

RESEARCH ARTICLE

Open Access

A trivariate meta-analysis of diagnostic studies accounting for prevalence and non-evaluable subjects: re-evaluation of the meta-analysis of coronary CT angiography studies

Xiaoye Ma^{1*}, Muhammad Fareed K Suri² and Haitao Chu¹

Abstract

Background: A recent paper proposed an intent-to-diagnose approach to handle non-evaluable index test results and discussed several alternative approaches, with an application to the meta-analysis of coronary CT angiography diagnostic accuracy studies. However, no simulation studies have been conducted to test the performance of the methods.

Methods: We propose an extended trivariate generalized linear mixed model (TGLMM) to handle non-evaluable index test results. The performance of the intent-to-diagnose approach, the alternative approaches and the extended TGLMM approach is examined by extensive simulation studies. The meta-analysis of coronary CT angiography diagnostic accuracy studies is re-evaluated by the extended TGLMM.

Results: Simulation studies showed that the intent-to-diagnose approach under-estimate sensitivity and specificity. Under the missing at random (MAR) assumption, the TGLMM gives nearly unbiased estimates of test accuracy indices and disease prevalence. After applying the TGLMM approach to re-evaluate the coronary CT angiography meta-analysis, overall median sensitivity is 0.98 (0.967, 0.993), specificity is 0.875 (0.827, 0.923) and disease prevalence is 0.478 (0.379, 0.577).

Conclusions: Under MAR assumption, the intent-to-diagnose approach under-estimate both sensitivity and specificity, while the extended TGLMM gives nearly unbiased estimates of sensitivity, specificity and prevalence. We recommend the extended TGLMM to handle non-evaluable index test subjects.

Keywords: Meta-analysis, Diagnostic test, Non-evaluable subjects

Background

In studies of meta-analysis of diagnostic test comparing an index test with a reference test, non-evaluable test outcome is an important issue that could potentially lead to biased estimates of index test accuracy. Many papers in the literature discussed missing reference test outcome (missing disease status) and how to correct such bias, so called partial verification bias or work up bias [1-4]. However, index test outcomes can be non-evaluable as

well, especially for tests yielding dichotomous results. Different situations were discussed where index test result can be non-evaluable: uninterpretable, intermediate and indeterminate [5,6].

For a single study, there are many discussions about how to deal with non-evaluable index test outcomes, such as excluding them [7], grouping them with positive or negative outcomes [5,7], or use 3×2 table to report them as an extension of the standard 2×2 table [7]. On the other hand, in meta-analysis, there is little discussion on how to deal with missing index test outcomes [6]. The "classic" 2×2 table models such as the bivariate linear mixed models [8-13], bivariate generalized linear mixed model (GLMM) [14-16] and TGLMM

*Correspondence: xiaoye1043@gmail.com

¹Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware St. SE, 55455 Minneapolis, MN, USA
Full list of author information is available at the end of the article

[17] ignore missing index test outcomes. Recently, a paper by Schuetz et al. [6] discussed this issue by studying different approaches dealing with index test non-evaluable subjects. The paper conducted a meta-analysis of coronary CT angiography studies and presented an intent-to-diagnose approach together with three commonly applied alternative approaches. The intent-to-diagnose approach takes non-evaluable diseased subjects as false positives and non-diseased subjects as false negatives such that sensitivity and specificity won't be over-estimated. We name the other three alternative approaches in Schuetz et al. [6] as Model 1 (non-evaluable subjects are excluded from the study), Model 2 (non-evaluable diseased subjects are taken as true positives and non-diseased subjects are taken as false positives) and Model 3 (non-evaluable diseased subjects are taken as false negatives and non-diseased subjects are taken as true negatives). We use Model 1-3 to denote the above three approaches throughout the rest of this paper. The authors concluded that excluding the index test non-evaluable subjects (Model 1) leads to overestimation of sensitivity and specificity and recommended the conservative intent-to-diagnose approach by treating non-evaluable diseased subjects as false negatives and non-evaluable non-diseased subjects as false positives. However, no simulation studies have been conducted to evaluate the performance of these approaches. Moreover, the above conclusions can be misleading.

We can treat index test non-evaluable subjects as missing data. Schuetz et al. [6] concluded that sensitivity and specificity could be over-estimated by excluding non-evaluable subjects. In fact, under a reasonable general assumption, missing at random (MAR), excluding non-evaluable subjects can provide unbiased estimates of sensitivity (Se) and specificity (Sp). Under MAR assumption, the probability of missing only depend on observed information, such as patient characteristics and known true disease status [18,19]. For example, when diagnosing extrahepatic cholestasis using percutaneous transhepatic cholangiography, non-diseased subjects can have uninterpretable results more often than diseased patients [5]. A special case of MAR is missing completely at random (MCAR), where missing is independent of both observed and unobserved variables [18]. E.g., accidental contamination of a urine sample such that the test result is discarded. Under MAR, T and M are independent given disease status D , where $M = 1, 0$ indicates missingness of index test outcome, $D = 1, 0$ indicates diseased or non-diseased and $T = 1, 0$ represents index test positive or negative. Hence, excluding non-evaluable subjects will have unbiased estimates of Se and Sp : $\widehat{Se} = Pr(T = 1|D = 1, M = 0) = Pr(T = 1|D = 1)$ and $\widehat{Sp} = Pr(T = 0|D = 0, M = 0) = Pr(T = 0|D = 0)$. Similarly, positive and negative likelihood ratios (LR+ and LR-) and area

under the curve (AUC) are unbiased too. Under MCAR, $Pr(M = 1|D = 1) = Pr(M = 1|D = 0)$, and hence disease prevalence (π) estimate is also unbiased if non-evaluable subjects are excluded. However, when missing probabilities are not equal between diseased and non-diseased participants, disease prevalence estimate can be biased if non-evaluable subjects are excluded, leading overall estimates of positive predictive value (PPV) and negative predictive value (NPV) biased. PPV and NPV are generally preferred by clinicians as measurements of how well a test predicts true disease status because their interpretations are more intuitive: PPV is the probability that a subject with positive index test result is truly diseased and NPV is the probability that a subject with negative index test result is truly non-diseased [19]. However, none of the approaches discussed in Schuetz et al. [6] can correct bias in their estimates.

In this article, we propose to extend the TGLMM approach [17] by treating non-evaluable subjects as missing data to adjust for potential bias. The TGLMM was proposed by Chu et al. [17] as an extension of the bivariate GLMM [9,10,14]. Sensitivities and specificities are found to be potentially dependent on disease prevalence [20-22]. The TGLMM models disease prevalence together with sensitivity and specificity to account for potential correlations among them. Moreover, once overall disease prevalence is evaluated, other test accuracy indices such as PPV and NPV can be calculated. By extending the TGLMM to account for missing data, potential bias in disease prevalence estimate can be adjusted and thus, bias in PPV and NPV estimates can be avoided.

In the rest of this paper, we first present the extended TGLMM approach in the "Methods" section. Next, in section "Results", simulation studies are carried out to systematically evaluate the performance of the extended TGLMM, Model 1-3 and the intent-to-diagnose approach when there are non-evaluable index test subjects. The meta-analysis of coronary CT angiography studies is re-evaluated by the extended TGLMM approach. The SAS code for the extended TGLMM is available in the Appendix. Finally, we conclude the paper with some discussions in section "Conclusions".

Methods

Assume there are $i = 1, \dots, N$ studies in one meta-analysis data set. We generalize the TGLMM approach to account for missing index test outcomes by extending the "classic" 2×2 table to Table 1. Each cell in Table 1 reports the cell count and cell probability corresponding to a combination of index test and disease outcomes in study i . Let n_{itd} denote the cell counts in study i with index test outcome $T = t$ and reference test outcome $D = d$, where $t = 1, 0, m$ stands for positive, negative and missing, and $d = 1, 0$ denotes positive and negative. Se_i, Sp_i

Table 1 3 × 2 table accounting for prevalence and missing index test results

Index test	Gold standard		Total
	+	−	
+	n_{i11}	n_{i10}	n_{i1+}
	$(1 - \omega_{im1})\pi_i Se_i$	$(1 - \omega_{im0})(1 - \pi_i)(1 - Sp_i)$	$(1 - \omega_{im1})\pi_i Se_i + (1 - \omega_{im0})(1 - \pi_i)(1 - Sp_i)$
−	n_{i01}	n_{i00}	n_{i0+}
	$(1 - \omega_{im1})\pi_i(1 - Se_i)$	$(1 - \omega_{im0})(1 - \pi_i)Sp_i$	$(1 - \omega_{im1})\pi_i(1 - Se_i) + (1 - \omega_{im0})(1 - \pi_i)Sp_i$
Missing	n_{im1}	n_{im0}	n_{im+}
	$\omega_{im1}\pi_i$	$\omega_{im0}(1 - \pi_i)$	$\omega_{im1}\pi_i + \omega_{im0}(1 - \pi_i)$
Total	n_{i+1}	n_{i+0}	n_{i++}
	π_i	$1 - \pi_i$	1

Each cell reports the cell count and cell probability corresponding to a combination of index test and disease outcomes in study i . n_{imd} denotes the cell counts in study i with index test outcome $T = t$ and reference test outcome $D = d$, where $t = 1, 0, m$ stands for positive, negative and missing, and $d = 1, 0$ denotes positive and negative. Se_i, Sp_i and π_i are sensitivity, specificity and prevalence of study i , respectively. ω_{imd} denotes the missing probability of index test given disease status d in study i .

and π_i are sensitivity, specificity and prevalence of study i , respectively. Let ω_{imd} denote the missing probability of index test given disease status d in study i : $\omega_{imd} = Pr(T = m | D = d)$. The missing probabilities and disease prevalence are incorporated in the cell probabilities in Table 1. Assuming a multinomial distribution, the likelihood for $\theta_i = (Se_i, Sp_i, \pi_i)$ and $\omega_i = (\omega_{im1}, \omega_{im0})$ given data (cell counts) is:

$$L(\theta_i, \omega_i | \text{Data}) \propto \{(1 - \omega_{im1})\pi_i Se_i\}^{n_{i11}} \{(1 - \omega_{im0})(1 - \pi_i)(1 - Sp_i)\}^{n_{i10}} \{(1 - \omega_{im1})\pi_i(1 - Se_i)\}^{n_{i01}} \{(1 - \omega_{im0})(1 - \pi_i)Sp_i\}^{n_{i00}} \{\pi_i \omega_{im1}\}^{n_{im1}} \{(1 - \pi_i)\omega_{im0}\}^{n_{im0}} \quad (1)$$

It is straight forward to tell from (1) that $L(\theta_i, \omega_i | \text{Data}) \propto L(\theta_i | \text{Data}) \times L(\omega_i | \text{Data})$, where the log-likelihood of θ_i is:

$$\begin{aligned} \log L(\theta_i | \text{Data}) &= n_{i11} \{\log(\pi_i) + \log(Se_i)\} \\ &+ n_{i10} \{\log(1 - \pi_i) + \log(1 - Sp_i)\} \\ &+ n_{i01} \{\log(\pi_i) + \log(1 - Se_i)\} \\ &+ n_{i00} \{\log(1 - \pi_i) + \log(Sp_i)\} \\ &+ n_{im1} \log(\pi_i) + n_{im0} \log(1 - \pi_i) \end{aligned}$$

Let $\theta = \{\theta_i\}$. Assuming independence among studies conditional on θ_i , the total log likelihood of θ is:

$$\log L(\theta | \text{Data}) = \sum_{i=1}^N \log L(\theta_i | \text{Data}) \quad (2)$$

Let $\text{logit}(\pi_i) = \eta + \varepsilon_i$, $\text{logit}(Se_i) = \alpha + \mu_i$ and $\text{logit}(Sp_i) = \beta + \nu_i$, where $\text{logit}(\cdot)$ is the logit link function

such that $\text{logit}(p) = \log(p/(1 - p))$, for $0 < p < 1$. (η, α, β) are the fixed effect parameters such that median π, Se and Sp can be approximated as $\text{logit}^{-1}(\eta)$, $\text{logit}^{-1}(\alpha)$ and $\text{logit}^{-1}(\beta)$, respectively, where $\text{logit}^{-1}(\cdot)$ is the inverse logit function such that $\text{logit}^{-1}(x) = 1/(1 + \exp(-x))$. The random effect vector $(\varepsilon_i, \mu_i, \nu_i)$ is assumed to be trivariate normally distributed:

$$(\varepsilon_i, \mu_i, \nu_i)^T \sim MVN(\mathbf{0}, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \rho_{\varepsilon\mu}\sigma_\mu\sigma_\varepsilon & \rho_{\varepsilon\nu}\sigma_\nu\sigma_\varepsilon \\ & \sigma_\mu^2 & \rho_{\mu\nu}\sigma_\nu\sigma_\mu \\ & & \sigma_\nu^2 \end{bmatrix},$$

where the diagonal elements in Σ account for between-study variations of π, Se and Sp and the off-diagonal elements take care of potential correlations among the three parameters.

Median PPV, NPV, LR+ and LR− and median area under the curve (AUC_M) can be approximated as [16]:

$$PPV = \frac{\text{logit}^{-1}(\eta)\text{logit}^{-1}(\alpha)}{\text{logit}^{-1}(\eta)\text{logit}^{-1}(\alpha) + \{1 - \text{logit}^{-1}(\eta)\}\{1 - \text{logit}^{-1}(\beta)\}},$$

$$NPV = \frac{\{1 - \text{logit}^{-1}(\eta)\}\text{logit}^{-1}(\beta)}{\{1 - \text{logit}^{-1}(\eta)\}\text{logit}^{-1}(\beta) + \text{logit}^{-1}(\eta)\{1 - \text{logit}^{-1}(\alpha)\}},$$

$$LR+ = \text{logit}^{-1}(\alpha) / \{1 - \text{logit}^{-1}(\beta)\},$$

$$LR- = \{1 - \text{logit}^{-1}(\alpha)\} / \text{logit}^{-1}(\beta),$$

$$\begin{aligned} AUC_M &= \int_0^1 \text{logit}^{-1} \{(\alpha - \rho_{\mu\nu}\beta\sigma_\mu) / \sigma_\nu \\ &+ \rho_{\mu\nu}\sigma_\mu / \sigma_\nu [\text{logit}(1 - Sp)]\} dSp. \end{aligned}$$

The extended TGLMM can be fitted by standard software like SAS NLMIXED procedure, which implements an adaptive Gaussian quadrature to approximate the log-likelihood in (2) integrated on random effects with dual

quasi-Newton optimization techniques. The NLMIXED procedure directly outputs fixed effects estimates $\hat{\eta}$, $\hat{\alpha}$ and $\hat{\beta}$ and can provide median prevalence, Se, Sp, PPV, NPV, LR+, LR- estimates and their confidence intervals through the “estimate” statements. Sample SAS code is available in the Appendix.

Results

Simulations

Simulation scenarios

We conduct simulation studies under three missing scenarios to systematically evaluate the performance of the proposed extended TGLMM approach and the approaches discussed in Schuetz et al. [6]: missing probabilities for diseased and non-diseased subjects are same (0.1), or missing probability of diseased group (0.1) is smaller than non-diseased group (0.2), or missing probability of diseased group (0.2) is larger than non-diseased group (0.1). All three scenarios satisfy the MAR assumption, and the first scenario is in fact MCAR [18]. True sensitivity and specificity are 0.7 and 0.9, disease prevalence is 0.25 and variances of Se, Sp and prevalence are 1 on logit scale. These assumptions mimic a diagnostic test with relatively low sensitivity, high specificity and a disease with moderate prevalence. A moderate positive correlation of 0.3 is assumed between Se and π , and moderate negative correlations of -0.3 are assumed between Sp and π and between Se and Sp, on logit scales. Such correlation directions were observed in some meta-analysis studies [11,20]. Intuitively, a population with higher prevalence may have more diseased cases with clear disease symptoms, leading to increased sensitivity. Under each setting, 5000 meta-analysis data sets are simulated with 30 studies in each data set. π_i , Se_i and Sp_i for each study were generated according to the trivariate assumption described in the Methods section. True and false positives, true and false negatives and non-evaluable counts are sampled from the multinomial distribution in Table 1. For each simulated meta-analysis data set, the extended TGLMM, Model 1-3 and the intent-to-diagnose approach are fitted. Bias in percentage, mean standard error (SE) and 95% confidence interval coverage probability (CP) are collected and compared for estimates of sensitivity, specificity, prevalence, PPV, NPV, LR+ and LR-. Bias in percentage is calculated by $(\hat{\delta} - \delta) \times 100/\delta$, where δ is the true value and $\hat{\delta}$ is the estimator.

Simulation results

Table 2 shows the simulation results under different scenarios. When MCAR ($\omega_{m1} = \omega_{m0} = 0.1$), disease prevalence estimates from all five models are nearly unbiased (bias less than 1%). The extended TGLMM and Model 1 both give nearly unbiased estimates (bias less than 1.6%) and nominal coverage probabilities around 93% for Se,

Sp, PPV, NPV, LR+ and LR- estimates. Model 2 over-estimates sensitivity and under-estimates specificity: bias of sensitivity estimate is 4.6% and bias of specificity estimates is 11.9%. Estimates of PPV and LR+ are more biased (22.6% bias for PPV and 49.2% bias for LR+). Using Model 3 sensitivities are largely under-estimated (12.6% bias) and specificities are over-estimated (1.1% bias). The intent-to-diagnose approach largely under-estimates both sensitivity and specificity (12.6% and 11.9% bias, respectively). The CPs for some estimates from Model 2 and 3 and the intent-to-diagnose approach can be as low as 0 (e.g., specificity estimates from Model 2), indicating that none of the confidence intervals cover the true values. When missing probability of the diseased group is smaller than the non-diseased group ($\omega_{m1} = 0.1$, $\omega_{m0} = 0.2$), the extended TGLMM and Model 1 both give nearly unbiased estimates (bias around 0.1%) of sensitivity and specificity. However, Model 1 over-estimates disease prevalence (9.6% bias) while the extended TGLMM gives nearly unbiased (bias within 1%) estimate of prevalence. As a consequence, Model 1 gives biased estimates of PPV and NPV (3.1% and 1.3%, respectively), while the extended TGLMM provides nearly unbiased estimates for all parameters (within 2%). Again, under this scenario, the intent-to-diagnose approach largely under-estimates sensitivity, specificity, PPV, NPV and LR+ and over-estimates LR-, with CPs less than 40% and some as low as 0. On the other hand, when $\omega_{m1} = 0.2$ and $\omega_{m0} = 0.1$, the extended TGLMM and Model 1 again give nearly unbiased estimates (bias around 0.1%) of sensitivity and specificity. Model 1 under-estimates disease prevalence (8.4% bias) while the extended TGLMM provides nearly unbiased estimates. The intent-to-diagnose approach largely under-estimates sensitivity, specificity, PPV, NPV and LR+ and over-estimates LR- and some CPs are as low as 0. When the missing probabilities for diseased and non-diseased subjects are more unbalanced, we expect the estimates from Model 1-3 and the intent-to-diagnose approach to have larger bias and smaller CP. In practice, however, depending on the test performance and missing probabilities, the direction and magnitude of the bias from the four approaches discussed in Schuetz et al. [6] can be different from what we observed in these simulation studies.

Re-evaluation of the meta-analysis of coronary CT angiography studies

Cardiac CT scans can be used to rule out stenoses, however, are found to be subject to non-evaluable results. Schuetz et al. [6] performed a systematic search for diagnostic accuracy studies of coronary CT angiography. The authors searched Medline, Embase and ISI Web of Science databases for prospective studies using conventional coronary angiography as the gold standard and have patients with non-evaluable CT images. Eventually, 26 studies were

Table 2 Simulation results under MAR assumption

Model Estimate	TGLMM			Model 1			Model2			Model3			Intent-to-diagnose		
	Bias%	meanSE	CP	Bias%	meanSE	CP	Bias%	meanSE	CP	Bias%	meanSE	CP	Bias%	meanSE	CP
$\omega_{m1} = \omega_{m0} = 0.1$															
Se	-0.3	0.041	0.94	-0.3	0.041	0.94	4.6	0.036	0.81	-12.6	0.037	0.33	-12.6	0.036	0.33
Sp	-0.1	0.017	0.93	-0.1	0.017	0.93	-11.9	0.018	0	1.1	0.015	0.84	-11.9	0.017	0
Prev	0.8	0.034	0.93	0.8	0.034	0.93	0.8	0.034	0.93	0.8	0.034	0.93	0.8	0.034	0.93
PPV	-0.1	0.046	0.94	-0.3	0.046	0.94	-22.6	0.047	0.08	-0.9	0.046	0.94	-29	0.049	0.01
NPV	-0.1	0.018	0.93	-0.1	0.018	0.93	-0.2	0.018	0.93	-2.9	0.020	0.81	-4.6	0.022	0.59
LR+	1.6	1.188	0.92	1.6	1.189	0.93	-49.2	0.307	0	-0.5	1.160	0.92	-57.6	0.271	0
LR-	0.9	0.044	0.94	0.9	0.044	0.94	1.5	0.044	0.94	27.9	0.039	0.33	46.8	0.045	0.04
$\omega_{m1} = 0.1, \omega_{m0} = 0.2$															
Se	-0.1	0.041	0.94	-0.1	0.041	0.94	4.7	0.036	0.80	-12.3	0.036	0.34	-12.3	0.036	0.34
Sp	-0.1	0.017	0.94	-0.1	0.017	0.94	-22.3	0.017	0	2.2	0.014	0.62	-22.3	0.017	0
Prev	0.4	0.034	0.93	9.6	0.036	0.90	0.4	0.034	0.93	0.4	0.034	0.93	0.4	0.034	0.93
PPV	-0.3	0.046	0.93	3.1	0.044	0.88	-36	0.047	0	2.7	0.044	0.89	-42.1	0.047	0
NPV	-0.1	0.018	0.94	-1.3	0.020	0.93	-1.4	0.020	0.92	-2.7	0.020	0.83	-6.3	0.024	0.36
LR+	1.4	1.195	0.94	1.4	1.194	0.94	-65.1	0.159	0	12.3	1.312	0.95	-70.8	0.147	0
LR-	0.6	0.044	0.93	0.6	0.044	0.93	14.7	0.050	0.85	26.1	0.038	0.39	66.1	0.051	0
$\omega_{m1} = 0.2, \omega_{m0} = 0.1$															
Se	-0.1	0.023	0.93	-0.1	0.023	0.93	8.7	0.018	0.12	-21	0.020	0	-21	0.019	0
Sp	0	0.009	0.93	0	0.009	0.93	-10.6	0.009	0	1.1	0.008	0.74	-10.6	0.009	0
Prev	0	0.018	0.93	-8.4	0.017	0.72	0	0.017	0.91	0	0.017	0.91	0	0.0168	0.89
PPV	-0.1	0.025	0.93	-3.7	0.027	0.83	-19.1	0.025	0	-4	0.026	0.8	-30.6	0.025	0
NPV	0	0.010	0.92	1.1	0.009	0.76	1.1	0.009	0.74	-4.6	0.011	0.05	-6.2	0.012	0
LR+	0.3	0.655	0.93	0.3	0.653	0.93	-44.1	0.196	0	-11.7	0.570	0.62	-59.3	0.154	0
LR-	0.3	0.025	0.93	0.3	0.025	0.93	-10.8	0.022	0.62	47.4	0.021	0	66.7	0.024	0

Three scenarios are studied: equal or unequal missing probabilities for the diseased and non-diseased groups. Bias in percentage(Bias%), mean standard error (meanSE) and 95% confidence interval coverage probability (CP) are summarized for estimates of sensitivity (Se), specificity (Sp), prevalence (Prev), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+) and negative likelihood ratio (LR-) from different models. "TGLMM" stands for the extended TGLMM. Model 1 excludes non-evaluable subjects, Model 2 takes non-evaluable subjects as index test positives, Model 3 takes non-evaluable subjects as index test negatives and the intent-to-diagnose approach takes non-evaluable subjects as false positives and false negatives.

included that reports cell counts in a 3×2 table as Table 1. The authors mentioned that the 3×2 table can be extended to a 3×3 table for non-evaluable results of the gold standard, however such cases were rare (0.1%) in this systematic review. We re-evaluate the 26 studies by the extended TGLMM and compare to the estimates following the four approaches discussed in Schuetz et al. [6].

The fitted median estimates and 95% confidence intervals are reported in Table 3. The extended TGLMM accounting for missing subjects gives median sensitivity, specificity, LR+, LR- and AUC estimates close to the estimates when non-evaluable subjects are excluded as in Model 1. The median disease prevalence estimated from the extended TGLMM is slightly lower than the estimate from Model 1. Model 2 gives significantly lower specificity estimate and Model 3 gives lower sensitivity estimate. The intent-to-diagnose approach provides lower estimates for sensitivity, specificity and AUC as it is the most conservative approach. Figure 1 presents the estimated PPV and NPV with 95% confidence bands versus prevalence, based on the overall sensitivity and specificity estimates from the extended TGLMM and the intent-to-diagnose approach. Figure 1 shows that as disease prevalence changes, PPV and NPV estimates from the latter approach are not ever included in the 95% confidence band of the estimates from the extended TGLMM, which suggests potential underestimation of PPV and NPV.

Discussions

Adequate reporting of the missing outcomes in study reports is essential to apply the discussed models. As shown in the simulation studies, different missing

scenarios can have different impact on how estimates are biased and more importantly, missing mechanism can indicate whether the MAR assumption holds. When the MAR assumption is violated, i.e., the probability of non-evaluation depends on unobserved index test outcomes, the direction and magnitude of bias are hard to predict. Few sensitivity analysis methods using pattern mixture models and selection models are available for this scenario [23,24]. These approaches can be explored in further research. On the other hand, number of non-evaluable results need to be known in order to apply the proposed methods. However, a recent study shows that they are not consistently or adequately reported in published studies [25].

A reviewer has pointed out that as long as number of non-evaluable subjects are known, disease prevalence can be estimated unbiasedly through an univariate meta-analysis. Consequently, together with unbiased sensitivity and specificity estimates, PPV and NPV estimates are unbiased too. This approach is a simpler method than the proposed extended TGLMM to estimate prevalence, however, can be less efficient by ignoring the potential correlation between prevalence, sensitivity and specificity, which may result in wider confidence intervals.

For an individual patient, different approaches of treating a missing result can have different impact. For example, if index test results are missing due to the same reason of returning a negative result (and thus is MNAR), then treating such patients as disease negatives can yield unbiased estimate of prevalence for a study, and also won't affect the patients' diagnosis. On the contrary, if index test missing patients are treated as positives for reasons such as suspicious of serious disease like cancer [26],

Table 3 Median estimates and 95% CI (in brackets) for parameter estimates using different methods

Method	Sensitivity	Specificity	Prevalence	PPV
TGLMM	98.0 (96.7,99.3)	87.5 (82.7,92.3)	47.8 (37.9,57.7)	87.8 (83.3,92.3)
Model 1	98.0 (96.7,99.3)	87.4 (82.5, 92.3)	49.3 (38.9,59.7)	88.4 (84,92.7)
Model 2	98.1 (96.9,99.3)	75.9 (69.3,82.5)	47.8 (37.9,57.8)	78.9 (71.9,85.9)
Model 3	91.7 (88.1,95.4)	89 (85.4,92.7)	47.8 (37.9,57.7)	88.4 (84.1,92.7)
Intent-to-diagnose	91.7 (88.1,95.3)	76.2 (69.7,82.6)	47.9 (37.9,57.9)	78 (70.2,85.7)
Method	NPV	LR+	LR-	AUC
TGLMM	97.9 (96.4,99.5)	7.8 (4.8,10.9)	0.02 (0.01,0.04)	0.99 (0.96,1)
Model 1	97.8 (96.1,99.4)	7.8 (4.8,10.9)	0.02 (0.01,0.04)	0.99 (0.96,1)
Model 2	97.8 (96.2, 99.4)	4.1 (2.9,5.2)	0.02 (0.01,0.04)	0.98 (0.97,1)
Model 3	92.1 (88.4,95.8)	8.4 (5.5,11.3)	0.09 (0.05,0.14)	0.96 (0.93,0.99)
Intent-to-diagnose	90.9 (86.4,95.5)	3.8 (2.7,5.0)	0.11 (0.06,0.16)	0.93 (0.89,0.96)

"TGLMM" stands for the extended TGLMM. Model 1 excludes non-evaluable subjects, Model 2 takes non-evaluable subjects as index test positives, Model 3 takes non-evaluable subjects as index test negatives and the intent-to-diagnose approach takes non-evaluable subjects as false positives and false negatives. Positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), negative likelihood ratio (LR-) and area under the curve (AUC) are summarized.

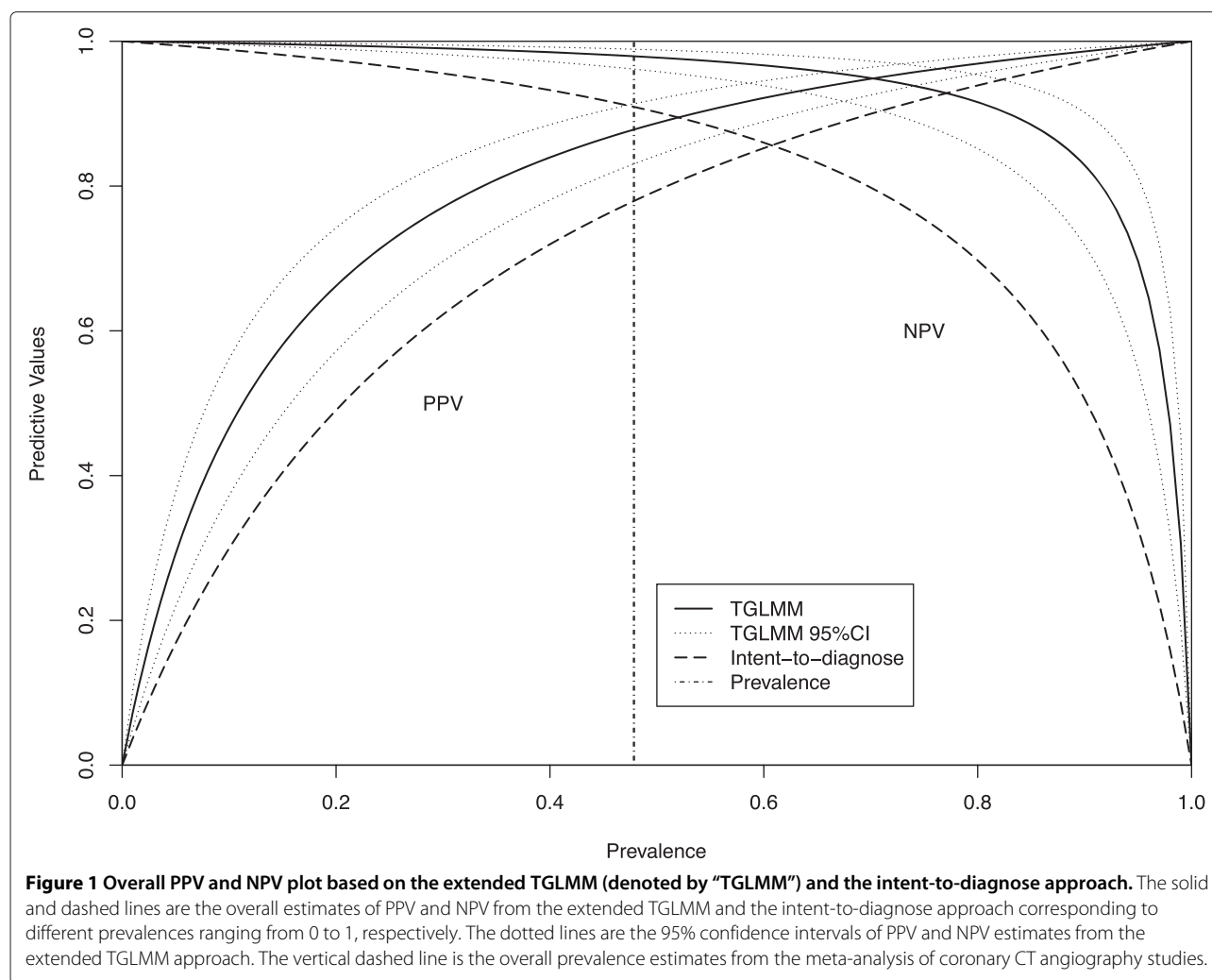


Figure 1 Overall PPV and NPV plot based on the extended TGLMM (denoted by “TGLMM”) and the intent-to-diagnose approach. The solid and dashed lines are the overall estimates of PPV and NPV from the extended TGLMM and the intent-to-diagnose approach corresponding to different prevalences ranging from 0 to 1, respectively. The dotted lines are the 95% confidence intervals of PPV and NPV estimates from the extended TGLMM approach. The vertical dashed line is the overall prevalence estimates from the meta-analysis of coronary CT angiography studies.

it may result in over-estimation of disease prevalence and unnecessary medial cost for the patient. For another example, if index test is repeatable and repeated for subjects with non-evaluable results, then it is appropriate to ignore missing results.

Conclusions

In this paper, we propose an extended TGLMM approach to handle non-evaluable index test subjects in meta-analysis of diagnostic tests. The extended TGLMM is compared to an intent-to-diagnose approach and three alternative approaches proposed by Schuetz et al. [6] through simulation studies and re-evaluation of the meta-analysis of coronary CT angiography studies.

In summary, by simulation studies we showed that under MAR assumption, excluding index test non-evaluable subjects (Model 1) will not lead to biased estimates of sensitivity, specificity, LR+, LR– and AUC. Thus in practice, researchers can be confident to apply Model

1 when there is a belief in the MAR assumption. However, when disease prevalence or PPV and NPV are of interest, excluding non-evaluable subjects could lead to biased estimates of these parameters. Under this situation, the extended TGLMM accounting for missingness should be preferred. Even though the extended TGLMM is more theoretically complex than the widely used bivariate random effects model, it is easy to program use SAS NLMIXED procedure. Sample SAS code with an application to the meta-analysis of coronary CT angiography studies is provided in the Appendix. Model 2, Model 3 and the intent-to-diagnose approach all largely under- or over-estimate sensitivity and specificity, so that they should not be recommended when MAR assumption is not seriously violated.

Claims

Ethical approvals and informed consents are not applicable to this paper.

Appendix: SAS code of the extended TGLMM approach: meta-analysis of coronary CT angiography studies

```
/*Read in data for the coronary CT
angiography studies*/ data observe;
input sid year ntp1 nfp1 nfn1 ntn1 nm1 nm0;
datalines;
1 2005 25 4 0 19 2 0
2 2008 57 3 2 79 0 9
... ..
26 2008 32 6 4 6 10 0
; run;

/*Macro to compute AUC*/
%macro auclogit;
estimate "AUC" 0
%do sp=1 %to 1000;
+0.001/(1+exp(-alpha0-RhoSeSp*exp(sigse)/
exp(sigsp)*
(log(1-0.001*&sp+0.0005)
-log(0.001*&sp-0.0005)+beta0)) )
%end;
;
%mend auclogit;

/*TGLMM using NLMIXED procedure*/
proc nlmixed data=observe fd df=1000
cov corr ecorr gtol=1e-10;
parms beta0=2 alpha0=1.5 eta0=0.2
sigse=0.4 sigsp=-0.3 sigpi=0.5
fZ=0 fZ1=0 fZ2=0;
lsei = alpha0 + muse ;
lspi = beta0 + musp ;
lppi = eta0 + mupi ;
Sei=1/(1+exp(-lsei));
Spi=1/(1+exp(-lspi));
ppi=1/(1+exp(-lppi));
RhoSeSp = (exp(2*fZ)-1)/(exp(2*fZ)+1);
RhoSePi = (exp(2*fZ1)-1)/(exp(2*fZ1)+1);
RhoSpPi = (exp(2*fZ2)-1)/(exp(2*fZ2)+1);

/* The log likelihood accounting for
missing cell counts by nm1*log(ppi)
and nm0*log(1-ppi)*/

logL= ntp1 * (log(ppi) + log(Sei)) +
nfp1 * (log(1-ppi) + log(1-Spi)) +
nfn1 * (log(ppi) + log(1-Sei)) + ntn1 *
(log(1-ppi) + log(Spi)) +nm1 * log(ppi) +
nm0 * log(1-ppi);
model ntp1 ~ general(logL);
random muse musp mupi~normal([0, 0, 0],
[exp(2*sigse),
```

```
RhoSeSp*exp(sigse+sigsp), exp(2*sigsp),
RhoSePi*exp(sigse+sigpi),
RhoSpPi*exp(sigsp+sigpi), exp(2*sigpi)] )
subject=sid;
%auclogit;
```

```
/*estimate statements in proc NLMIXED
can give point estimates of functions
of parameters as well as variances
from delta method*/
```

```
estimate "Se" 1/(1+exp(-alpha0));
estimate "Sp" 1/(1+exp(-beta0));
estimate "Prev" 1/(1+exp(-eta0));
estimate "PPV" exp(eta0+alpha0)*
(1+exp(beta0))/(exp(eta0+alpha0)*
(1+exp(beta0))+(1+exp(alpha0)));
estimate "NPV" exp(beta0)*(1+exp(alpha0))/
(exp(beta0)*(1+exp(alpha0))+exp(eta0)*
(1+exp(beta0)));
estimate "LR+" exp(alpha0)/(1+exp(alpha0))/
(1-exp(beta0)/(1+exp(beta0)));
estimate "LR-" (1-exp(alpha0))/
(1+exp(alpha0))/(exp(beta0)/
(1+exp(beta0))); run;
```

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All the authors contributed substantively to the study and approved the manuscript submitted for review. XM and HC conceived of the idea of the study, XM was responsible for data analysis. FS, XM and HC all contributed in drafting and revising the manuscript. All authors read and approved the final manuscript.

Acknowledgements

XM and HC were supported in part by the US NIAID A1103012, NCI P01CA142538, NCI P30CA077598, and U54-MD008620. XM, FS and HC were supported by NHLBI 1R01HL105626. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources

Author details

¹Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware St. SE, 55455 Minneapolis, MN, USA. ²Department of Neurology, University of Minnesota, MMC 295, 420 Delaware St. SE, 55455 Minneapolis, MN, USA.

Received: 29 September 2014 Accepted: 18 November 2014

Published: 4 December 2014

References

1. Begg CB, Greenes RA: **Assessment of diagnostic tests when disease verification is subject to selection bias.** *Biometrics* 1983, **39**(1):207-215.
2. de Groot JA, Dendukuri N, Janssen KJ, Reitsma JB, Brophy J, Joseph L, Bossuyt PM, Moons KG: **Adjusting for partial verification or workup bias in meta-analyses of diagnostic accuracy studies.** *Am J Epidemiol* 2012, **175**(8):847-853.
3. Harel O, Zhou X: **Multiple imputation for correcting verification bias.** *Stat Med* 2006, **25**(22):3769-3786.

4. Ransohoff DF, Feinstein AR: **Problems of spectrum and bias in evaluating the efficacy of diagnostic tests.** *N Engl J Med* 1978, **299**(17):926–930.
5. Begg CB, Greenes RA, Iglewicz B: **The influence of uninterpretability on the assessment of diagnostic tests.** *J Chronic Dis* 1986, **39**(8):575–584.
6. Schuetz GM, Schlattmann P: **Use of 3 × 2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary ct angiography studies.** *BMJ* 2012, **345**(2):6717–6717.
7. Simel DL, Feussner JR, Delong ER, Matchar DB: **Intermediate, indeterminate, and uninterpretable diagnostic test results.** *Med Decis Making* 1987, **7**(2):107–114.
8. Rutter CM, Gatsonis CA: **A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations.** *Stat Med* 2001, **20**(19):2865–2884.
9. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH: **Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.** *J Clin Epidemiol* 2005, **58**(10):982–990.
10. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA: **A unification of models for meta-analysis of diagnostic accuracy studies.** *Biostatistics* 2007, **8**(2):239–251.
11. Ma X, Nie L, Cole SR, Chu H: **Statistical methods for multivariate meta-analysis of diagnostic tests: An overview and tutorial.** *Stat Methods Med Res* 2013. in press.
12. Van Houwelingen HC, Arends LR, Stijnen T: **Advanced methods in meta-analysis: multivariate approach and meta-regression.** *Stat Med* 2002, **21**(4):589–624.
13. Zwinderman AH, Bossuyt PM: **We should not pool diagnostic likelihood ratios in systematic reviews.** *Stat Med* 2008, **27**(5):687–697.
14. Chu H, Cole SR: **Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach.** *J Clin Epidemiol* 2006, **59**(12):1331–1332.
15. Hamza TH, Reitsma JB, Stijnen T: **Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary roc approaches.** *Med Decis Making* 2008, **28**(5):639–649.
16. Chu H, Guo H, Zhou Y: **Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models.** *Med Decis Making* 2010, **30**(4):499–508.
17. Chu H, Nie L, Cole SR, Poole C: **Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection.** *Stat Med* 2009, **28**(18):2384–2399.
18. Little RJ, Rubin D: *Statistical Analysis with Missing Data, 2nd Edn.* New Jersey: John Wiley & Sons; 2002.
19. Pepe MS: *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford: Oxford University Press; 2003.
20. Brenner H, Gefeller O: **Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence.** *Stat Med* 1997, **16**(9):981–991.
21. Choi BC: **Causal modeling to estimate sensitivity and specificity of a test when prevalence changes.** *Epidemiology* 1997, **1**:80–86.
22. Leeflang MM, Bossuyt PM, Irwig L: **Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis.** *J Clin Epidemiol* 2009, **62**(1):5–12.
23. Little RJ: **Modeling the drop-out mechanism in repeated-measures studies.** *J Am Stat Assoc* 1995, **90**(431):1112–1121.
24. Scharfstein DO, Rotnitzky A, Robins JM: **Adjusting for nonignorable drop-out using semiparametric nonresponse models.** *J Am Stat Assoc* 1999, **94**(448):1096–1120.
25. Shinkins B, Thompson M, Mallett S, Perera R: **Diagnostic accuracy studies: how to report and analyse inconclusive test results.** *BMJ: Br Med J* 2013, **346**:2778.
26. Blick CG, Nazir SA, Mallett S, Turney BW, Onwu NN, Roberts IS, Crew JP, Cowan NC: **Evaluation of diagnostic strategies for bladder cancer using computed tomography (ct) urography, flexible cystoscopy and voided urine cytology: results for 778 patients from a hospital haematuria clinic.** *BJU Int* 2012, **110**(1):84–94.

doi:10.1186/1471-2288-14-128

Cite this article as: Ma et al.: A trivariate meta-analysis of diagnostic studies accounting for prevalence and non-evaluable subjects: re-evaluation of the meta-analysis of coronary CT angiography studies. *BMC Medical Research Methodology* 2014 **14**:128.

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

