Received: 11 June 2017 | Revised: 1 August 2017 | Accepted: 6 August 2017

DOI: 10.1002/sim.7449

RESEARCH ARTICLE

WILEY Statistics in Medicine

Modeling conditional dependence among multiple diagnostic tests

Zhuoyu Wang[1] | Nandini Dendukuri[1,3] | Heather J. Zar[2] | Lawrence Joseph[1,3]

[1]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec H3A 1A2, Canada

[2]Department of Paediatrics and Child Health and MRC Unit on Child and Adolescent Health, Red Cross Childrens Hospital, University of Cape Town, Cape Town, South Africa

[3]Division of Clinical Epidemiology, McGill University Health Center, Montreal, Quebec H3A 1A1, Canada

Correspondence
Nandini Dendukuri, Centre for Outcomes Research and Evaluation (CORE) 5252 boul. de Maisonneuve Bureau 3F.50 Montreal, Quebec, H4A 3S5.
Email: nandini.dendukuri@mcgill.ca

Funding information
Canadian Institutes of Health Research; Natural Sciences and Engineering Research Council of Canada; Canadian Institutes of Health Research (CIHR); Natural Sciences and Engineering Research Council (NSERC) of Canada; National Institutes of Health, Grant/Award Number: R01HD058971; Medical Research Council South Africa

When multiple imperfect dichotomous diagnostic tests are applied to an individual, it is possible that some or all of their results remain dependent even after conditioning on the true disease status. The estimates could be biased if this conditional dependence is ignored when using the test results to infer about the prevalence of a disease or the accuracies of the diagnostic tests. However, statistical methods correcting for this bias by modelling higher-order conditional dependence terms between multiple diagnostic tests are not well addressed in the literature. This paper extends a Bayesian fixed effects model for 2 diagnostic tests with pairwise correlation to cases with 3 or more diagnostic tests with higher order correlations. Simulation results show that the proposed fixed effects model works well both in the case when the tests are highly correlated and in the case when the tests are truly conditionally independent, provided adequate external information is available in the form of fixed constraints or prior distributions. A data set on the diagnosis of childhood pulmonary tuberculosis is used to illustrate the proposed model.

KEYWORDS
Bayesian inference, childhood pulmonary tuberculosis, correlations, fixed effects model, higher-order conditional dependence, latent class model

1 | INTRODUCTION

When multiple imperfect dichotomous diagnostic tests are applied to an individual in a medical study, it is possible that some or all of their results remain dependent even after conditioning on the dichotomous latent disease status. For example, diagnosis of childhood pulmonary tuberculosis (CPTB) relies on different microbiological, immune response, and imaging tests, as there is no single perfect test for this disease. In 1 data set, 749 hospitalized children suspected of CPTB from South Africa were tested by culture, smear microscopy, Xpert MTB/RIF (Xpert), tuberculin skin test (TST), and chest X-Ray (CXR).[13] It is of interest to estimate not only the prevalence of TB in this cohort, but also the diagnostic properties (rate of false positive and false negative results) for each test. With no perfect test, this problem is non-trivial and is typically handled using a latent class model that assumes that the observed data are a mixture of

disease positive and disease negative groups. The model would also need to consider the possible inter-dependence between these tests, 4 of which are influenced by the unknown bacterial load.

A variety of approaches for incorporating conditional dependence between imperfect tests have been proposed. These different approaches typically extend the conditional independence model (CIM) with 2 latent classes,[10,17] either by adding covariance terms or by sub-dividing the disease positive and disease negative latent classes. For example, Vacek[16] extended the CIM of Hui and Walter[9] by adding covariance terms between a pair of tests, within each of the disease positive and disease negative groups. Espeland and Handelman[5] increased the number of latent classes by 2 by adding unequivocally negative and unequivocally positive categories. Dendukuri et al[2] showed how the increase in the number of latent classes can be determined by the types of tests observed, recognizing that tests based on different biological mechanisms may be detecting different targets. Qu et al[12] proposed a random effects approach, accounting for conditional dependence by modeling the sensitivities and specificities of tests as functions of subject-specific Gaussian random variables. In other words, their model potentially included as many latent classes as the sample size. This was extended by Xu and Craig[19] who proposed a probit latent class model with a general correlation matrix structure for modeling the pairwise correlation between multiple tests. Dendukuri and Joseph[3] proposed Bayesian versions of the fixed effects model (FEM) described by Vacek[16] and the random effects model of Qu et al.[12]

Amongst all of these models, the one described by Vacek[16] and its variants are especially attractive because the pairwise correlation terms are easy to interpret. However, these models are limited in considering only pairwise correlations and hence are unsuitable for our motivating CPTB dataset. Though higher-order correlations may be modeled using random effects,[2] these models are computationally more complex as they substantially increase the number of unknown parameters to be estimated. Further, several assumptions are required about the distribution of the random effects and the magnitude of their association with different tests.

The literature to date has not devoted much attention to models that include higher-order covariance terms. Some authors have modeled only pairwise correlations reasoning that higher order dependence terms are less likely to occur in practice,[15] some have relied on a series of conditional probabilities which avoids specification of higher order correlation terms,[1] while others have entirely ignored these terms resulting in an overly simplistic specification of the likelihood.[11] In this paper, we will extend the FEM proposed by Vacek[16] to cases with 3 or more diagnostic tests with higher order correlations.

In Section 2, we introduce a general FEM for multiple diagnostic tests, with accompanying notation. In Section 3, we describe Bayesian inference for the FEM. Section 4 presents some simulation results, and we apply the FEM to the CPTB data set in Section 5. In Section 6, we summarize and discuss the main contributions of the new model.

2 | MODEL AND NOTATION

2.1 | Fixed effects model for conditional dependence among multiple tests

Assume K dichotomous diagnostic tests, T_i, $i = 1, 2, ..., K$, are applied to an individual to ascertain the presence or absence of a disease or condition D. Let $T_i = 0$ for a negative diagnosis and $T_i = 1$ for a positive diagnosis, and $D = 0$ for truly non-diseased subjects, otherwise $D = 1$. Denote the sensitivity and specificity of test T_i as S_i and C_i, respectively. $S_i = P(T_i = 1|D = 1)$ is the sensitivity of test i, with $P(T_i = 0|D = 1)$ the false negative rate. Similarly, $C_i = P(T_i = 0|D = 0)$ is the specificity of test i, with corresponding false positive rate given by $P(T_i = 1|D = 0)$. The joint conditional probability of the K tests on an individual given true disease status $D = d$ can be written in the form

$$\mathrm{P}(T_1 = t_1, T_2 = t_2, ...T_K = t_K|D = d) = \prod_{i=1}^{K} P(T_i = t_i|D = d) + \varphi_{t_1,t_2,...,t_K|d}, \quad (1)$$

where the parameter $\varphi_{t_1,t_2,...,t_K|d}$ is a conditional dependence term, which can be considered as the difference between the true joint probability and the joint probability under the conditional independence assumption.

Denote the prevalence of the disease as $\pi = P(D = 1)$. Then, the joint probability of test results ($T_1 = t_1$, $T_2 = t_2$, ..., $T_K = t_K$) is given by

$$\begin{aligned}
\mathrm{p}_{t_1\ldots t_K} &= \mathrm{P}(T_1 = t_1, T_2 = t_2, \ldots T_K = t_K) \\
&= \sum_{d=0}^{1} \{P(T_1 = t_1, T_2 = t_2, \ldots T_K = t_K | D = d) P(D = d)\} \\
&= \pi \left\{ \prod_{i=1}^{K} P(T_i = t_i | D = 1) + \varphi_{t_1, t_2, \ldots, t_K | 1} \right\} \\
&\quad + (1-\pi) \left\{ \prod_{i=1}^{K} P(T_i = t_i | D = 0) + \varphi_{t_1, t_2, \ldots, t_K | 0} \right\}.
\end{aligned}$$

Given a sample of N subjects, a multinomial distribution can be used to represent the likelihood function. Denote the number of subjects with test results (t_1, t_2, ..., t_K) by $\mathrm{n}_{t_1,\ldots,t_K}$. Then

$$\left(n_{0\ldots0}, \ldots, n_{1,\ldots,1}\right) \sim multinomial(N, (p_{0\ldots0}, p_{1\ldots1})).$$

The FEM described by Vacek[16] is a special case of this likelihood, when $K = 2$. In this case, Equation 1 becomes

$$\mathrm{P}(T_1 = t_1, T_2 = t_2 | D = d) = \prod_{i=1}^{2} P(T_i = t_i | D = d) + \varphi_{t_1, t_2 | d},$$

and the joint probability for an observed test result reduces to

$$\mathrm{p}_{t_1 t_2} = \pi \left\{ \prod_{i=1}^{2} P(T_i = t_i | D = 1) + \varphi_{t_1, t_2 | 1} \right\} + (1-\pi) \left\{ \prod_{i=1}^{2} P(T_i = t_i | D = 0) + \varphi_{t_1, t_2 | 0} \right\}.$$

When $K = 3$, Equation 1 becomes

$$\mathrm{P}(T_1 = t_1, T_2 = t_2, T_3 = t_3 | D = d) = \prod_{i=1}^{3} P(T_i = t_i | D = d) + \varphi_{t_1, t_2, t_3 | d},$$

and the joint probability is

$$\mathrm{p}_{t_1 t_2 t_3} = \pi \left\{ \prod_{i=1}^{3} P(T_i = t_i | D = 1) + \varphi_{t_1, t_2, t_3 | 1} \right\} + (1-\pi) \left\{ \prod_{i=1}^{3} P(T_i = t_i | D = 0) + \varphi_{t_1, t_2, t_3 | 0} \right\}.$$

Of course, similar expressions can be written for $K > 3$. As the number of tests increases, the number of conditional dependence terms increases substantially. With 2 tests, the number of dependence terms is 8, and with 3 tests, it increases to 16. In general, we have 2^{K+1} dependence terms for K tests.

2.2 | Equality constraints

In total, there are $2^{K+1} + 2K + 1$ parameters in the model, including 2^{K+1} dependence terms, K sensitivities, K specificities, and the disease prevalence. As we will show later, this large number of parameters to be estimated can be reduced by considering a series of equality constraints relating the conditional dependence terms. Under the parameterization introduced in Section 2.1, it can be shown that the conditional dependence term among a subset of the K tests, say, T_1, T_2, ..., T_k, where $k < K$, with testing results [t_1, t_2, ..., t_k], is the marginal sum of $\varphi_{t_1, t_2, \ldots, t_K | d}$ across the remaining tests, ie, T_{k+1}, T_{k+2}, ..., T_K. For example, when $K = 3$, the joint conditional probability of the first 2 tests can be written as

$$\mathrm{P}(T_1 = t_1, T_2 = t_2 | D = d) = \prod_{i=1}^{2} P(T_i = t_i | D = d) + \varphi_{t_1, t_2 | d}, \tag{2}$$

However, it can also be written as

$$\begin{aligned} \mathrm{P}(T_1 = t_1, T_2 = t_2 | D = d) &= \sum_{t_3=0}^{1} P(T_1 = t_1, T_2 = t_2, T_3 = t_3 | D = d) \\ &= \sum_{t_3=0}^{1} \left[\prod_{i=1}^{3} P(T_i = t_i | D = d) + \varphi_{t_1,t_2,t_3|d} \right] \\ &= \prod_{i=1}^{2} P(T_i = t_i | D = d) + \varphi_{t_1,t_2,0|d} + \varphi_{t_1,t_2,1|d}. \end{aligned} \tag{3}$$

Equating 2 and 3, we have $\varphi_{t_1,t_2|d} = \varphi_{t_1,t_2,0|d} + \varphi_{t_1,t_2,1|d}$. The pairwise conditional dependence between any 2 of the K tests can be expressed similarly.

Further, the marginal conditional probability $P(T_i = t_i | D = d)$ can be written as

$$\begin{aligned} P(T_i = t_i | D = d) &= \sum_{t_1=0}^{1} \cdots \sum_{t_{i-1}=0}^{1} \sum_{t_{i+1}=0}^{1} \cdots \sum_{t_K=0}^{1} P(T_1 = t_1, ..., T_i = t_i, .., T_K = t_K | D = d) \\ &= \sum_{t_1=0}^{1} \cdots \sum_{t_{i-1}=0}^{1} \sum_{t_{i+1}=0}^{1} \cdots \sum_{t_K=0}^{1} \left[\prod_{j=1}^{K} P\left(T_j = t_j | D = d\right) + \varphi_{t_1,t_2,...,t_K|d} \right] \\ &= P(T_i = t_i | D = d) + \sum_{t_1=0}^{1} \cdots \sum_{t_{i-1}=0}^{1} \sum_{t_{i+1}=0}^{1} \cdots \sum_{t_K=0}^{1} \varphi_{t_1,t_2,...,t_K|d}. \end{aligned}$$

Then, it must follow that

$$\sum_{t_1=0}^{1} \cdots \sum_{t_{i-1}=0}^{1} \sum_{t_{i+1}=0}^{1} \cdots \sum_{t_K=0}^{1} \varphi_{t_1,t_2,...,t_K|d} = \varphi_{\bullet,\cdots t_i,\cdots,\bullet|d} = 0, \tag{4}$$

where $\bullet$ means summing over all possible results of the corresponding test.

In addition, by summing the left and right-hand sides of Equation 1 across all possible test results for a given disease status d, we can derive 1 more constraint,

$$\sum_{t_1=0}^{1} \cdots \sum_{t_K=0}^{1} \varphi_{t_1,t_2,...,t_K|d} = \varphi_{\bullet,\cdots,\bullet|d} = 0, \tag{5}$$

In Equation 4, because T_i can be any of the K tests and t_i could equal 0 or 1, we have a total of 2 K constraints under true disease status D = d. However, from (5) we know that the term $\varphi_{\bullet,\cdots,\bullet|d}$ can be written as the sum of $\varphi_{\bullet,\cdots t_i,\cdots,\bullet|d}$ across over $t_i = 0$ and $t_i = 1$. Because the sum of the 2 terms must be zero, any of the terms being zero implies that the other is also zero. Therefore, K of the $2\,K$ constraints from (4) are redundant, reducing the number of independent constraints to $K + 1$ for each disease status, that is, $2\,K + 2$ overall.

When $K = 2$, we have $2\,K + 2 = 6$ constraints, 2 for each disease status following from Equation 4 and 1 for each disease status following from Equation 5:

$$\begin{aligned} \varphi_{0,0|d} + \varphi_{0,1|d} = 0 \text{ or } \varphi_{1,0|d} + \varphi_{1,1|d} = 0, \\ \varphi_{0,0|d} + \varphi_{1,0|d} = 0 \text{ or } \varphi_{0,1|d} + \varphi_{1,1|d} = 0, \end{aligned}$$

and

$$\varphi_{0,0|d} + \varphi_{0,1|d} + \varphi_{1,0|d} + \varphi_{1,1|d} = 0.$$

These constraints can be re-expressed as $\varphi_{1,1|d} = -\varphi_{0,1|d} = -\varphi_{1,0|d} = \varphi_{0,0|d}$, leaving us with only 2 dependence terms to be estimated, one for each disease status.[3]

When $K = 3$, we have 8 independent constraints, 4 for each disease status, including:

$$\varphi_{0,0,0|d} + \varphi_{0,1,0|d} + \varphi_{1,0,0|d} + \varphi_{1,1,0|d} = 0 \text{ or } \varphi_{0,0,1|d} + \varphi_{0,1,1|d} + \varphi_{1,0,1|d} + \varphi_{1,1,1|d} = 0,$$

$$\begin{aligned} \varphi_{0,0,0|d} + \varphi_{0,0,1|d} + \varphi_{1,0,0|d} + \varphi_{1,0,1|d} = 0 \text{ or } \varphi_{0,1,0|d} + \varphi_{0,1,1|d} + \varphi_{1,1,0|d} + \varphi_{1,1,1|d} = 0, \\ \varphi_{0,0,0|d} + \varphi_{0,0,1|d} + \varphi_{0,1,0|d} + \varphi_{0,1,1|d} = 0 \text{ or } \varphi_{1,0,0|d} + \varphi_{1,0,1|d} + \varphi_{1,1,0|d} + \varphi_{1,1,1|d} = 0, \end{aligned}$$

following from Equation 4, and

$$\varphi_{0,0,0|d} + \varphi_{0,0,1|d} + \varphi_{0,1,0|d} + \varphi_{0,1,1|d} + \varphi_{1,0,0|d} + \varphi_{1,0,1|d} + \varphi_{1,1,0|d} + \varphi_{1,1,1|d} = 0,$$

following from Equation 5. This implies that, for each disease status, any 4 of the conditional dependence terms can be expressed as a function of the other 4 conditional dependence terms, such that

$$\begin{aligned}
\varphi_{1,0,0|d} &= \varphi_{0,1,1|d} + \varphi_{1,1,1|d} - \varphi_{0,0,0|d}, \\
\varphi_{1,0,1|d} &= -\varphi_{0,0,1|d} - \varphi_{0,1,1|d} - \varphi_{1,1,1|d}, \\
\varphi_{1,1,0|d} &= -\varphi_{0,0,0|d} + \varphi_{0,0,1|d} - \varphi_{0,1,1|d}, \\
\varphi_{0,1,0|d} &= -\varphi_{0,0,0|d} - \varphi_{0,0,1|d} - \varphi_{0,1,1|d}.
\end{aligned} \tag{6}$$

This leaves us with 8 conditional dependence terms to be estimated, 4 for each disease status.

When there are K tests, the FEM includes 2^K dependence terms for each disease status, with $K + 1$ constraints on these dependence terms. Therefore, in total, $2^{K+1} - (2K + 2)$ dependence terms in the model need to be estimated, while the remaining $(2K + 2)$ dependence terms can be deterministically calculated from equations arising from the constraints. For example, when $K = 3$, we can estimate $\varphi_{0,0,0|d}$, $\varphi_{0,0,1|d}$, $\varphi_{0,1,1|d}$, and $\varphi_{1,1,1|d}$, and then calculate the other 4 dependence terms using the equations in 6. The choice of which terms are to be estimated and which terms are to be calculated is arbitrary.

In addition to the dependence terms to be estimated, the FEM also includes K sensitivities, K specificities, and 1 prevalence. Thus, in total, there are $2^{K+1} - (2K + 2) + 2K + 1 = 2^{K+1} - 1$ parameters to be estimated. However, for K imperfect tests, the degrees of freedom provided by the observed data set is only $2^K - 1$. Therefore, informative priors are required for at least 2^K parameters in order to reasonably estimate all parameters.[6]

The FEM we have described is flexible in allowing for the addition of further equality constraints into the model. For example, in some cases where the specificities of some tests are close to 1, the conditional dependence of these specificities will necessarily be very small and can in practice be assumed to be zero. These additional equality constraints can further decrease the number of parameters to be estimated in the model, as we will illustrate in both our simulated data and real example applied to TB data.

2.3 | Inequality constraints

The dependence terms in the FEM also have to satisfy some inequalities. Because test results are discrete, the joint probability is always bounded by any of the corresponding marginal probabilities. So, we have

$$0 \leq \mathrm{P}(T_1 = t_1, T_2 = t_2, \ldots T_K = t_K | D = d) \leq \min_{i=i \text{ to } K} \mathrm{P}(T_i = t_i | D = d).$$

By substituting the joint probability from Equation 1, we have

$$0 \leq \prod_{i=1}^{K} P(T_i = t_i | D = d) + \varphi_{t_1,t_2,\ldots,t_K|d} \leq \min_{i=i \text{ to } K} \mathrm{P}(T_i = t_i | D = d).$$

Therefore,

$$-\prod_{i=1}^{K} P(T_i = t_i | D = d) \leq \varphi_{t_1,t_2,\ldots,t_K|d} \leq \min_{i=i \text{ to } K} P(T_i = t_i | D = d) - \prod_{i=1}^{K} P(T_i = t_i | D = d). \tag{7}$$

To illustrate, when $K = 2$, after considering the relationships between the dependence terms, the inequalities become

$$-S_1 S_2 + \max(0, S_1 + S_2 - 1) \leq \varphi_{1,1|1} \leq \min(S_1, S_2) - S_1 S_2, \text{ and}$$

$$-C_1 C_2 + \max(0, C + C_2 - 1) \leq \varphi_{0,0|0} \leq \min(C_1, C_2) - C_1 C_2.$$

The above bounds on the conditional dependence terms are the same as those previously derived in Vacek[16] and Dendukuri and Joseph.[3]

The inequalities in (7) can be applied similarly when $K \geq 3$, but they cannot be as easily expressed in terms of individual dependence parameters. These inequalities define the support regions of the dependence parameters in the FEM.

3 | BAYESIAN INFERENCE

Let the unobserved latent variable $Y_{t_1,\cdots,t_K}$ be the number of true positive subjects out of the observed number of subjects, $n_{t_1\cdots t_K}$, who have testing results $[t_1, ..., t_K]$. Then, the likelihood function of the observed and latent data is given

$$\begin{aligned} L &= l\left(Y_{0,\cdots,0}, \cdots, Y_{1,\cdots,1} | \pi, S_1, C_1, \cdots, S_K, C_K, \varphi_{0,\cdots,0|0}, \cdots, \varphi_{1,\cdots,1|1}\right) \\ &= \prod_{t_1=0}^{1} \cdots \prod_{t_K=0}^{1} I\left(\varphi_{t_1,\cdots,t_K|1}\right) \left\{ \pi \left[\prod_{i=1}^{K} P(T_i = t_i | D = 1) + \varphi_{t_1,t_2,\ldots,t_K|1} \right] \right\}^{Y_{t_1,\cdots,t_K}} \\ &\quad I\left(\varphi_{t_1,\cdots,t_K|0}\right) \left\{ (1-\pi) \left[\prod_{i=1}^{K} P(T_i = t_i | D = 0) + \varphi_{t_1,t_2,\ldots,t_K|0} \right] \right\}^{n_{t_1,\cdots,t_K} - Y_{t_1,\cdots,t_K}}, \end{aligned} \tag{8}$$

where $I\left(\varphi_{t_1,\cdots,t_K|d}\right)$ is an indicator function which equals 1 when $\varphi_{t_1,\cdots,t_K|d}$ is within its domain shown in (7), and 0 otherwise.

Prior information in the form of *Beta* densities will be assumed for π, S_i and C_i, $I = 1, ..., K$. Let α_π and β_π represent the *Beta* prior distribution parameters for π. Let $(\alpha_{S_i}, \beta_{S_i})$ and $(\alpha_{C_i}, \beta_{C_i})$ represent the parameters of the *Beta* prior distribution for S_i and C_i, respectively. RStan (Version 2.9.0-3)[14] is used to generate random samples from the joint posterior density via an MCMC algorithm. In Stan, all of the conditional dependence terms are declared as constrained variables with the support ranges shown in (7). If a separate prior distribution is not defined for the conditional dependence terms, uniform prior distributions are used over the support ranges for the conditional dependence terms. The support ranges for the constrained conditional dependence terms are automatically respected by rejecting values outside of the support ranges in (7) during Stan's MCMC sampling process. By Bayes' theorem, over the region of support, the joint posterior distribution is proportional to

$$L \times \pi^{\alpha_\pi - 1} (1-\pi)^{\beta_\pi - 1} \prod_{i=1}^{K} \left[S_i^{\alpha_{S_i} - 1} (1-S_i)^{\beta_{S_i} - 1} C_i^{\alpha_{C_i} - 1} (1-C_i)^{\beta_{C_i} - 1} \right], \tag{9}$$

where L is given by (8).

No closed-form solutions for the marginal posterior distributions of the parameters are available, and inferences will be drawn by applying the No-U-Turn sampler (NUTS)[7,8] in Stan. For each model fit, 5 chains with random initial values were used, and each chain consisted of at least 50 000 iterations with half used as burn-in.

4 | SIMULATIONS

Given the large number of parameters in our models, it is clearly not possible to run exhaustive simulations covering all situations of potential interest. We carried out a series of illustrative simulations with the goal of showing that the FEM generally works well and to provide practical guidelines for applying these models. We considered 2 scenarios, one based on 3 tests and the other on 5 tests with settings motivated by our real-life example. In the 3-test scenario, we assessed how the FEM performs both when the tests are actually conditionally independent and when there is high conditional dependence among the tests. In order to carry out a typical simulation study, data can be simulated by random sampling from the appropriate multinomial distributions. However, sampling variations in the data can affect the parameter estimates in the latent class models, making it difficult to separate the effects of conditional dependence from random variation. Therefore, in addition to a series of randomly sampled data sets, we also considered the expected data set, or "mean data set", which eliminates the effect of random sampling variation. For simulations involving 5 tests, only the expected data set was considered.

4.1 | 3-test scenario

In the 3-test scenario, we considered 3 settings. These settings are motivated by the problem of modeling results from 3 microbiological tests for childhood TB with only moderate sensitivity but high specificity. For each setting, 500 random data sets were generated with the following specifications. In the first setting, we assumed the 3 tests T_1, T_2, and T_3 to be conditionally independent, with $S_1 = 0.6$, $S_2 = 0.7$, $S_3 = 0.8$, and $C_1 = C_2 = C_3 = 0.99$. We also assume the prevalence of the disease is $\pi = 0.4$ and the sample size is 1000. In this case, the expected test results are $(n_{111}, n_{110}, ..., n_{000}) = (134, 34, 58, 20, 90, 28, 44, 592)$.

In the second setting, all the parameters are the same as in the first setting except that the 3 tests are now assumed to be highly conditionally dependent. To select the values of the dependence terms amongst the sensitivities for our simulation, we first calculated the inequality constraints, based on (7) in Section 2.3, for all dependence terms. We then enumerated all possible combinations of the dependence terms $\varphi_{1,1,1|1}$, $\varphi_{0,1,1|1}$, $\varphi_{0,0,1|1}$, and $\varphi_{0,0,0|1}$. For each possible combination, we calculated the values of the other 5 dependence terms, as discussed in Section 2.2, and checked if they satisfied their inequality constraints. Among all the valid combinations, we selected a combination which produced high dependencies amongst sensitivities. The selected values for our simulation were $\varphi_{1,1,1|1} = 0.25$, $\varphi_{0,1,1|1} = -0.12$, $\varphi_{0,0,1|1} = 0.01$, and $\varphi_{0,0,0|1} = 0.16$. The pairwise conditional dependencies between S_1 and S_2, S_1 and S_3, and S_2 and S_3 are 0.17, 0.11, and 0.13, respectively, and the relative dependencies between each pair of tests (ie, the ratio of the pairwise covariance to the maximum possible covariance) among the true disease positives are 0.94, 0.92, and 0.93, respectively. Because all of the specificities are very high, the maximum possible conditional dependencies among the specificities are trivially small and are thus assumed to be zero.[18] In this setting, the expected test results are $(n_{111}, n_{110}, ..., n_{000}) = (234, 2, 2, 8, 42, 8, 48, 656)$.

In the third setting, we modified the dependence terms among the sensitivities to be $\varphi_{1,1,1|1} = 0.14$, $\varphi_{0,1,1|1} = -0.01$, $\varphi_{0,0,1|1} = 0.01$, and $\varphi_{0,0,0|1} = 0.05$, such that the covariance between S_1 and S_2 is 0.06 and the covariance between S_2 and S_3 is 0.13, but S_1 and S_3 are independent. The relative dependencies between S_1 and S_2, and S_2 and S_3 are 0.33 and 0.93, respectively. In this setting, the expected test results are $(n_{111}, n_{110}, ..., n_{000}) = (190, 2, 2, 52, 86, 8, 48, 612)$.

For each setting, each of the 500 random data sets and the "expected" data set were fit by both a CIM and a FEM. In the FEM, we modeled all possible dependence terms among the sensitivities but set the conditional dependence amongst specificities to be zero, leaving $2^{3+1} - 1 - 4 = 11$ parameters to be estimated. Because data on 3 tests can be classified into 8 possible cells and thus 7 degrees of freedom once the sample size is fixed, there is a need for informative prior distributions on a minimum of 4 parameters in order to obtain reasonable inferences. For both the CIM and FEM models, we used informative *Beta*(113.25,0.42) priors, with 95% coverage range of (0.98, 0.9999), for C_1, C_2, and C_3, and a *Beta* (54.53,36.03) prior, with 95% coverage range of (0.5, 0.7) for S_1. Non-informative *Beta*(1,1) priors were used for π, S_2, and S_3. As mentioned in Section 3, in Stan, all of the conditional dependence terms were declared as constrained variables with domains given by (7). The lower and upper limit constraints for $\varphi_{1,1,1|1}$, $\varphi_{0,1,1|1}$, $\varphi_{0,0,1|1}$, and $\varphi_{0,0,0|1}$ were recursively used to define a uniform prior over their range.

Results comparing the conditional independence and FEMs across all 3 scenarios are listed in Tables 1 and 2. When the tests are conditionally independent, posterior median estimates for the prevalence and accuracy provided by the CIM are very close to their true values and the 95% credible intervals (CrIs) are reasonably narrow, as expected. By contrast, the prevalence and sensitivity estimates under the FEM are slightly removed from their true values (Table 1), although with wide 95% CrIs that contain the true values. The quantiles of the 500 medians of the parameters and the 95% CrI coverage are listed in Web Table A-1 in the Appendix. The distributions of the medians of the prevalence and test properties, as well as the coverage probabilities of the 95% CrIs are similar for both the conditional independence and FEMs. It should be noted that because the FEM is non-identifiable, there is no guarantee that the coverage of individual parameters will be close to 95% as the joint posterior distribution will not converge to a point even with an infinite sample size.[4,6] Nonetheless, we find the coverage is high, exceeding 80% for all parameters and exceeding 95% for some parameters.

In the scenarios where there is conditional dependence, the estimates based on the CIM depart considerably from the true values, with 95% CrIs that do not contain the true values. For example, for the expected data set simulated under high conditional dependence between all 3 tests, the median (95% CrI) of π is 0.2817 (0.2541, 0.3104) although its true value is 0.4. In addition, although an informative prior was used for C_3, its estimate is also far from the true value, at 0.9424 (0.925, 0.9573). In comparison, the estimates provided by the FEM are closer to the true values. Similar results are obtained for the conditional dependence setting where only 2 pairs of tests are conditionally dependent. Table 2

TABLE 1 Posterior median (95% credible intervals) for the prevalence and test accuracy parameters obtained when fitting the conditional independence model (CIM) and fixed effects model (FEM) to the expected data set for scenarios involving 3 tests

	True value	Scenario 1: Data simulated under conditional independence		Scenario 2: Data simulated under conditional dependence between all 3 tests		Scenario 3: Data simulated under conditional dependence between 2 pairs of tests	
		CIM	FEM	CIM	FEM	CIM	FEM
π	0.4	0.414 (0.380, 0.448)	0.425 (0.380, 0.495)	0.282 (0.254, 0.310)	0.391 (0.334, 0.470)	0.282 (0.254, 0.311)	0.410 (0.363, 0.485)
S_1	0.6	0.592 (0.547, 0.637)	0.578 (0.505, 0.634)	0.786 (0.741, 0.826)	0.62 (0.528, 0.700)	0.664 (0.613, 0.711)	0.595 (0.514, 0.656)
S_2	0.7	0.686 (0.637, 0.734)	0.666 (0.573, 0.731)	0.989 (0.967, 0.999)	0.722 (0.611, 0.821)	0.992 (0.966, 0.9997)	0.693 (0.589, 0.766)
S_3	0.8	0.782 (0.736, 0.827)	0.760 (0.655, 0.825)	0.987 (0.964, 0.997)	0.823 (0.690, 0.935)	0.984 (0.955, 0.998)	0.787 (0.669, 0.861)
C_1	0.99	0.998 (0.985, 1)	0.999 (0.981, 1)	0.990 (0.981, 0.995)	0.999 (0.988, 1)	0.936 (0.917, 0.951)	0.999 (0.979, 1)
C_2	0.99	0.999 (0.985, 1)	0.998 (0.978, 1)	0.991 (0.983, 0.997)	0.999 (0.988, 1)	0.992 (0.982, 0.9997)	0.999 (0.987, 1)
C_3	0.99	0.999 (0.983, 1)	0.998 (0.977, 1)	0.942 (0.925, 0.957)	0.999 (0.980, 1)	0.942 (0.924, 0.957)	0.999 (0.978, 1)

TABLE 2 Posterior median (95% credible intervals) for the conditional dependence parameters obtained when fitting the conditional independence model (CIM) and fixed effects model (FEM) model to the expected data set for scenarios involving 3 tests

	Scenario 1: Data simulated under conditional independence			Scenario 2: Data simulated under conditional dependence between all 3 tests			Scenario 3: Data simulated under conditional dependence between 2 pairs of tests		
Parameter	True value	CIM	FEM	True value	CIM	FEM	True value	CIM	FEM
$\Phi_{111\|1}$	0	-	0.021 (−0.023, 0.087)	0.25	-	0.218 (0.128, 0.281)	0.14	-	0.133 (0.088, 0.195)
$\Phi_{011\|1}$	0	-	−0.005 (−0.031, 0.024)	−0.12	-	−0.115 (−0.135, −0.091)	−0.01	-	−0.012 (−0.041, 0.019)
$\Phi_{001\|1}$	0	-	−0.009 (−0.057, 0.023)	0.01	-	0.032 (−0.034, 0.093)	0.01	-	0.014 (−0.041, 0.052)
$\Phi_{000\|1}$	0	-	0.017 (−0.022, 0.107)	0.16	-	0.113 (0.012, 0.215)	0.05	-	0.038 (−0.013, 0.134)
$\Phi_{100\|1}$	0	-	−0.003 (−0.037, 0.023)	−0.03	-	−0.012 (−0.049, 0.017)	0.08	-	0.082 (0.034, 0.117)
$\Phi_{101\|1}$	0	-	−0.009 (−0.035, 0.019)	−0.14	-	−0.133 (−0.15, −0.104)	−0.14	-	−0.136 (−0.154, −0.114)
$\Phi_{110\|1}$	0	-	−0.012 (−0.037, 0.016)	−0.08	-	−0.072 (−0.099, −0.027)	−0.08	-	−0.081 (−0.103, −0.056)
$\Phi_{010\|1}$	0	-	−0.007 (−0.048, 0.022)	−0.05	-	−0.03 (−0.074, 0.009)	−0.05	-	−0.042 (−0.079, −0.017)
$\Phi^{T1T2}{}_{11\|1}$	0	-	0.011 (−0.023, 0.058)	0.17	-	0.146 (0.095, 0.186)	0.06	-	0.054 (0.017, 0.101)
$\Phi^{T1T3}{}_{11\|1}$	0	-	0.013 (−0.021, 0.065)	0.11	-	0.084 (0.017, 0.144)	0	-	−0.003 (−0.042, 0.059)
$\Phi^{T1T3}{}_{11\|1}$	0	-	0.017 (−0.02, 0.076)	0.13	-	0.101 (0.023, 0.169)	0.13	-	0.122 (0.079, 0.176)

shows that for both conditional dependence settings, the various dependence parameters are well estimated by the 2 FEMs.

The quantiles of the 500 medians of the parameter and the 95% CrI coverage are listed in Web Tables A-2 and A-3 in the Appendix, respectively. As we would expect, the FEM performs better than the CIM in these settings. When there is high conditional dependence between all 3 sensitivities, the medians of π and C_3 are underestimated, and all of the sensitivities are overestimated under the CIM, and the 95% CrI coverage probabilities for all of these parameters are zero. However, the medians of these parameters under the FEM are distributed near the true values, and the 95% CrI coverage probabilities for them are close to 1 for most parameters. These simulations suggest that the FEM successfully corrects for the bias observed when ignoring conditional dependence.

4.2 | 5-test scenario

We consider 3 settings for simulating and analyzing data for 5 tests, now using only the expected datasets. In the first setting, we assume the 5 tests T_1, T_2, ..., T_5 are conditionally independent. In the second and third settings, we considered

2 alternative dependence structures inspired by our applied example. We considered these 2 structures to enable study of the possible impact of the mis-specification of the conditional dependence structure. The prevalence, sample size, and the accuracies of the first 3 tests are the same as in the 3-test scenario of Section 4.1 in all 3 settings. Tests T_4 and T_5 are assumed to have $S_4 = 0.65$; $C_4 = 0.75$, and $S_5 = 0.68$; $C_5 = 0.65$, ie, these tests have moderate sensitivity and specificity. The expected dataset under conditional independence is (n_{11111}, n_{11110}, ..., n_{00000}) = (59, 28, 32, 15, 15, 7, 8, 4, 25, 12, 14, 6, 7, 4, 5, 4, 40, 19, 21, 10, 10, 6, 7, 5, 17, 9, 11, 7, 55, 97, 155, 285). The counts are rounded to the nearest integer so that the total is 999 rather than 1000.

In the second setting, we assume that the sensitivities of the first 4 tests are correlated similar to our CPTB example. We assume the vector Φ has components $\varphi^{T_1T_2T_3T_4}_{1111|1} = 0.1$, $\varphi^{T_1T_2T_3T_4}_{0111|1} = -0.1$, $\varphi^{T_1T_2T_3T_4}_{1011|1} = 0.06$, $\varphi^{T_1T_2T_3T_4}_{1101|1} = -0.05$, $\varphi^{T_1T_2T_3T_4}_{1110|1} = -0.1$, $\varphi^{T_1T_2T_3T_4}_{0011|1} = 0$, $\varphi^{T_1T_2T_3T_4}_{0101|1} = 0$, $\varphi^{T_1T_2T_3T_4}_{0110|1} = 0.11$, $\varphi^{T_1T_2T_3T_4}_{1001|1} = 0$, $\varphi^{T_1T_2T_3T_4}_{1010|1} = -0.05$, and $\varphi^{T_1T_2T_3T_4}_{1100|1} = 0.03$, but the sensitivity of T_5 is independent of the other tests. The pairwise conditional dependencies between S_1 and S_2, S_1 and S_3, S_1 and S_4, S_2 and S_3, S_2 and S_4, and S_3 and S_4 are −0.02, 0.01, 0.11, 0.01, −0.05, and 0.06, respectively, and pairwise relative conditional dependencies of these values are 0.17, 0.08, 0.53, 0.07, 0.48, and 0.46. When the pairwise conditional dependence is negative, the corresponding relative conditional dependence is defined as the ratio of the pairwise covariance to the lower limit of the possible covariance. Specificities between all 5 tests are assumed to be conditionally independent. The expected dataset is (n_{11111}, n_{11110}, ..., n_{00000}) = (87, 41, 5, 2, 1, 1, 16, 8, 42, 20, 0, 0, 7, 4, 8, 6, 12, 6, 51, 24, 10, 6, 10, 7, 17, 9, 5, 5, 52, 95, 158, 286). In the third setting, the true values of the prevalence and accuracy parameters remain the same as in the first 2 settings. We assume additionally that the sensitivities of T_1, T_2, and T_3 are correlated, with $\varphi^{T_1T_2T_3}_{111|1} = 0.25$, $\varphi^{T_1T_2T_3}_{011|1} = 0.12$, $\varphi^{T_1T_2T_3}_{001|1} = 0.01$, and $\varphi^{T_1T_2T_3}_{000|1} = 0.16$, and that the sensitivities of T_4 and T_5 as well as their specificities are correlated, with $\varphi^{T_4T_5}_{11|1} = 0.1$ and $\varphi^{T_4T_5}_{00|1} = 0.05$, with the relative conditional dependencies of 0.48 and 0.31, respectively. The pairwise conditional dependencies between S_1 and S_2, S_1 and S_3, and S_2 and S_3 are 0.17, 0.11, and 0.13, respectively, and the pairwise relative conditional dependencies of these values are 0.94, 0.92, and 0.93, respectively. The expected data set is now (n_{11111}, n_{11110}, ..., n_{00000}) = (127, 25, 32, 50, 1, 0, 0, 0, 1, 0, 0, 0, 2, 1, 2, 4, 23, 4, 6, 9, 2, 1, 2, 4, 24, 5, 7, 12, 120, 73, 134, 329).

The first FEM (FEM1) that was fit to the data sets above allows dependence between 2 clusters of tests—(T_1, T_2, and T_3) which have a high specificity and (T_4, T_5) which do not. This model allows for dependencies among S_1, S_2, and S_3, among C_1, C_2, and C_3, between S_4 and S_5, and between C_4 and C_5. The conditional joint probability can then be simplified to

$$\begin{aligned}\mathrm{P}(T_1 = t_1, \cdots, T_5 = t_5|D) &= \mathrm{P}(T_1 = t_1, \cdots, T_3 = t_3|D)\mathrm{P}(T_4 = t_4, T_5 = t_5|D)\\ &= \left[\prod_{i=1}^{3}\mathrm{P}(T_i = t_i|D) + \varphi^{T_1T_2T_3}_{t_1t_2t_3|D}\right]\left[\prod_{j=4}^{5}\mathrm{P}\left(T_j = t_j|D\right) + \varphi^{T_4T_5}_{t_4t_5|D}\right].\end{aligned}$$

The second FEM (FEM2) allows for dependence among sensitivities of the first 4 tests. The conditional joint probability of FEM2 can be written as

$$\mathrm{P}(T_1 = t_1, \cdots, T_5 = t_5|D) = \mathrm{P}(T_1 = t_1, \cdots, T_4 = t_4|D)\mathrm{P}(T_5 = t_5|D) = \left[\prod_{i=1}^{4}P(T_i = t_i|D) + \varphi^{T_1T_2T_3T_4}_{t_1t_2t_3t_4|D}\right]\mathrm{P}(T_5 = t_5|D),$$

where $\varphi^{T_1T_2T_3T_4}_{t_1t_2t_3t_4|1} = 0$ for any value of t_1, t_2, t_3, and t_4. There are 11, 21, and 22 parameters to be estimated in the CIM, FEM1, and FEM2 models, respectively. With 5 tests, there are $2^5 - 1 = 31$ degrees of freedom in the data sets. Therefore, all of the models are identifiable. The same *Beta*(113.25,0.42) prior was used for C_1, C_2, and C_3 and uniform priors were used for all the other parameters across all 3 models.

Results of fitting these 3 models to the expected data sets under conditional independence and under conditional dependence among the first 4 tests are listed in Tables 3 and 4. As with 3 tests, the best performance was observed when the true model was fit to the data. When the tests were truly conditionally independent, most estimates provided by both FEMs are close to the true values, and all of the 95% CrIs contain the true values, although the lengths of the 95% CrIs are larger than under the CIM. Only the estimate of S_3 provided by FEM2 is an underestimate by 0.08 with the 95% CrI not covering the true value. For the data set generated under the assumption that the sensitivities of the first 4 tests are dependent, the main parameters are well estimated by the CIM and the model that specifies the conditional dependence structure correctly (FEM2). However, π, S_2, S_4, and all the specificities are biased under the mis-specified conditional dependence structure (FEM1), highlighting the importance of specifying the conditional dependence structure correctly.

TABLE 3 Posterior median (95% credible intervals) for the prevalence and test accuracy parameters obtained when fitting the conditional independence model (CIM) and 2 different fixed effects models (FEM) to the expected data set for scenarios involving 5 tests

		Scenario 1: Data simulated under conditional independence			Scenario 2: Data simulated under conditional dependence between (T_1, T_2, T_3, T_4)		
Parameter	**True value**	**CIM**	**FEM1**	**FEM2**	**CIM**	**FEM1**	**FEM2**
Π	0.4	0.412 (0.379, 0.445)	0.436 (0.376, 0.530)	0.443 (0.391, 0.565)	0.400 (0.366, 0.435)	0.290 (0.236, 0.356)	0.445 (0.392, 0.57)
S1	0.6	0.589 (0.539, 0.639)	0.545 (0.444, 0.629)	0.547 (0.429, 0.619)	0.596 (0.545, 0.646)	0.701 (0.578, 0.827)	0.55 (0.43, 0.622)
S2	0.7	0.689 (0.639, 0.737)	0.637 (0.524, 0.731)	0.636 (0.501, 0.712)	0.694 (0.644, 0.741)	0.545 (0.456, 0.641)	0.635 (0.497, 0.71)
S3	0.8	0.783 (0.737, 0.828)	0.725 (0.598, 0.818)	0.722 (0.569, 0.801)	0.811 (0.76, 0.858)	0.789 (0.667, 0.911)	0.72 (0.564, 0.8)
S4	0.65	0.640 (0.592, 0.687)	0.647 (0.597, 0.696)	0.61 (0.515, 0.676)	0.661 (0.612, 0.709)	0.960 (0.919, 0.988)	0.614 (0.513, 0.681)
S5	0.68	0.671 (0.624, 0.717)	0.677 (0.628, 0.725)	0.67 (0.622, 0.715)	0.673 (0.625, 0.718)	0.687 (0.624, 0.747)	0.666 (0.618, 0.711)
C1	0.99	0.997 (0.985, 1)	0.988 (0.970, 0.998)	0.998 (0.984, 1)	0.987 (0.972, 0.999)	0.947 (0.930, 0.962)	0.998 (0.982, 1)
C2	0.99	0.998 (0.985, 1)	0.987 (0.969, 0.997)	0.998 (0.983, 1)	0.988 (0.970, 0.9998)	0.857 (0.832, 0.880)	0.998 (0.981, 1)
C3	0.99	0.998 (0.984, 1)	0.987 (0.967, 0.997)	0.998 (0.982, 1)	0.999 (0.991, 1)	0.894 (0.873, 0.914)	0.998 (0.982, 1)
C4	0.75	0.750 (0.714, 0.785)	0.774 (0.719, 0.853)	0.751 (0.699, 0.804)	0.758 (0.723, 0.792)	0.816 (0.759, 0.885)	0.756 (0.704, 0.808)
C5	0.65	0.651 (0.611, 0.69)	0.672 (0.62, 0.742)	0.67 (0.621, 0.766)	0.648 (0.609, 0.685)	0.603 (0.561, 0.648)	0.67 (0.62, 0.767)

FEM1: Fixed effects model with dependence between (T_1, T_2, and T_3) and between (T_4 and T_5) FEM2: Fixed effects model with dependence between (T_1, T_2, T_3, and T_4)

TABLE 4 Posterior median (95% credible intervals) for selected conditional dependence parameters obtained when fitting the conditional independence (CIM) and 2 different fixed effects (FEM) models to the expected data set for scenarios involving 5 tests

	Scenario 1: Data simulated under conditional independence				Scenario 2: Data simulated under conditional dependence between (T_1, T_2, T_3, T_4)			
Parameter	True value	CIM	FEM1	FEM2	True value	CIM	FEM1	FEM2
$\Phi^{T1T2T3}{}_{111\|1}$	0	-	0.049 (−0.016, 0.111)		0	-	0.147 (0.069, 0.199)	
$\Phi^{T1T2T3}{}_{011\|1}$	0	-	−0.006 (−0.031, 0.02)		0.01	-	−0.086 (−0.11, −0.057)	
$\Phi^{T1T2T3}{}_{001\|1}$	0	-	−0.027 (−0.082, 0.019)		−0.02	-	−0.019 (−0.074, 0.041)	
$\Phi^{T1T2T3}{}_{000\|1}$	0	-	0.055 (−0.018, 0.151)		0	-	0.092 (−0.002, 0.175)	
$\Phi^{T1T2T3}{}_{111\|0}$	0	-	0.001 (0.0001, 0.008)	-	0	-	0.002 (−0.001, 0.008)	-
$\Phi^{T1T2T3}{}_{011\|0}$	0	-	0.0017 (−0.0001, 0.011)	-	0	-	0.073 (0.059, 0.088)	-
$\Phi^{T1T2T3}{}_{001\|0}$	0	-	−0.006 (−0.018, −0.001)	-	0	-	−0.072 (−0.086, −0.058)	-
$\Phi^{T1T2T3}{}_{000\|0}$	0	-	0.011 (0.003, 0.027)	-	0	-	0.095 (0.079, 0.111)	-
$\Phi^{T1T2T3}{}_{100\|1}$	0	-	−0.015 (−0.054, 0.02)		0.01	-	−0.034 (−0.074, 0.012)	
$\Phi^{T1T2T3}{}_{101\|1}$	0	-	−0.015 (−0.039, 0.015)		0.01	-	−0.038 (−0.069, −0.008)	
$\Phi^{T1T2T3}{}_{110\|1}$	0	-	−0.019 (−0.043, 0.011)		−0.02	-	−0.074 (−0.095, −0.037)	
$\Phi^{T1T2T3}{}_{010\|1}$	0	-	−0.022 (−0.07, 0.018)		0.01	-	0.014 (−0.023, 0.054)	
$\Phi^{T1T2T3}{}_{100\|0}$	0	-	−0.006 (−0.018, −0.001)	-	0	-	−0.019 (−0.029, −0.011)	-
$\Phi^{T1T2T3}{}_{101\|0}$	0	-	0.002 (−0.0001, 0.01)	-	0	-	−0.004 (−0.006, −0.0003)	-
$\Phi^{T1T2T3}{}_{110\|0}$	0	-	0.002 (−0.0001, 0.009)	-	0	-	0.021 (0.013, 0.031)	-
$\Phi^{T1T2T3}{}_{010\|0}$	0	-	−0.006 (−0.018, −0.001)	-	0	-	−0.097 (−0.112, −0.081)	-
$\Phi^{T1T2}{}_{11\|1}$	0	-	0.03 (−0.018, 0.078)	0.026 (−0.013, 0.081)	−0.02	-	0.073 (0.022, 0.113)	0.008 (−0.032, 0.067)
$\Phi^{T1T3}{}_{11\|1}$	0	-	0.034 (−0.016, 0.087)	0.03 (−0.012, 0.091)	0.01	-	0.108 (0.034, 0.159)	0.039 (−0.002, 0.099)
$\Phi^{T1T4}{}_{11\|1}$	0	-	-	0.016 (−0.018, 0.057)	0.11	-	-	0.11 (0.078, 0.14)
$\Phi^{T2T3}{}_{11\|1}$	0	-	0.042 (−0.014, 0.103)	0.012 (−0.03, 0.057)	0.01	-	0.059 (−0.001, 0.108)	0.035 (−0.001, 0.074)
$\Phi^{T2T4}{}_{11\|1}$	0	-	-	0.02 (−0.016, 0.066)	−0.05		-	−0.028 (−0.065, 0.029)
$\Phi^{T3T4}{}_{11\|1}$	0	-	-	0.022 (−0.015, 0.074)	0.06		-	0.072 (0.035, 0.118)
$\Phi^{T1T2}{}_{11\|0}$	0	-	0.004 (0.0004, 0.013)	-	0	-	0.023 (0.014, 0.034)	-
$\Phi^{T1T3}{}_{11\|0}$	0	-	0.004 (0.0004, 0.014)	-	0	-	−0.002 (−0.005, 0.004)	-
$\Phi^{T2T3}{}_{11\|0}$	0	-	0.004 (0.001, 0.014)	-	0	-	0.075 (0.061, 0.09)	-
$\Phi^{T4T5}{}_{11\|1}$	0	-	−0.002 (−0.026, 0.021)	-	0	-	−0.003 (−0.015, 0.013)	-

(Continues)

TABLE 4 (Continued)

Parameter	Scenario 1: Data simulated under conditional independence				Scenario 2: Data simulated under conditional dependence between (T_1, T_2, T_3, T_4)			
	True value	CIM	FEM1	FEM2	True value	CIM	FEM1	FEM2
$\Phi^{T4T5}{}_{11\|0}$	0	-	−0.01 (−0.04, 0.013)	-	0	-	−0.022 (−0.045, 0)	-
$\Phi^{T1T2T3T4}{}_{1111\|1}$	0	-	-	0.037 (−0.008, 0.091)	0.1		-	0.124 (0.084, 0.166)
$\Phi^{T1T2T3T4}{}_{0111\|1}$	0	-	-	0.005 (−0.021, 0.031)	−0.1		-	−0.084 (−0.11, −0.049)
$\Phi^{T1T2T3T4}{}_{1011\|1}$	0	-	-	−0.004 (−0.025, 0.018)	0.06		-	0.048 (0.024, 0.074)
$\Phi^{T1T2T3T4}{}_{1101\|1}$	0	-	-	−0.009 (−0.026, 0.011)	−0.05		-	−0.052 (−0.065, −0.039)
$\Phi^{T1T2T3T4}{}_{1110\|1}$	0	-	-	0.008 (−0.019, 0.035)	−0.1		-	−0.079 (−0.105, −0.042)
$\Phi^{T1T2T3T4}{}_{0011\|1}$	0	-	-	−0.015 (−0.045, 0.011)	0		-	−0.016 (−0.044, 0.01)
$\Phi^{T1T2T3T4}{}_{0101\|1}$	0	-	-	−0.014 (−0.04, 0.009)	0		-	−0.014 (−0.04, 0.008)
$\Phi^{T1T2T3T4}{}_{0110\|1}$	0	-	-	−0.011 (−0.033, 0.013)	0.11		-	0.085 (0.047, 0.119)
$\Phi^{T1T2T3T4}{}_{1001\|1}$	0	-	-	−0.009 (−0.032, 0.01)	0		-	−0.01 (−0.033, 0.009)
$\Phi^{T1T2T3T4}{}_{1010\|1}$	0	-	-	−0.011 (−0.03, 0.011)	−0.05		-	−0.053 (−0.064, −0.042)
$\Phi^{T1T2T3T4}{}_{1100\|1}$	0	-	-	−0.01 (−0.027, 0.009)	0.03		-	0.016 (−0.007, 0.04)
$\Phi^{T1T2T3T4}{}_{0001\|1}$	0	-	-	0.005 (−0.025, 0.057)	0.01		-	0.001 (−0.027, 0.05)
$\Phi^{T1T2T3T4}{}_{0010\|1}$	0	-	-	−0.009 (−0.047, 0.016)	0.01		-	−0.026 (−0.062, −0.005)
$\Phi^{T1T2T3T4}{}_{0100\|1}$	0	-	-	−0.007 (−0.039, 0.014)	−0.02		-	0.003 (−0.032, 0.028)
$\Phi^{T1T2T3T4}{}_{1000\|1}$	0	-	-	−0.004 (−0.03, 0.015)	−0.01		-	0.006 (−0.023, 0.028)
$\Phi^{T1T2T3T4}{}_{0000\|1}$	0	-	-	0.036 (−0.009, 0.156)	0.01		-	0.04 (−0.009, 0.164)

FEM1: Fixed effects model with dependence between (T_1, T_2, and T_3) and between (T_4 and T_5)FEM2: Fixed effects model with dependence between (T_1, T_2, T_3, and T_4)

The better performance of the CIM compared with FEM2 may be attributed to the fact that the pairwise correlation between T_1 and T_4 was high, but all other pairwise correlations were negligible conditional on the true disease status. We should not conclude that the conditional independence will generally perform better than a mis-specified dependence model in other datasets. In a simulated dataset of course the true dependence structure is known. In a real data analysis, it may not be possible to distinguish between 2 models using statistical criteria alone.[2] The results obtained in these simulations highlight the need for expert opinion regarding which tests may be correlated as well as robustness checks across plausible correlation structures. We illustrate both of these issues in our TB example in Section 5. Results obtained under the simulation based on the second conditional dependence structure were similar and are provided in Web Table A-4.

5 | ANALYSIS OF THE CHILDHOOD PULMONARY TB DATA

Based on the opinion of clinical experts, among the 5 childhood diagnostic tests for TB, the sensitivities of culture, smear, Xpert, and TST tests are all functions of the severity of CPTB.[13] The conditional dependence between the image-based CXR and the remaining tests can reasonably be assumed to be negligible as they are based on different biological mechanisms. The culture, Xpert, and smear tests are all microbiological tests based on the same induced sputum sample. These tests are therefore likely to be positively correlated because children with a higher bacillary load are more likely to test positive. The correlation between TST and the microbiological tests is expected to be negative among children who have a very severe infection, but positive otherwise. In an earlier analysis of these data, the conditional dependence between the 4 tests was modeled using a normally distributed random effect to represent the bacterial load, which acts as a proxy for the severity of infection.[13] In other words, the random effect created the dependence between tests. The sensitivities of culture, Xpert, and smear were assumed to be linear functions of the random effect, while the sensitivity of TST was assumed to be a quadratic function. The assumptions regarding the probability distribution of the random effect and its association with the sensitivities of the individual tests cannot be verified from the data. It is therefore of interest to compare these earlier results with those from a FEM that does not impose these distributional assumptions.

Our primary objective was to fit a FEM assuming the sensitivities of culture, smear, Xpert, and TST are dependent, but CXR is conditionally independent of the remaining tests. Because the specificities of culture, smear, and Xpert are all close to 100%, it is also reasonable to assume that the dependence amongst the test specificities are small and can be ignored in the modelling. The observed data were as follows (n_{11111}, n_{11110}, ..., n_{00000}) = ((7, 5, 21, 8, 27, 17, 4, 1, 0, 0, 0, 0, 20, 8, 1, 3, 0, 0, 0, 0, 2, 7, 2, 5, 0, 1, 0, 0, 78, 149, 87, 296)), where T_1 = Culture, T_2 = Xpert, T_3 = Smear, T_4 = TST, and T_5 = CXR.

For comparison, we also report the results of fitting 3 alternative models, 2 of which are mis-specified based on the expert opinion:

1. A CIM assuming the 5 tests are all conditionally independent.
2. An alternative FEM assuming culture, smear, and Xpert to have dependencies on both sensitivities and specificities, and TST and CXR are also dependent on both sensitivities and specificities.
3. A normal random effects model with conditional dependence structure proposed by the experts as in the earlier analysis.[13]

Across these 4 models, informative *Beta*(238.178,0.457) priors, with 95% prior CrIs of (0.99, 0.999999), were used for the specificity of culture, and a *Beta*(113.249,0.419) prior, with 95% prior CrI of (0.98, 0.999999), was used for the specificity of Xpert, matching Schumacher et al's analysis. Uniform priors were used for all remaining parameters. The posterior estimates of these parameters from all models are listed in Tables 5 and 6.

The model that is in keeping with expert opinion gives similar results to those reported earlier by Schumacher et al.[13] Of note, the point estimates of the sensitivity of culture, Xpert, and smear are higher under the new model, while the point estimate for TST sensitivity is lower. This suggests that the shrinkage of the sensitivities is greater when using a normal random effects model than with our new model that does not impose any distributional assumptions on the random effects. The 95% CrIs from both models overlap considerably. Interestingly, Table 6 shows that there is a strong pairwise dependence between the 3 microbiological tests.

The models that are not in keeping with expert opinion would lead us to noticeably different results. Both the CIM and the model assuming an alternative fixed effects structure estimate that prevalence is considerably lower and culture

TABLE 5 Posterior median and 95% credible intervals for the prevalence and accuracy parameters for the childhood pulmonary tuberculosis tests dataset

Parameter	Conditional independence model	Fixed effects model with alternative dependence structure	Fixed effects model with dependence structure based on expert opinion	Random effects model with dependence structure based on expert opinion[a]
CPTB prevalence	0.166 (0.139, 0.196)	0.192 (0.142, 0.318)	0.227 (0.167, 0.316)	0.25 (0.193, 0.341)
Sensitivity of culture	0.971 (0.89, 0.999)	0.785 (0.485, 0.943)	0.689 (0.507, 0.875)	0.641 (0.482, 0.788)
Specificity of culture	0.999 (0.989, 1)	0.989 (0.972, 0.997)	0.997 (0.986, 1)	0.996 (0.987, 1)
Sensitivity of Xpert	0.744 (0.659, 0.823)	0.619 (0.386, 0.766)	0.572 (0.42, 0.731)	0.519 (0.391, 0.64)
Specificity of Xpert	0.982 (0.969, 0.992)	0.977 (0.956, 0.992)	0.988 (0.973, 1)	0.984 (0.971, 0.994)
Sensitivity of smear	0.335 (0.254, 0.424)	0.256 (0.145, 0.367)	0.267 (0.182, 0.376)	0.238 (0.166, 0.317)
Specificity of smear	0.998 (0.992, 0.9998)	0.989 (0.976, 0.997)	1 (0.995, 1)	0.997 (0.99, 1)
Sensitivity of TST	0.69 (0.604, 0.768)	0.714 (0.62, 0.811)	0.707 (0.599, 0.798)	0.739 (0.592, 0.827)
Specificity of TST	0.623 (0.585, 0.662)	0.642 (0.598, 0.708)	0.659 (0.603, 0.726)	0.678 (0.621, 0.734)
Sensitivity of CXR	0.655 (0.566, 0.739)	0.658 (0.565, 0.752)	0.647 (0.559, 0.733)	0.661 (0.574, 0.743)
Specificity of CXR	0.731 (0.696, 0.765)	0.747 (0.704, 0.823)	0.76 (0.713, 0.815)	0.777 (0.731, 0.839)

[a]We did not include any covariates as was previously done in the work of Schumacher et al (2016). The results are nonetheless very similar.

sensitivity is considerably higher. This suggests that these models place greater weight on culture, which is strongly in agreement with Xpert and Smear, and they place less weight on the complementary information available from TST.

Web Table A-5 in the Appendix presents the predicted test results given by each model. We report the sum of squared errors (SSE) as a guide to how closely the predictions under each model match the observed data. The CIM has the largest SSE value, the alternative FEM has the smallest SSE value, and the other 2 models have similar SSE values, although the differences among the last 3 models are not large. These figures should be interpreted keeping in mind that models with more parameters will in general have smaller SSE values. This explains in part why the model based on expert opinion has a larger SSE than the so-called "mis-specified" model as the latter has 2 additional parameters.

6 | DISCUSSION

This paper introduces a general FEM for multiple diagnostic tests with higher order correlations, an extension of the FEM for 2 tests originally introduced by Vacek.[16] In this more general FEM, the conditional dependence terms among multiple tests is defined as the departure of the observed conditional joint probability from the expected conditional joint probability under the assumption of conditional independence. In the special case when 2 tests are considered, this conditional dependence term becomes the covariance between the sensitivities or the specificities of the 2 tests. Although the conditional dependence terms among multiple tests may not have a straightforward meaning, we can use them to calculate and draw inferences about the pairwise covariance between any pair of tests. Although it was previously hypothesized[15] that higher order terms may be small in magnitude and therefore ignored, our simulations and real-life application suggest otherwise.

This FEM is very flexible in incorporating additional assumptions about the dependence structure. For example, as our applied example illustrates, some covariance terms can be fixed a priori to be 0. It is also possible to model a clustered conditional dependence structure among different sets of tests. The simulations show that the fixed effect model provides better inferences compared with the CIM when higher-order conditional dependencies truly exist in the data. In addition, when the tests are truly conditionally independent, the fixed effect model can still provide reasonable estimates, although it will result in wider credible results from the need to estimate a larger number of parameters from the same data.

There is no easy way to check the dependence structure in these complex and non-identifiable models. Therefore, discussions with subject matter experts to understand the source and likely form of dependence is important, as is checking the robustness of estimates across a series of plausible models, as we have done in our analysis of the CPTB

TABLE 6 Posterior median and 95% credible intervals for the conditional dependence parameters for the childhood pulmonary tuberculosis dataset

Parameter	Conditional independence model	Fixed effects model with alternative dependence structure	Fixed effects model with dependence structure based on expert opinion
$\Phi^{Culture,Xpert,Smear}{}_{111\|1}$	-	0.105 (0.056, 0.153)	
$\Phi^{Culture,Xpert,Smear}{}_{111\|0}$	-	0.006 (0.0002, 0.019)	
$\Phi^{Culture,Xpert,Smear}{}_{110\|1}$	-	−0.029 (−0.098, 0.057)	
$\Phi^{Culture,Xpert,Smear}{}_{110\|0}$	-	0.001 (−0.0004, 0.006)	
$\Phi^{Culture,Xpert,Smear}{}_{101\|1}$	-	−0.066 (−0.096, −0.037)	
$\Phi^{Culture,Xpert,Smear}{}_{101\|0}$	-	0.001 (−0.0003, 0.004)	
$\Phi^{Culture,Xpert,Smear}{}_{100\|1}$	-	−0.007 (−0.123, 0.08)	
$\Phi^{Culture,Xpert,Smear}{}_{100\|0}$	-	−0.008 (−0.021, −0.001)	
$\Phi^{Culture,Xpert,Smear}{}_{011\|1}$	-	−0.026 (−0.046, −0.001)	
$\Phi^{Culture,Xpert,Smear}{}_{011\|0}$	-	0.001 (−0.001, 0.004)	
$\Phi^{Culture,Xpert,Smear}{}_{010\|1}$	-	−0.051 (−0.138, 0.026)	
$\Phi^{Culture,Xpert,Smear}{}_{010\|0}$	-	−0.008 (−0.021, −0.001)	
$\Phi^{Culture,Xpert,Smear}{}_{001\|1}$	-	−0.012 (−0.048, 0.018)	
$\Phi^{Culture,Xpert,Smear}{}_{001\|0}$	-	−0.008 (−0.02, −0.001)	
$\Phi^{Culture,Xpert,Smear}{}_{000\|1}$	-	0.089 (−0.01, 0.213)	
$\Phi^{Culture,Xpert,Smear}{}_{000\|0}$	-	0.015 (0.003, 0.04)	
$\Phi^{CXR,TST}{}_{11\|1}$	-	−0.025 (−0.066, 0.022)	
$\Phi^{CXR,TST}{}_{11\|0}$	-	0.021 (−0.002, 0.042)	
$\Phi^{Culture,Xpert,Smear,TST}{}_{1111\|1}$	-	-	−0.005 (−0.064, 0.034)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0111\|1}$	-	-	−0.027 (−0.04, −0.009)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1011\|1}$	-	-	−0.049 (−0.067, −0.03)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1101\|1}$	-	-	0.039 (−0.004, 0.085)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1110\|1}$	-	-	0.127 (0.086, 0.175)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0011\|1}$	-	-	−0.016 (−0.036, 0.011)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0101\|1}$	-	-	−0.057 (−0.101, 0.001)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0110\|1}$	-	-	−0.009 (−0.018, 0.007)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1001\|1}$	-	-	0 (−0.065, 0.088)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1010\|1}$	-	-	−0.017 (−0.032, −0.003)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1100\|1}$	-	-	−0.052 (−0.107, −0.012)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0001\|1}$	-	-	0.115 (−0.003, 0.207)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0010\|1}$	-	-	−0.006 (−0.016, 0.011)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0100\|1}$	-	-	−0.018 (−0.045, 0.017)
$\Phi^{Culture,Xpert,Smear,TST}{}_{1000\|1}$	-	-	−0.043 (−0.072, −0.015)
$\Phi^{Culture,Xpert,Smear,TST}{}_{0000\|1}$	-	-	0.008 (−0.027, 0.096)
$\Phi^{Culture,Xpert}{}_{11\|1}$	-	0.072 (−0.01, 0.17)	0.112 (0.021, 0.166)
$\Phi^{Culture,Xpert}{}_{11\|0}$	-	0.007 (0.001, 0.02)	-
$\Phi^{Culture,Smear}{}_{11\|1}$	-	0.039 (−0.007, 0.085)	0.056 (0.006, 0.09)
$\Phi^{Culture,Smear}{}_{11\|0}$	-	0.007 (0.001, 0.02)	-
$\Phi^{Smear,Xpert}{}_{11\|1}$	-	0.08 (0.043, 0.114)	0.087 (0.052, 0.117)

(Continues)

TABLE 6 (Continued)

Parameter	Conditional independence model	Fixed effects model with alternative dependence structure	Fixed effects model with dependence structure based on expert opinion
$\Phi^{Smear,Xpert}{}_{11\|0}$	-	0.007 (0.001, 0.02)	-
$\Phi^{Culture,TST}{}_{11\|1}$	-	-	−0.016 (−0.061, 0.047)
$\Phi^{Xpert,TST}{}_{11\|1}$	-	-	−0.052 (−0.091, 0.004)
$\Phi^{Smear,TST}{}_{11\|1}$	-	-	−0.098 (−0.139, −0.057)

data. Examination of the posterior predicted values has also been shown to be useful for comparison with experts' prior beliefs to determine the appropriateness of the dependence structure.[2]

The analysis of the childhood TB data shows that when the model is based on experts' opinion, the pairwise dependencies between the sensitivities of any 2 of culture, Xpert, and smear are positive. When the dependence terms among these tests are not included in the model, the estimates of some parameters change greatly. The dependence conditioning on disease negative between culture, smear, and Xpert was small and negligible.

An important practical limitation to our model is that the number of estimated parameters in the full FEM increases exponentially with the number of diagnostic tests. Therefore, considerable prior information is needed to draw reasonable inferences. This is not a deficiency of the model itself, but rather reflects the difficulty of the problem. In the application of the proposed model, close consultation with clinicians familiar with the test properties becomes increasingly important in order to gather sufficient prior information for reasonable estimation. For example, the prior knowledge that 1 test is conditionally independent of the remaining tests in a data set could reduce the number of parameters dramatically. In our application, we found that the number of parameters being estimated and the number of informative priors required could be reduced by making appropriate assumptions based on experts' opinions on the dependence structure.

ACKNOWLEDGEMENTS

This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canadian Institutes of Health Research (CIHR). Nandini Dendukuri is supported by a salary award from the Fonds de Recherche Québec—Santé. Funding for the childhood tuberculosis study was from the National Institutes of Health (R01HD058971); Author Heather Zar is supported by the Medical Research Council South Africa.

ORCID

Zhuoyu Wang http://orcid.org/0000-0002-8940-8226
Nandini Dendukuri http://orcid.org/0000-0002-2330-0976

REFERENCES

1. Berkvens D, Speybroeck N, Praet N, Adel A, Lesaffre E. Estimating disease prevalence in a Bayesian framework using probabilistic constraints. *Epidemiology*. 2006;17:145-153.
2. Dendukuri N, Hadgu A, Wang L. Modeling conditional dependence between diagnostic tests: a multiple latent variable model. *Stat Med*. 2009;28:441-461.
3. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*. 2001;57:158-167.
4. Dendukuri N, Rahme E, Bélisle P, Joseph L. Bayesian sample size determination for prevalence and diagnostic test studies in the absence of a gold standard test. *Biometrics*. 2004;60:388-397.
5. Espeland MA, Handelman SL. Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics*. 1989;45:587-599.
6. Gustafson P. On model expansion, model contraction, identifiability, and prior information: two illustrative scenarios involving mismeasured variables. *Statistical Science*. 2005;20:111-140.
7. Hoffman MD, Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv*. 2011;1111.4246.

8. Hoffman MD, Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*. 2014;15:1593-1623.
9. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics*. 1980;36:167-171.
10. 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;141:263-272.
11. Morton JM, McCoy RJ, Kann RK, Gardner IA, Meers J. Validation of real-time polymerase chain reaction tests for diagnosing feline immunodeficiency virus infection in domestic cats using Bayesian latent class models. *Prev Vet Med*. 2012;104:136-148.
12. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating acmracy of diagnostic tests. *Biornetrics*. 1996;52:797-810.
13. Schumacher S, van Smeden M, Dendukuri N, et al. Diagnostic test accuracy in childhood pulmonary tuberculosis: a Bayesian latent class analysis. *Am J Epidemiol*. 2016;184(9):690-700.
14. Stan Development Team. Stan modeling language user's guide and reference manual, version 2.9.0-3. URL http://mc-stan.org/. 2015.
15. Torrance-Rynard VL, Walter SD. Effects of dependent errors in the assessment of diagnostic test performance. *Stat Med*. 1997;16:2157-2175.
16. Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*. 1985;41:953-968.
17. Walter SD, Irwig LM. Estimation of error rates, disease prevalence, and relative risk misclassified data: a review. *J Clin Epidemiol*. 1998;41:923-937.
18. Wang Z, Dendukuri N, Joseph L. Understanding the effects of conditional dependence in research studies involving imperfect diagnostic tests. *Stat Med*. 2017;36:466-480.
19. Xu H, Craig BA. A probit latent class model with general correlation structures for evaluating accuracy of diagnostic tests. *Biometrics*. 2009;65:1145-1155.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Wang Z, Dendukuri N, Zar HJ, Joseph L. Modeling conditional dependence among multiple diagnostic tests. *Statistics in Medicine*. 2017;1–17. https://doi.org/10.1002/sim.7449