

RESEARCH

Open Access



Evaluation of the accuracy of diagnostic scales for a syndrome in Chinese medicine in the absence of a gold standard

Xiao Nan Wang¹, Vanessa Zhou³, Qiang Liu⁴, Ying Gao^{5*} and Xiao-Hua Zhou^{1,2*}

Abstract

Background: The concept of syndromes (*zhengs*) is unique to Chinese medicine (CM) and difficult to measure. Expert consensus is used as a gold standard to identify *zhengs* and evaluate the accuracy of existing diagnostic scales for *zhengs*. But, the use of expert consensus as a gold standard is problematic because the diagnosis of *zhengs* by expert consensus is not 100 % accurate. This study aimed to evaluate the accuracy of standardized diagnostic scales for a syndrome *zhengs* in the absence of a gold standard, with application to internal wind (*nei feng*) syndrome in ischemic stroke patients.

Methods: A total of 204 participants (age 41–84 years) with ischemic stroke were assessed by the stroke syndrome differentiation diagnostic criterion (SSDC), ischemic stroke TCM syndrome diagnostic scale (ISDS), and expert syndrome differentiation (ESD). The diagnostic tests and data collection process were conducted over a 10-month period (February 2008 to November 2008) in 10 hospitals across nine cities in China. The Bayesian method was used to estimate the accuracy of the SSDC, ISDS, and ESD.

Results: For internal wind syndrome, the estimated sensitivities and specificities of the SSDC, ISDS, and ESD without use of a gold standard were respectively: $\hat{Se}_1 = 0.687$, $\hat{Sp}_1 = 0.776$; $\hat{Se}_2 = 0.884$, $\hat{Sp}_2 = 0.875$; and $\hat{Se}_3 = 0.813$, $\hat{Sp}_3 = 0.922$

Conclusion: After adjusting for imperfect gold standard bias, we found that both the sensitivity and specificity of the ISDS were higher than those of the SSDC for diagnosis of internal wind syndrome in ischemic stroke patients.

Keywords: CM, Sensitivity, Specificity, Diagnostic scale, Syndrome, Bayesian method, Without a gold standard

Background

The concept of syndromes (*zhengs*) is unique to Chinese medicine (CM). Syndromes are identifiable from a holistic understanding of a patient's clinical presentation using the four CM diagnostic methods: observation, listening/smelling, questioning, and pulse analyses [1]. Identification of a syndrome can differ from one CM practitioner to another because of varying medical experience and other related factors. In recent years, the CM community has

developed several standardized diagnostic scales for syndromes [2–5]. The accuracies of these scales have been assessed by the diagnostic opinion of CM practitioners as the gold standard. However, an expert diagnosis is largely dependent on clinical experience and educational background, leading to different syndrome differentiation for the same patient by different expert CM practitioners. This results in biased estimates for the accuracy of diagnostic scales because the expert syndrome differentiation (ESD) is imperfect. Such bias is called an imperfect gold standard bias [6, 7]. If the diagnostic test and imperfect gold standard are conditionally independent of the true disease status, the sensitivity and specificity of the diagnostic test are underestimated. However, if the diagnostic test and imperfect gold standard are conditionally

*Correspondence: gaoying973@126.com; azhou@uw.edu

² Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA 98195, USA

⁵ Department of Neurology, Dongzhimen Hospital of Beijing University of Chinese Medicine, Beijing 100700, China

Full list of author information is available at the end of the article

dependent, the estimated sensitivity and specificity of the diagnostic test can be biased in either direction. The direction of the bias is determined by the degree to which the diagnostic tests and imperfect gold standard misclassify the same patients. When this tendency is slight, the accuracy of the diagnostic test is generally underestimated; when the tendency is strong, the accuracy of the diagnostic test is generally overestimated [6].

In recent years, several statistical methods have been developed to correct imperfect gold standard bias. Hui and Waiter [8] developed a model for two diagnostic tests within two populations and introduced a maximum likelihood approach when assuming the existence of two populations strata with different prevalence rates. In that model, they also assumed that the two tests were conditionally independent. However, the assumption of conditional independence may not be realistic in some applications owing to some common factors that can influence both diagnostic tests and true disease status. Sinclair and Gastwirth [9] extended the Hui and Waiter model to allow for conditional dependence. Espeland and Handelman [10] and Yang and Becker [11] proposed latent class modeling for conditional dependence, Qu et al. [12], Hadgu and Qu [13] proposed random effects models, and Albert and Dodd [14] developed latent class modeling approaches for binary tests. Pepe and Janes [15] discussed the latent class analysis method when assessing the multiple diagnostic tests without a gold standard, and concluded that a latent class model required careful justification of assumptions made about the conditional dependence structure. These researchers also stressed that a formal clinical definition of the disease should be given before evaluating the accuracy of diagnostic tests with the latent class method. Only when the disease has been clearly defined can the estimated parameters be meaningful for diagnostic tests; otherwise, the results of the estimators were meaningless. The above-mentioned methods used the frequentist approach to estimate the parameters in the model when the diagnostic tests were conditionally independent, given the true disease status or given the true disease status and a random effect.

Joseph et al. [16] used Bayesian methods to assess the accuracy of diagnostic tests under conditional independence without a gold standard. Dendukuri [17], Georgiadis et al. [18], and Branscum et al. [19] developed Bayesian models to evaluate the accuracy of diagnostic tests with two conditionally dependent tests. These methods have been widely used for estimation of the accuracy of diagnostic tests without a gold standard in Western medicine research [20–27]. However, they have not been applied for estimation of the accuracy of diagnostic tests for CM syndromes. This study aimed to evaluate the accuracy of standardized diagnostic scales for a syndrome in the

absence of a gold standard, with application to internal wind (nei feng) syndrome in ischemic stroke patients.

Methods

Study design and approval

In this study, we evaluated the accuracy of the stroke syndrome differentiation diagnostic criterion (SSDC), ischemic stroke TCM syndrome diagnostic scale (ISDS), and ESD for detecting “internal wind” in ischemic stroke patients, without assuming that the ESD is the gold standard. We mainly focused on comparing the accuracy of the two diagnostic scales (SSDC and ISDS). This study used data from the second round of a diagnostic test study of the ISDS. The diagnostic test and data collection process were performed over a 10-month period (February 2008 to November 2008), after receiving approval (ECSL-BDY-2008-012) from the Ethics Committee of the Dongzhimen Hospital of Beijing University of Chinese Medicine (Additional files 1 and 2).

Inclusion and exclusion criteria

Individuals who had a confirmed diagnosis of acute ischemic stroke by computed tomography and magnetic resonance imaging examinations, were aged between 35 and 85 years, and were informed of the objectives and research procedures of the study (details of study please see Additional file 3) and provided signed consent forms themselves (consent forms please see Additional file 4) were selected as the participants in this study [5]. We excluded individuals with the following symptoms: transient ischemic attack; cerebral hemorrhage or subarachnoid hemorrhage; stroke caused by brain tumor, traumatic brain injury, or blood disease; severe heart, liver, kidney, or hematopoietic system comorbidity and complication; mental disorder or severe dementia; and severe aphasia that could affect data collection [5].

Study subjects

The final data set comprised 204 patients from 10 hospitals across nine cities in China [4, 5]. All of the participants (age 41–84 years) were diagnosed with ischemic stroke. The mean age of the patients was 65 years, and the mode age was 74 years. The subjects were diagnosed as “0” or “1” by each of the SSDC, ISDS, and ESD. The detailed results of the cross-classification of the three diagnostic tests for internal wind syndrome in the 204 ischemic stroke participants are shown in Table 1. The CM syndrome factor scales (SSDC and ISDS) of the symptoms and signs, and the ESD were separately completed on the same day. In this study, an expert was defined as a physician, who had the clinical title of deputy director or above and also had more than 10 years of

Table 1 Cross-classified test results of T_1 , T_2 and T_3 for internal wind syndrome

T_1	$T_2 = 1$		$T_2 = 0$	
	$T_3 = 1$	$T_3 = 0$	$T_3 = 1$	$T_3 = 0$
$T_1 = 1$	69	19	7	12
$T_1 = 0$	32	9	5	51

clinical work experience in diagnosing and curing stroke disease with traditional CM.

CM syndrome factor scales and syndrome differentiation

The SSDC and ESD were used to diagnose the status of a patient in place of a gold standard, before the development of the ISDS. The SSDC was the first recognized scale for diagnosing a CM syndrome in ischemic stroke patients, and has been widely used since its publication in 1994 [2, 3]. The development of the ISDS was based on the SSDC. Essentially, the ISDS is an updated version of the SSDC [3], and was first developed in 2007. The simple process for developing the ISDS has been described in the published literature [3, 4]. Briefly, the ISDS was developed from a two-round Delphi study, which generated a pool of draft items with 288 items in six syndrome factor dimensions [4]. From this pool of items, six syndrome factor diagnostic scales were constructed according to logistic regression functions and receiver-operating curve analysis. Each syndrome factor diagnostic scale consisted of 10–20 “yes” or “no” statements. The ESD was completed by three senior physicians with over 10 years of work experience [4, 5]. When the practitioners failed to reach a unanimous decision about a patient’s diagnosis, the majority opinion was used.

Statistical methods

Descriptive statistics were utilized to summarize the characteristics of the subjects in the data set. The latent class model was fitted to the results of the SSDC, ISDS, and ESD for the ischemic stroke patients when a gold standard was not available. The Bayesian method was used to estimate the sensitivity and specificity for every CM diagnostic scale. We followed the guidelines for reporting Bayesian analyses in biomedical journals, as described by Lang and Altman [28].

Using the reporting guidelines, we first described the general Bayesian statistical model. Next, we specified the pre-trial probabilities (prior distributions) for the parameters in the proposed model based on the data we wanted to analyze and also explained how the prior distributions were selected. Subsequently, we used Markov chain Monte Carlo (MCMC) techniques to obtain the Bayesian

estimated parameters, based on the posterior distribution. The median and credibility interval were used as the posterior summary measures in this study. Finally, we illustrated the sensitivity of the analyses to different prior distributions in the Bayesian model.

IBM SPSS Statistics for Windows [version: 21.0; IBM Corp; NY] was utilized for the descriptive statistics. WinBUGS software [version: 1.4.3; BUGS project; UK] was used for the Bayesian data analysis (WinBUGS code for this study could be found in Additional file 5). A detailed description of the proposed Bayesian method for evaluating the accuracy of the diagnostic tests without a gold standard is given as below.

Notation

Let T_1 , T_2 , and T_3 denote the diagnostic results of the two CM diagnostic tests (SSDC and ISDS) and ESD for one syndrome factor in ischemic stroke patients, where T_1 , T_2 , and $T_3 = 0, 1$, with “1” indicating the presence of the syndrome factor and “0” indicating the absence of the syndrome factor. Let D denote the true status of the syndrome factor in an ischemic stroke patient, which is not observed in the study. The parameters of interest include: the prevalence of the syndrome factor in the population, π , defined as $\pi = P(D = 1)$; the sensitivity of the i th diagnostic test in detecting the syndrome factor, Se_i , defined as $Se_i = P(T_i = 1|D = 1)$; and the specificity of the i th diagnostic test for detecting the syndrome factor, Sp_i , defined as $Sp_i = P(T_i = 0|D = 0)$, where $i = 1, 2, 3$.

Bayesian model

Assume that there are n participants in the sample and three test results for every subject. We represent the observed data as $Y = (Y_{t_1, t_2, t_3})$, where Y_{t_1, t_2, t_3} is the number of subjects with $T_1 = t_1$, $T_2 = t_2$, and $T_3 = t_3$; here $t_1, t_2, t_3 = 0, 1$. For example, Y_{111} denotes the number of subjects whose diagnostic results for all three tests indicate that the syndrome factor is present. Correspondingly, p_{t_1, t_2, t_3} represents the joint probability of the outcome $(T_1 = t_1, T_2 = t_2, T_3 = t_3)$, which is defined as follows:

$$\begin{aligned}
 p_{t_1, t_2, t_3} &= P(T_1 = t_1, T_2 = t_2, T_3 = t_3) \\
 &= P(T_1 = t_1, T_2 = t_2, T_3 = t_3|D = 1) \times P(D = 1) \\
 &\quad + P(T_1 = t_1, T_2 = t_2, T_3 = t_3|D = 0) \times P(D = 0) \\
 &= P(T_1 = t_1, T_2 = t_2, T_3 = t_3|D = 1) \times \pi \\
 &\quad + P(T_1 = t_1, T_2 = t_2, T_3 = t_3|D = 0) \times (1 - \pi).
 \end{aligned} \tag{1}$$

Among the three tests in this study, the first two tests represent the diagnostic results of the SSDC and ISDS, respectively, and the last test represents the diagnostic result of the expert opinion, called the ESD. Since the

two diagnostic scales consist of standardized questionnaires, while the ESD is based on the individual opinion of expert CM practitioners, it is reasonable to assume that the CM expert and the diagnostic scales err independently (i.e., they are conditionally independent, given the true CM syndrome status). Nevertheless, the two diagnostic scales do not err independently (i.e., they are conditionally dependent, given the true CM syndrome status). Such dependence is measured by the conditional dependence correlations, given the true CM syndrome status. Hence, we assume that T_3 is independent of T_1 and T_2 conditional on D , while we allow T_1 and T_2 to be conditionally dependent, given D . Let C_+ and C_- denote the covariance between T_1 and T_2 among the CM syndrome positive and negative individuals, respectively. In other words, $C_+ = \text{cov}(T_1, T_2|D = 1)$ and $C_- = \text{cov}(T_1, T_2|D = 0)$. Such a model has also been studied by Dendukuri and Joseph [17]. To present the Bayesian method, we need to compute the likelihood function of the observed data. Note that we can respectively write $P(T_1 = t_1, T_2 = t_2|D = 1)$ and $P(T_1 = t_1, T_2 = t_2|D = 0)$ as follows:

$$\begin{aligned} P(T_1 = t_1, T_2 = t_2|D = 1) &= \prod_{i=1}^2 Se_i^{t_i} (1 - Se_i)^{(1-t_i)} + (-1)^{t_1+t_2} C_+, \\ P(T_1 = t_1, T_2 = t_2|D = 0) &= \prod_{i=1}^2 Sp_i^{(1-t_i)} (1 - Sp_i)^{t_i} + (-1)^{t_1+t_2} C_-. \end{aligned}$$

Consequently, we can rewrite the joint probability of the outcome $(T_1 = t_1, T_2 = t_2, T_3 = t_3)$ as follows:

$$\begin{aligned} p_{t_1, t_2, t_3} &= P(T_1 = t_1, T_2 = t_2|D = 1) \\ &\quad \times P(T_3 = t_3|D = 1) \times \pi \\ &\quad + P(T_1 = t_1, T_2 = t_2|D = 0) \\ &\quad \times P(T_3 = t_3|D = 0) \times (1 - \pi) \\ &= \pi \left[\prod_{i=1}^2 Se_i^{t_i} (1 - Se_i)^{(1-t_i)} + (-1)^{t_1+t_2} C_+ \right] \\ &\quad \times [Se_3^{t_3} (1 - Se_3)^{(1-t_3)}] \\ &\quad + (1 - \pi) \left[\prod_{i=1}^2 Sp_i^{(1-t_i)} (1 - Sp_i)^{t_i} + (-1)^{t_1+t_2} C_- \right] \\ &\quad \times [(1 - Sp_3)^{t_3} Sp_3^{(1-t_3)}]. \end{aligned} \quad (2)$$

Let $Y = (Y_{111}, Y_{110}, Y_{101}, Y_{100}, Y_{011}, Y_{010}, Y_{001}, Y_{000})$, the observed data, and $\theta = (Se_1, Sp_1, Se_2, Sp_2, Se_3, Sp_3, \pi, C_+, C_-)$, which represents the set of parameters in the model. According to (2), the likelihood function based on the observed data is:

$$\begin{aligned} L(\theta|Y) &= \prod_{t_1, t_2, t_3} p_{t_1, t_2, t_3}^{Y_{t_1, t_2, t_3}} \\ &= \prod_{t_1, t_2, t_3} \left\{ \pi \left[\prod_{i=1}^2 Se_i^{t_i} (1 - Se_i)^{(1-t_i)} + (-1)^{t_1+t_2} C_+ \right] \right. \\ &\quad \times [Se_3^{t_3} (1 - Se_3)^{(1-t_3)}] + (1 - \pi) \left[\prod_{i=1}^2 Sp_i^{(1-t_i)} (1 - Sp_i)^{t_i} \right. \\ &\quad \left. \left. + (-1)^{t_1+t_2} C_- \right] \times [(1 - Sp_3)^{t_3} Sp_3^{(1-t_3)}] \right\}^{Y_{t_1, t_2, t_3}} \end{aligned} \quad (3)$$

To use the Bayesian method to estimate the vector of the parameters, θ , we need to specify a prior distribution for θ . Let $f(\theta)$ denote the prior distribution of θ . The Bayesian method combines the prior information about θ with the data we have collected, and then uses the Bayes theorem to obtain an interpretable posterior distribution for θ . We can use the median of the posterior distribution to estimate θ . According to the Bayes theorem, the joint posterior distribution $f(\theta|Y)$ of the parameter θ given the observed data Y can be written as follows:

$$f(\theta|Y) = \frac{L(\theta|Y)f(\theta)}{\int L(\theta|Y)f(\theta)d\theta} = \frac{\mathcal{A}}{\mathcal{B}}, \quad (4)$$

where

$$\begin{aligned} \mathcal{A} &= f(\theta) \prod_{t_1, t_2, t_3} \left\{ \pi \left[\prod_{i=1}^2 Se_i^{t_i} (1 - Se_i)^{(1-t_i)} + (-1)^{t_1+t_2} C_+ \right] \right. \\ &\quad \times [Se_3^{t_3} (1 - Se_3)^{(1-t_3)}] + (1 - \pi) \\ &\quad \times \left[\prod_{i=1}^2 Sp_i^{(1-t_i)} (1 - Sp_i)^{t_i} + (-1)^{t_1+t_2} C_- \right] \\ &\quad \left. \times [(1 - Sp_3)^{t_3} Sp_3^{(1-t_3)}] \right\}^{Y_{t_1, t_2, t_3}}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{B} &= \underbrace{\int \int \cdots \int}_{\theta} f(\theta) \prod_{t_1, t_2, t_3} \left\{ \pi \left[\prod_{i=1}^2 Se_i^{t_i} (1 - Se_i)^{(1-t_i)} \right. \right. \\ &\quad \left. \left. + (-1)^{t_1+t_2} C_+ \right] [Se_3^{t_3} (1 - Se_3)^{(1-t_3)}] \right. \\ &\quad \left. + (1 - \pi) \left[\prod_{i=1}^2 Sp_i^{(1-t_i)} (1 - Sp_i)^{t_i} + (-1)^{t_1+t_2} C_- \right] \right. \\ &\quad \left. \times [(1 - Sp_3)^{t_3} Sp_3^{(1-t_3)}] \right\}^{Y_{t_1, t_2, t_3}} \underbrace{dSe_1 dSe_2 \cdots dC_+ dC_-}_{\theta} \end{aligned}$$

Consequently, the marginal posterior density function for any component in θ , such as Sp_2 , given the data, can be expressed as:

$$f(Sp_2|Y) = \underbrace{\int \int \cdots \int}_8 \int f(\theta|Y) dSe_1 \\ \times dSe_2 dSe_3 dSp_1 dSp_3 d\pi dC_+ dC_-$$

By estimating the median of the margin distribution about Sp_2 , denoted by $\widehat{Sp_2}$, we obtain a Bayesian estimate, $\widehat{Sp_2}$, for Sp_2 .

Procedure of the analysis

Here T_1, T_2 , and T_3 denote the CM diagnostic scales (SSDS and ISDS) and ESD for detecting internal wind syndrome, respectively. The observed data can be represented by $Y = (69, 19, 7, 12, 32, 9, 5, 51)$, as shown in Table 1. We denoted the proposed model as model (I). For comparison purposes, we also included the results obtained by the commonly used naive method, which assumed the ESD as the gold standard, and denoted this method as model (II). In the Bayesian analysis, a prior distribution for θ , which was defined in the Bayesian model, had to be chosen.

Selecting the prior distribution

A prior distribution for θ consisted of three sensitivities, three specificities, one prevalence rate, and two conditional covariances. Since the first six parameters have a range between 0 and 1, we chose a beta distribution $Beta(\alpha, \beta)$ for each of them, where α and β were hyper-parameters. We used the method proposed by Dendukuri [17] and Enøe et al. [27] to choose these hyper-parameter values by the priori moment information. According to the published literature describing the three diagnostic tests (SSDC, ISDS, and ESD) [2–5], the most probable value of the sensitivities of T_1 and T_2 for detecting internal wind syndrome was determined as 0.7, and we were 95 % sure that these sensitivities were less than 0.5. Thus, the prior distribution for the sensitivities T_1, T_2 was chosen to be the beta distribution, $Beta(13.322, 6.281)$. For the specificities of the diagnostic scales T_1, T_2 for detecting internal wind syndrome, the most probable value was determined as 0.8, and we were 95 % sure that these specificities were less than 0.5. Therefore, the prior distribution for the specificities T_1 and T_2 was chosen to be the beta distribution, $Beta(7.549, 2.637)$. The best guess value for the sensitivity of T_3 was 0.8, and the experts were 95 % sure that the sensitivity of T_3 was at least 0.7; hence, the prior distribution for the sensitivity of T_3 was chosen to be the beta distribution, $Beta(48.283, 12.821)$.

The best guess value for the specificity of T_3 was 0.85, and the experts were 95 % sure that the specificity of T_3 was at least 0.6; thus the prior distribution for the specificity of T_3 was chosen to be the beta distribution, $Beta(10.657, 2.704)$. The uniform distribution on $[0, 1]$ was used for the prior distribution of the internal wind prevalence rate. For the last two conditional covariances, C_+ and C_- , which measured the dependence of T_1 and T_2 among the diseased and non-diseased statuses, respectively, we have the following constraints: $(Se_1 - 1)(1 - Se_2) \leq C_+ \leq \min(Se_1, Se_2) - Se_1 Se_2$ and $(Sp_1 - 1)(1 - Sp_2) \leq C_- \leq \min(Sp_1, Sp_2) - Sp_1 Sp_2$, respectively. Hence, we chose two uniform distributions for C_+ and C_- : $U((Se_1 - 1)(1 - Se_2), (\min(Se_1, Se_2) - Se_1 Se_2))$ and $U((Sp_1 - 1)(1 - Sp_2), (\min(Sp_1, Sp_2) - Sp_1 Sp_2))$.

MCMC techniques for computing the posterior estimator

It was difficult to directly obtain the posterior estimator of each parameter through a numerical integration method in the Bayesian model. Since the joint posterior distribution $f(\theta | Y)$ was complicated and involved high-dimensional integral problems, which were often impossible to compute directly, we used the MCMC algorithm to draw a random sample from the joint posterior distribution. We then computed the sample median of the randomly drawn sample to estimate θ and its components of interest. In this study, the WinBUGS package was used to perform this MCMC process.

To use the MCMC technique in the Bayesian method, we specified the initial values of the model parameters, and the initial values were given as follows: $\pi = 0.623$, $Se_1 = 0.748$, $Se_2 = 0.945$, $Se_3 = 0.850$, $Sp_1 = 0.844$, $Sp_2 = 0.883$, $Sp_3 = 0.935$, respectively. We also chose different initial values and obtained similar results. The numbers of iterations and burn-ins were determined by the convergence of the Markov chain in estimating the parameters by WinBUGS.

Results and discussion

The sensitivity of the SSDC for internal wind syndrome (Table 2) was estimated as 0.687 by the Bayesian method in the absence of a gold standard, while the commonly used naive method, which uses the ESD as a gold standard, estimated the sensitivity of the SSDC as 0.673. The estimated sensitivity of the ISDS showed similar results. The Bayesian method estimated the specificity of the ISDS as 0.875, while the commonly used naive method estimated the specificity of the ISDS as 0.692. From these results, we can conclude that the commonly used naive method in CM for estimating the accuracy of diagnostic scales for this CM syndrome might be biased. Table 2

Table 2 Accuracy of diagnostic scales (median) for internal wind syndrome factor in 204 ischemic stroke patients under different models

	Model I	Model II
\hat{Se}_1	0.687 (0.605,0.765)	0.673 (0.587,0.759)
\hat{Se}_2	0.884 (0.815,0.938)	0.894 (0.837,0.951)
\hat{Se}_3	0.813 (0.742,0.880)	N.A.
\hat{Sp}_1	0.776 (0.652,0.885)	0.659 (0.562,0.756)
\hat{Sp}_2	0.875 (0.739,0.968)	0.692 (0.597,0.787)
\hat{Sp}_3	0.922 (0.831,0.981)	N.A.
$\hat{\pi}$	0.648 (0.556,0.731)	0.554 (0.486,0.622)

also shows the 95 % Bayesian confidence intervals for the sensitivity and specificity of the SSDC in detecting internal wind syndrome, which were (0.605,0.765) and (0.652,0.885), respectively. Similarly, the Bayesian confidence intervals for the sensitivity and specificity of the ISDS are also shown in Table 2.

As shown in Table 2, the respective Bayesian estimated sensitivities of the SSDC, ISDS, and ESD for diagnosing internal wind syndrome without a gold standard were as follows: $\hat{Se}_1 = 0.687$, $\hat{Se}_2 = 0.884$, and $\hat{Se}_3 = 0.813$. The respective estimated specificities of the SSDC, ISDS, and ESD for diagnosing internal wind syndrome in the absence of a gold standard were as follows: $\hat{Sp}_1 = 0.776$, $\hat{Sp}_2 = 0.875$, and $\hat{Sp}_3 = 0.922$. From these results, we concluded that the ISDS was more accurate than the SSDC in detecting internal wind syndrome. The Bayesian method also gave an estimate of $\hat{\pi} = 0.648$ for the prevalence rate of internal wind syndrome. Hence, we concluded that the sensitivity and specificity of the ISDS were both higher than those of the SSDC when diagnosing internal wind syndrome in ischemic stroke patients. We also found that the sensitivity and specificity of the ESD for internal wind syndrome were also high, but not perfect.

To assess the sensitivity of our results to chosen prior distributions, we selected several different prior distributions for parameters in model (I). The posterior estimates under the chosen prior distributions for the parameters led to consistent results with the previous posterior estimates.

Conclusion

After adjusting for imperfect gold standard bias, we found that both the sensitivity and specificity of the ISDS were higher than those of the SSDC for diagnosis of internal wind syndrome in ischemic stroke patients.

Additional files

Additional file 1. Approval document of ethics.

Additional file 2. Ethics committee members.

Additional file 3. Research Programme.

Additional file 4. Informed consent.

Additional file 5. WinBUGS code.

Abbreviations

CM: Chinese medicine; SSDC: stroke syndrome differentiation diagnostic criterion; ISDS: ischemic stroke TCM syndrome diagnostic scale; ESD: expert syndrome differentiation; MCMC: Markov chain Monte Carlo.

Authors' contributions

XZ, XW, YG, and QL conceived and designed the study. QL and YG facilitated the data collection in China. XW and XZ analyzed the data. XW, VZ, XZ, YG, and QL interpreted the results. XW, XZ, and VZ wrote the manuscript. XW and XZ revised the manuscript. All authors read and approved the final manuscript.

Author details

¹ School of Statistics, Renmin University of China, Beijing 100872, China.

² Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA 98195, USA. ³ UW Autism Center, University of Washington, Seattle, WA 98105, USA. ⁴ TCM International Cooperation Center of China, Beijing 100101, China. ⁵ Department of Neurology, Dongzhimen Hospital of Beijing University of Chinese Medicine, Beijing 100700, China.

Acknowledgements

The authors wish to acknowledge the support of the Ministry of Science and Technology of the PRC on a research project entitled "Significant New Drug Development-Construction of Technology Platform used for Original New Drug Research and Development" (2012ZX09303-010-002). The authors also wish to acknowledge the data provided by the "973 program," a basic research program supported by the Chinese Government Ministry of Science and Technology that promotes research in China. This research was also supported by the State Foundation for Studying Abroad.

Competing interests

The authors declare that they have no competing interests.

Received: 13 March 2015 Accepted: 27 May 2016

Published online: 28 July 2016

References

- Wang J, Wang P, Xiong X. Current situation and re-understanding of syndrome and formula syndrome in Chinese medicine. *Intern Med*. 2012;2:1–5. doi:10.4172/2165-8048.1000113.
- State Administration of TCM and Acute Encephalopathy Cooperation Group SSSS. TCM syndrome differentiation diagnosis criterion of stroke. *Beijing Zhong Yi Yao Da Xue Xue Bao*. 1994;17:42.
- Stroke Syndromes and Clinical Diagnosis SSSS. Clinical validation of TCM syndrome diagnostic criterion of stroke. *Beijing Zhong Yi Yao Da Xue Xue Bao*. 1994;17:41–3.
- Liu Q, Gao Y. Theory basis of syndrome diagnosis scale. *Zhong Hua Zhong Yi Yao Za Zhi*. 2010;25:989–92.
- Gao Y, Bin M, Liu Q, Wang Y. Methodological study and establishment of the diagnostic scale for TCM syndrome of ischemic stroke. *Zhong Yi Yao Za Zhi*. 2011;52:2097–101.
- Zhou X-H, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. New York: John Wiley and Sons; 2011.

7. Hui SL, Zhou X-H. Evaluation of diagnostic tests without gold standard. *Stat Methods Med Res.* 1998;7:354–70. doi:[10.1177/096228029800700404](https://doi.org/10.1177/096228029800700404).
8. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics.* 1980;36:167–71. doi:[10.2307/2530508](https://doi.org/10.2307/2530508).
9. Sinclair MD, Gastwirth JL. On procedures for evaluating the effectiveness of reinterview survey methods: application to labor force data. *J Am Stat Assoc.* 1996;91:961–9. doi:[10.1080/01621459.1996.10476966](https://doi.org/10.1080/01621459.1996.10476966).
10. Espeland MA, Handelman SL. Using latent class models to characterize and assess relative-error in discrete measurements. *Biometrics.* 1989;45:587–99. doi:[10.2307/2531499](https://doi.org/10.2307/2531499).
11. Yang J, Becker MP. Latent variable modeling of diagnostic accuracy. *Biometrics.* 1997;53:948–58. doi:[10.2307/2533555](https://doi.org/10.2307/2533555).
12. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic test. *Biometrics.* 1996;52:797–810. doi:[10.2307/2533043](https://doi.org/10.2307/2533043).
13. Hadgu A, Qu Y. A biomedical application of latent class models with random effects. *Appl Stat.* 1998;47:603–16. doi:[10.1111/1467-9876.00131](https://doi.org/10.1111/1467-9876.00131).
14. Albert PS, Dodd LE. A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics.* 2004;60:427–35. doi:[10.1111/j.0006-341X.2004.00187.x](https://doi.org/10.1111/j.0006-341X.2004.00187.x).
15. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. *Biostatistics.* 2007;8:474–84. doi:[10.1093/biostatistics/kxl038](https://doi.org/10.1093/biostatistics/kxl038).
16. Joseph L, Gyorkos T, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol.* 1995;141:263–72.
17. Dendukuri N, Joseph L. Bayesian approaches to modeling conditional dependence between multiple diagnostic tests. *Biometrics.* 2001;57:158–67. doi:[10.1111/j.0006-341X.2001.00158.x](https://doi.org/10.1111/j.0006-341X.2001.00158.x).
18. Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Appl Stat.* 2003;52:63–76. doi:[10.1111/1467-9876.00389](https://doi.org/10.1111/1467-9876.00389).
19. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med.* 2005;68:145–63. doi:[10.1016/j.prevetmed.2004.12.005](https://doi.org/10.1016/j.prevetmed.2004.12.005).
20. Rybicki BA, Peterson EL, Johnson CC, Kortsha GX, Cleary WM, Gorell JM. Intra- and inter-rater agreement in the assessment of occupational exposure to metals. *Int J Epidemiol.* 1998;27:269–73. doi:[10.1093/ije/27.2.269](https://doi.org/10.1093/ije/27.2.269).
21. McDermott J, Drews C, Green D, Berg C. Evaluation of prenatal care information on birth certificates. *Paediatr Perinat Epidemiol.* 1997;11:105–21. doi:[10.1046/j.1365-3016.1997.d01-4.x](https://doi.org/10.1046/j.1365-3016.1997.d01-4.x).
22. Line BR, Peters TL, Keenan J. Diagnostic test comparisons in patients with deep venous thrombosis. *J Nucl Med.* 1997;38:89–92.
23. Mahoney WJ, Szatmari P, Maclean JE, Bryson SE, Bartolucci G, Walter SD, Marshall BJ, Zwaigenbaum L. Reliability and accuracy of differentiating pervasive developmental disorder subtypes. *J Am Acad Child Adolesc Psychiatry.* 1998;37:278–85. doi:[10.1097/00004583-199803000-00012](https://doi.org/10.1097/00004583-199803000-00012).
24. Chriel M, Willeberg P. Dependency between sensitivity, specificity and prevalence analysed by means of Gibbs sampling. *Epidémiologie et Santé Animale.* 1997;31/32:12.03.1–3.
25. Georgiadis MP, Gardner IA, Hedrick RP. Field evaluation of sensitivity and specificity of a polymerase chain reaction (PCR) for detection of *N. salmonis* in rainbow trout. *J Aquat Anim Health.* 1998;10:372–80. doi:[10.1577/1548-8667\(1998\)010<0372:FEOSAS>2.0.CO;2](https://doi.org/10.1577/1548-8667(1998)010<0372:FEOSAS>2.0.CO;2).
26. Singer RS, Boyce WM, Gardner IA, Johnson WO, Fisher AS. Evaluation of bluetongue virus diagnostic tests in free-ranging bighorn sheep. *Prev Vet Med.* 1998;35:265–82. doi:[10.1016/S0167-5877\(98\)00067-1](https://doi.org/10.1016/S0167-5877(98)00067-1).
27. Enoe C, Georgiadis MP, Johnson WO. Estimation of the sensitivity and specificity of diagnostic tests and disease prevalence when true disease state is unknown. *Prev Vet Med.* 2000;45:61–81. doi:[10.1016/S0167-5877\(00\)00117-3](https://doi.org/10.1016/S0167-5877(00)00117-3).
28. Lang T, Altman D. Statistical Analyses and Methods in the Published Literature: the SAMPL Guidelines. Science Editors' Handbook, European Association of Science Editors; 2013.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

