics Press; 2005:105–110.
27. Spiegelhalter DJ, Thomas A, Best NG, Lunn D: **WinBUGS Version 1.4.3 user manual.** [http://www.mrc-bsu.cam.ac.uk/bugs]
28. Gelman A, Rubin DB: **Inference from iterative simulation using multiple sequences.** *Stat Sci* 1992, **7:**457–511.
29. Ntzoufras I: *Bayesian Modeling Using WinBUGS.* John Wiley & Sons, Inc: Hoboken, New Jersey; 2009.
30. Castiglione G, Zappa M, Grazzini G, Sani C, Mazzotta A, Mantellini P, Ciatto S: **Cost analysis in a population based screening programme for colorectal cancer: comparison of immunochemical and guaiac faecal occult blood testing.** *J Med Screen* 1997, **4**(3):142–146.
31. Liu P, Yang HT, Qiang LY, Xiao S, Shi ZX: **Estimation of the sensitivity and specificity of assays for screening antibodies to HIV: a comparison between the frequentist and Bayesian approaches.** *J Virol Met* 2012, **186:**89–93.

doi:10.1186/1471-2288-14-110
Cite this article as: Liu *et al.*: A two-stage Bayesian method for estimating accuracy and disease prevalence for two dependent dichotomous screening tests when the status of individuals who are negative on both tests is unverified. *BMC Medical Research Methodology* 2014 **14**:110.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central