Mixed latent Markov models for longitudinal multiple diagnostics data with an application to *Salmonella* in Malawi

Marc Y. R. Henrion[*,†] Angeziwa Chirambo[*,‡] Ndaru Jambo[*]
Tonney S. Nyirenda[*,§] Melita A. Gordon[*,¶]

Abstract

Latent Markov models (LMMs) are commonly used to analyze longitudinal data from multiple diagnostic tests. LMMs consist of a structural model for the latent infection state, defining probabilities for the initial state and transmission between states, and a measurement model for the observed test results, defining the item response probabilities and thus test sensitivities and specificities. LMMs typically assume that tests are independent conditional on the latent infection state. This is likely to be violated for tests using similar technologies. We introduce random effects to relax the conditional independence assumption and we derive a generalization of the basic LMM for an application to *Salmonella* infection data. We analyze longitudinal data from four molecular PCR tests and a stool culture test from patients in Blantyre, Malawi. To assess the tests' performances, we consider basic and mixed LMMs, both with time homogeneous and heterogeneous transition probabilities. We compare the different models and discuss technical considerations. A PCR assay using primers from the TTR gene achieves the best sensitivity / specificity trade-off.

Key Words: latent Markov model, latent variable model, mixed model, random effects, Bayesian modelling, *Salmonella*

1. Introduction

Latent variable models represent a large class of statistical models. As the name suggests they involve an unobserved (possibly unobservable) variable that the model tries to infer from the observed data. Such models can be very useful in the context of medical diagnostic tests. Latent Class Analysis (LCA; Lazarsfeld and Henry, 1968), a type of mixture model, has been widely used to estimate sensitivities and specifities of multiple diagnostic tests in situations where there is no perfect / gold standard reference test (Pan-ngum et al., 2013; van Smeden et al., 2014). Latent Markov models (LMMs; Bartolucci et al., 2013) represent an extension of LCA to longitudinal data. LMMs consist of a structural model, describing the states of the unobserved variable and the transitions between them, and a measurement model, describing the conditional distributions of the observed outcome variables. Figure 1 shows a representation of a basic LMM.

In addition to a first-order Markov assumption on the latent state (the distribution of the state at time t is completely specified by knowing the state at time $t-1$), a key assumption made by LMMs, and shared with LCA, is that, conditional on the unobserved state, the outcome variables are independent.

The models described later in this paper attempt to relax this conditional independence assumption by introducing subject-level random effects.

[*]Malawi - Liverpool - Wellcome Trust Clinical Research Programme, P.O. Box 30096, Blantyre, Malawi
[†]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK
[‡]London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, UK
[§]College of Medicine, University of Malawi, Private Bag 360, Blantyre, Malawi
[¶]University of Liverpool, Liverpool L69 3BX, UK

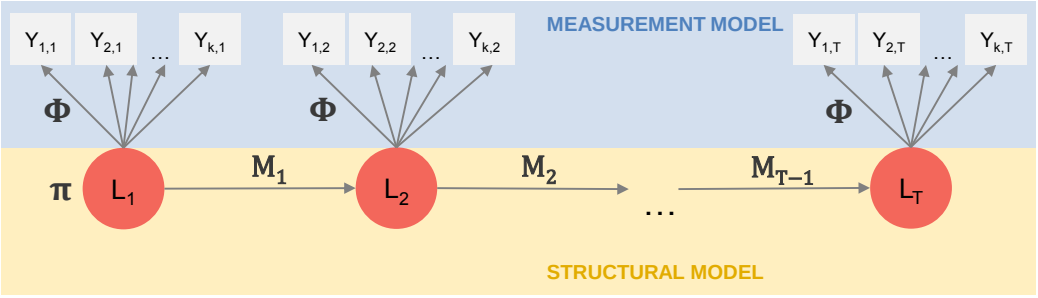


Figure 1: Graphical representation of the basic latent Markov model. Observed outcomes $Y_{k,t}$ depend on latent states L_t. The model is completely specified by a set of initial state probabilities $\boldsymbol{\pi}$, a set of $T-1$ transition probability matrices ($\boldsymbol{M}_t$, $t = 1, \ldots, T-1$) and conditional response probabilities ($\boldsymbol{\Phi}$). Highlighted in orange are the components of the structural model ($\boldsymbol{\pi}$, $\boldsymbol{M}$) and in blue the components of the measurement model ($\boldsymbol{\Phi}$).

1.1 Notation

We will consider a dataset D to consist of observations $\{\boldsymbol{y}_{t,i}\}_{i=1,\ldots,n;t=1,\ldots,T}$ for n individuals observed at T. Each observation $\boldsymbol{y}_{t,i} = (y_{1,t,i}, \ldots, y_{k,t,i})^T$ consists of responses for K variables ($k = 1, \ldots, K$). We write $l_{t,i}$ for the latent state of the i^{th} individual at time t. For our application the outcome variables are binary, but they could be categorical or continuous. Similarly, in our application the latent state variable is binary, but could be categorical. We will use lower case letters for observed values, upper case letters for random variables (i.e. $Y_{k,t,i}$, $L_{t,i}$) and boldface for vectors or matrices.

2. Model

2.1 Basic latent Markov model

Figure 1 shows a graphical representation of the basic LMM. This model is completely specified by the set of initial state $\boldsymbol{\pi}$, state transition $\boldsymbol{M}_t$, $t = 1, \ldots, T$ and conditional response probabilities $\boldsymbol{\Phi}_k$, $k = 1, \ldots, K$. For a binary latent state variable, the matrix $\boldsymbol{M}_t$ is the 2×2 matrix with transition probabilities from time t to time $t+1$ between the two states, $t = 1, \ldots, T-1$:

$$\boldsymbol{M}_t = \begin{bmatrix} 1 - \tau_{0,1}^{(t)} & \tau_{0,1}^{(t)} \\ 1 - \tau_{1,1}^{(t)} & \tau_{1,1}^{(t)} \end{bmatrix}$$

If $\boldsymbol{M}_t = \boldsymbol{M}_1 = \boldsymbol{M}$, for all $t = 1, \ldots, T-1$, then the LMM is said to have time homogeneous transition probabilities, otherwise the LMM is said to be time heterogenous. For a time homogeneous LMM, we drop the superscript $^{(t)}$ on the $\tau_{i,j}^{(t)}$. It also usually makes sense, in the time homogeneous LMM, to set $\boldsymbol{\pi}$ to the stationary distribution implied by $\boldsymbol{M}$. In the binary case where $\boldsymbol{\pi} = (1 - \pi_1, \pi_1)^T$, this is given by

$$\pi_1 = \frac{\tau_{0,1}}{(1 + \tau_{0,1} - \tau_{1,1})}$$

For binary outcome variables, the matrix $\boldsymbol{\Phi}$ is the $K \times 2$ matrix with the conditional response probabilities (CRPs), i.e. the probabilities of the outcomes conditional on the

latent state at time t:

$$\mathbf{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1^T \\ \dots \\ \boldsymbol{\phi}_K^T \end{bmatrix} = \begin{bmatrix} \phi_{1,0} & \phi_{1,1} \\ \dots & \\ \phi_{K,0} & \phi_{K,1} \end{bmatrix}$$

Specifically, for a binary latent state and binary outcomes, the basic LMM is given by:

$$\begin{aligned} L_{1,i} &\sim Bernoulli(\pi_1) \\ L_{t,i} &\sim Bernoulli(\tau^{(t-1)}_{l_{t-1,i},1}), \qquad t = 2, \dots, T \\ Y_{k,t,i} &\sim Bernoulli(\phi_{k,l_{t,i}}), \qquad k = 1, \dots, K, t = 1, \dots, T \end{aligned} \tag{1}$$

where $i = 1, \dots, n$.

In the basic LMM, the K outcome variables $Y_{k,t,i}$, for given t and i, are assumed independent conditional on $L_{t,i}$, the latent state variable at time t. In other words:

$$\begin{aligned} P(\boldsymbol{Y}_{t,i}|L_{t,i} = l_{t,i}) &= \prod_{k=1,\dots,K} P(Y_{k,t,i} = 1|L_{t,i} = l_{t,i}) \\ &= \prod_{k=1,\dots,K} \phi_{k,l_{t,i}} \end{aligned}$$

This conditional independence assumption (CIA), is shared with LCA models and considerably simplifies the expression for the likelihood of the LMM, see, e.g., Bartolucci et al. (2013).

2.2 Mixed latent Markov model

There have been a number of extensions to the basic LMM. For instance, covariates can be added to both the structural and measurement models (Bartolucci et al., 2013), and mixed latent Markov models with random effects in the structural model have been developed (Altman, 2007; Bartolucci et al., 2013; de Haan-Rietdijk et al., 2017; Koukounari et al., 2013b). There are a number of computational packages that can be used to fit LMMs, e.g. the R package `LMest`(Bartolucci et al., 2017) or the general purpose latent variable modelling software Mplus (Muthén and Muthén, 1998-2017). Overall LMMs remain quite popular, in particular for medical diagnostic test applications (Koukounari et al., 2013a,b).

Here we take a slightly different approach to generalise the basic LMM - one which, although described by some authors, has received less attention than the above extensions and for which little model fitting software exists. Specifically, we are concerned with situations where the CIA is unlikely to hold.

For example, the dataset we describe below has 5 outcome variables, each a different diagnostic test for *Salmonella*. Four of these tests use the same molecular technology, whereas the fifth is a stool culture based test (this is the reference diagnostic test, which, while highly specific, is known to be of low sensitivity). Further, two pairs of the molecular tests each share the same primers (DNA sequences that the tests detect). It seems unwise to assume that the molecular tests are — even conditionally — as independent of each other as they are compared to the stool culture test.

This example motivated us to introduce a subject-level random effect into the measurement model to capture the extra dependencies between the outcome variables and thereby relax the CIA. This is similar to Altman (2007), though the likelihood optimisation proposed therein would be too cumbersome for the models discussed below. Instead, here we adopt a Bayesian approach. Very similar models, also using a Bayesian approach, have been described by de Haan-Rietdijk et al. (2017), but the authors in that paper did not discuss the case where outcome variables share the same random effect to relax the CIA.

In the basic LMM, the CRPs $\phi_{k,l_{t,i}}$ are parameters that are directly estimated. To be able to introduce a subject-specific random effect (or adjust for covariates, if available), we follow Altman (2007) and model the logits of the CRPs instead through:

$$\phi_{k,l_{t,i},i} = \frac{exp(\alpha_{k,l_{t,i}} + \beta_k \cdot Z_i)}{1 + exp(\alpha_{k,l_{t,i}} + \beta_k \cdot Z_i)} \tag{2}$$

where $Z_i \sim \mathcal{N}(0, \sigma^2)$ is a subject-specific random effect. To keep the model identifiable, σ^2 is fixed to 1. From (2), it is clear that further random effects or covariates could be trivially added into the measurement model by simply adding terms to the expression $\alpha_{k,l_{t,i}} + \beta_k \cdot Z_i$. Likewise, the structural model can be extended through a similar expression for the transition probabilities (Altman, 2007; Bartolucci et al., 2013) and that extension is commonly used (e.g. Koukounari et al., 2013b).

With the above random effect added into the measurement model, and the CIA relaxed, we can now consider four models for the data described in the section 3. Table 1 lists the four models and their characteristics. For most of this paper, we focus on the two time homogeneous LMMs (models 1 and 3).

Table 1: Traditional and mixed latent Markov models for a binary latent state variable and binary outcome variables. For the time homogeneous models, we assume the initial state probabilities have been fixed to the stationary distribution.

model	transition probabilities	subject-level random effect	# parameters
1	time homogeneous	no	$2 \cdot (k+1)$
2	time heterogeneous	no	$2 \cdot (k+T-1)+1$
3	time homogeneous	yes	$3 \cdot k + 2$
4	time heterogeneous	yes	$3 \cdot k + 2 \cdot (T-1) + 1$

2.3 Implementation

Through the availability of R (R Core Team, 2018) and JAGS (Plummer, 2003, 2016), implementing the above models is relatively straightforward and our implementation is available from the first author's website (Henrion, 2018a) and GitHub (Henrion, 2018b).

2.3.1 Prior distributions

We used non-informative priors for all parameters. Specifically, for all probability parameters we used a beta(0.5,0.5) prior distribution (i.e. the Jeffrey's prior) and for the intercept and slope coefficients from (2) we used a Cauchy(0,10000) distribution.

3. Data

We evaluated the LMMs on several simulated datasets as well as a *Salmonella* dataset from Malawi. Table 2 summarises the different datasets we used and their characteristics.

3.1 Simulations

The simulations we ran are not meant to be a comprehensive evaluation of the LMMs listed in Table 1, but rather a tool to identify any convergence or identifiability issues of the

Table 2: Simulated and real datasets used in this paper.

dataset	n	T	k	CIA	transition probabilities
Sim1	75	8	5	cond. independent	time homogeneous
Sim2	75	8	5	cond. dependent	time homogeneous
Salmonella data	60	13	5	likely cond. dependent	likely time heterogeneous

models and to get a feel for how the basic and mixed LMMs compare in different scenarios. We focus only on the time homogeneous case. We will perform a fuller set of simulations in future work.

We generated two datasets, one (Sim1) for which the outcome variables are indepencent conditionally on the underlying latent state and another (Sim2) for which this conditional independence does not hold. Both datasets were simulated using the model desribed by (1) with $\phi_{k,l_{t,i}}$ raplaced with $\phi_{k,l_{t,i},i}$ for Sim2. Table 3 lists the parameter values used to generate the simulate data.

Table 3: Parameter values used to generate the simulated datasets. π_1 was not chosen, but computed to be the stationary distribution probability implied by $\tau_{0,1}$ and $\tau_{1,1}$.

Parameter	Sim1			Sim2		
π_1	$0.\overline{11}$			$0.\overline{11}$		
$\tau_{0,1}$	0.1			0.1		
$\tau_{1,1}$	0.2			0.2		
$\alpha_{1,0}, \alpha_{1,1}, \beta_1$	$\ln\frac{0.02}{1-0.02}$,	$\ln\frac{0.6}{1-0.6}$,	0	$\ln\frac{0.02}{1-0.02}$,	$\ln\frac{0.6}{1-0.6}$,	0
$\alpha_{2,0}, \alpha_{2,1}, \beta_2$	$\ln\frac{0.05}{1-0.05}$,	$\ln\frac{0.95}{1-0.95}$,	0	$\ln\frac{0.05}{1-0.05}$,	$\ln\frac{0.95}{1-0.95}$,	1.5
$\alpha_{3,0}, \alpha_{3,1}, \beta_3$	$\ln\frac{0.1}{1-0.1}$,	$\ln\frac{0.9}{1-0.9}$,	0	$\ln\frac{0.1}{1-0.1}$,	$\ln\frac{0.9}{1-0.9}$,	0.65
$\alpha_{4,0}, \alpha_{4,1}, \beta_4$	$\ln\frac{0.2}{1-0.2}$,	$\ln\frac{0.95}{1-0.95}$,	0	$\ln\frac{0.2}{1-0.2}$,	$\ln\frac{0.95}{1-0.95}$,	1.75
$\alpha_{5,0}, \alpha_{5,1}, \beta_5$	$\ln\frac{0.05}{1-0.05}$,	$\ln\frac{0.75}{1-0.75}$,	0	$\ln\frac{0.05}{1-0.05}$,	$\ln\frac{0.75}{1-0.75}$,	-1

3.2 Malawi *Salmonella* dataset

We have described this dataset in detail elsewhere (Chirambo et al., 2018; Nyirenda, 2015). In order to detect brief episodes of asymptomatic *Salmonella* stool carriage, monthly stool samples were collected from a cohort of 60 healthy children, recruited at 6 months of age, and followed up to 18 months of age. For the purpose of model fitting, this yielded data on 421 stool samples in total. The monthly visit times were recorded and these range from December 2013 to December 2014. For each child, the final dataset used for fitting the LMMs contains between 3 and 11 recorded stool samples.

Note that by study design, each child should have had up to 13 stool samples taken. Due to study withdrawals, not all children had in fact the full 13 stool samples taken and for some stool samples no data was available for analysis. This means that, unlike the simulated datasets, the *Salmonella* dataset contains a high proportion of missing values compared to the theoretical maximum number of participant-visit data points: $n \cdot T = 60 \cdot 13 = 780$ but

data was available for only 421 stool samples, which translates to 46% missing values. For every recorded participant-visit sample, data from all 5 tests were available.

Five methods for *Salmonella* detection in stool samples were used: stool culture, real time polymerase chain reaction (RT-PCR) using TTR and InvA primers tested as single PCR assay (TTR and InvA) or as part of Taqman Array Card assay (TTR-TAC and InvA-TAC). Stool culture was done on neat stool samples on the day of collection. All *Salmonella* isolates were stored at $-70°$C. DNA extraction was done on frozen neat samples for TTR-TAC and InvA-TAC assay and on selenite sub cultured samples for TTR and InvA RT-PCR test.

4. Results

We first use the simulated datasets to check model convergence, identifiability and performance. We then fit the LMMs to the Malawi *Salmonella* data.

4.1 Simulations

We simulated 2 datasets, both using identical parameters with the only difference between them being whether or not the CIA holds. We then fitted both the basic and mixed LMM to each dataset, allowing us to evaluate how the mixed LMM compares to the basic LMM when the CIA, in fact, holds and how the basic LMM fares when the CIA does not hold.

4.1.1 Convergence & identifiability

A first important check is that the mixed LMM is identifiable and that the MCMC chains of our Bayesian implementations of both LMMs converge. While we were careful with the parameterisation given above to insure that the model is theoretically identifiable (hence the constraint $\sigma^2 = 1$ for the random effect variance), it is still important to check this in practice. Also, since LMMs are a type of mixture model, identifiability is only guaranteed up to permutation of the latent states. This is a well-kown issue with mixture models and is particularly relevant for Bayesian models where it leads to label switching in the MCMC chains (Jasra et al., 2005). We observe such label switching also in our implementations of both the basic and the mixed LMMs. This can, however, be trivially fixed (at least in the binary case) after the MCMC algorithm has run as long as the probability parameters for the different latent states are sufficiently distinct. For most diseases and infections, where the proportion of infected is much less than the non-infected, this will be the case.

Figure 2 shows trace plots for the CRPs, transition and initial state probabilities of the mixed LMM fitted to dataset Sim2 after re-labelling latent states in chains that exhibited label switching. The chains show good mixing with no evidence of a lack of convergence. We have obtained similar results for the other LMMs fitted to both simulated datasets.

Further, the Gelman-Rubin potential scale reduction factors (Gelman and Rubin, 1992) converged to 1 for all parameters in all models after state re-labelling in chains with switched states.

Taken together, this suggests the models are identifiable and converge.

4.1.2 Comparison of basic and mixed LMMs

Regarding model fit, we want to investigate primarily whether there is a benefit in fitting the mixed LMM in situations where the CIA does not hold. However, we would also like to find out whether, in situations where the CIA does, in fact, hold, there is an appreciable disadvantage to fit the more complex mixed LMM to the simpler basic LMM.

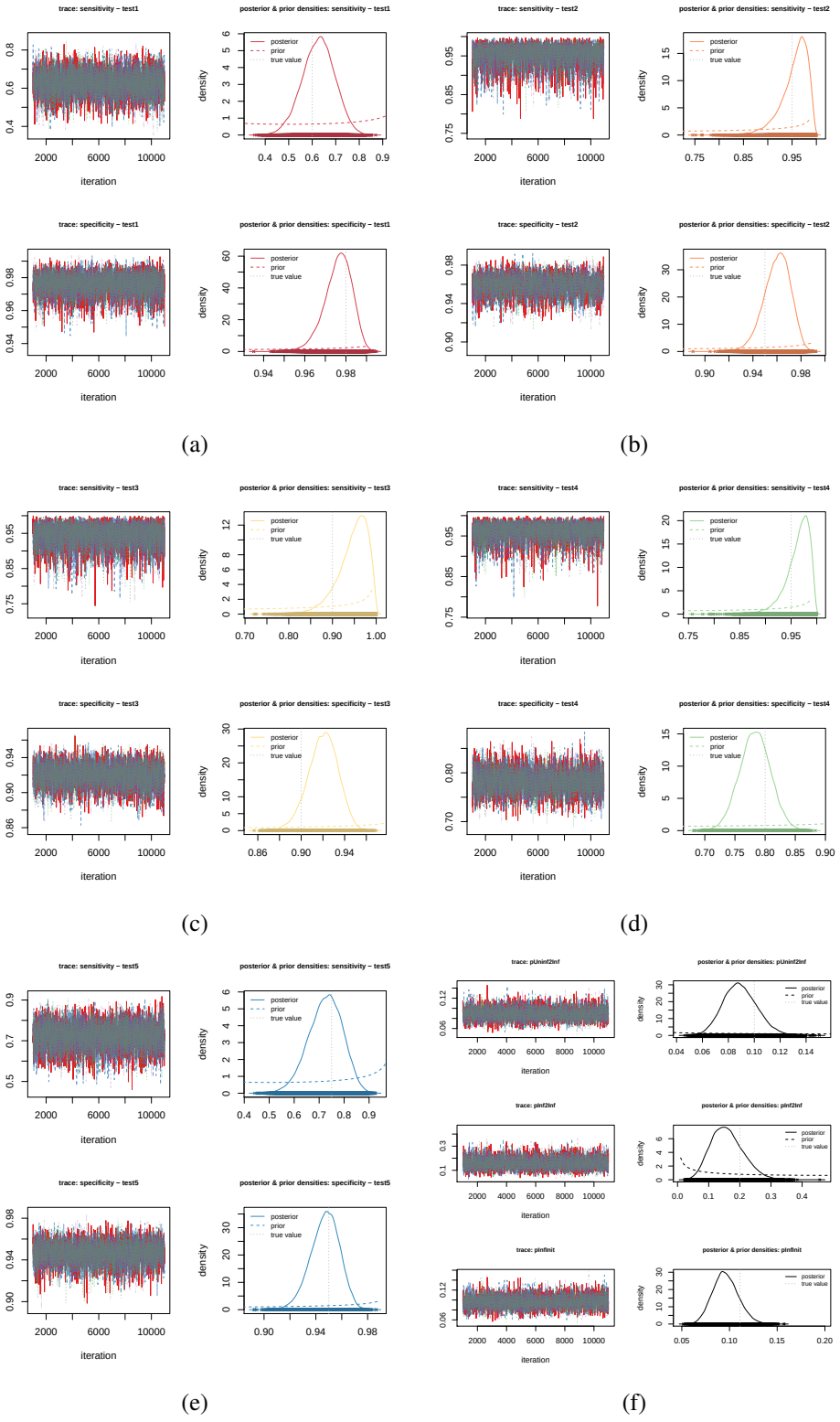


Figure 2: Trace plots and one-dimensional posterior distribution plots for the CRPs (a through e) and transition and initial probability parameters (f) of the mixed LMM from Figure 3b.

While we do not present a full and thorough investigations of this, we will look at each scenario for a specific simulated dataset: Sim1 (CIA holds) and Sim2 (CIA violated). We estimate deviance information criteria (DIC; Spiegelhalter et al., 2002) as well as penalized expected deviances (PED; Plummer, 2008) and compare posterior parameter densities and maximum a posteriori (MAP) estimates from the models to the actual values used to generate the data. DICs are more commonly used but are generally not considered to be suitable for mixture models where PEDs should be preferred.

Table 4: DIC / PED values for the basic and mixed LMMs fitted to the simulated datasets.

	basic LMM	mixed LMM
Sim1	1,575 / 1,732	1,576 / 1,784
Sim2	2,042 / 2,293	1,821 / 2,104

Table 4 shows the DIC and PED values for the basic and mixed LMM fitted to both simulated datasets. When the CIA is not valid, the (correct) mixed LMM gives an appreciably lower DIC (1,821) and PED (2,104) than the (wrong) basic LMM (DIC 2,042, PED 2,293). This is confirmed by Figure 3, which shows actual and estimated sensitivities and specificities for the Sim2 dataset: for most of the 5 tests, the mixed LMM's maximum a posterior (MAP) estimates of test sensitivities and specificities are closer to the actual values. In particular, for the mixed LMM the actual parameter values always lie within the rectangle given by the 95% credible intervals for sensitivity and specificity of each test whereas this is not the case for the basic LMM on this dataset.

When the CIA does hold (Sim1 dataset) however, then both models perform similarly well, with the basic LMM achieving DIC and PED values only marginally lower than those for the mixed LMM (DIC 1,575 compared to 1,576; PED 1,732 compared to 1,784). The analogous figure to Figure 3 for this dataset shows no appreciable difference between the posterior sensitivity and specificity distributions from both models.

These results suggest that the mixed LMM offers real benefits when the CIA does not hold and that there is little harm in using the more complex mixed LMM when the CIA is true and the basic LMM would have been sufficient. For real world datasets where it can be difficult to know whether the CIA does really hold or not, this is an important result.

4.2 Malawi *Salmonella* data

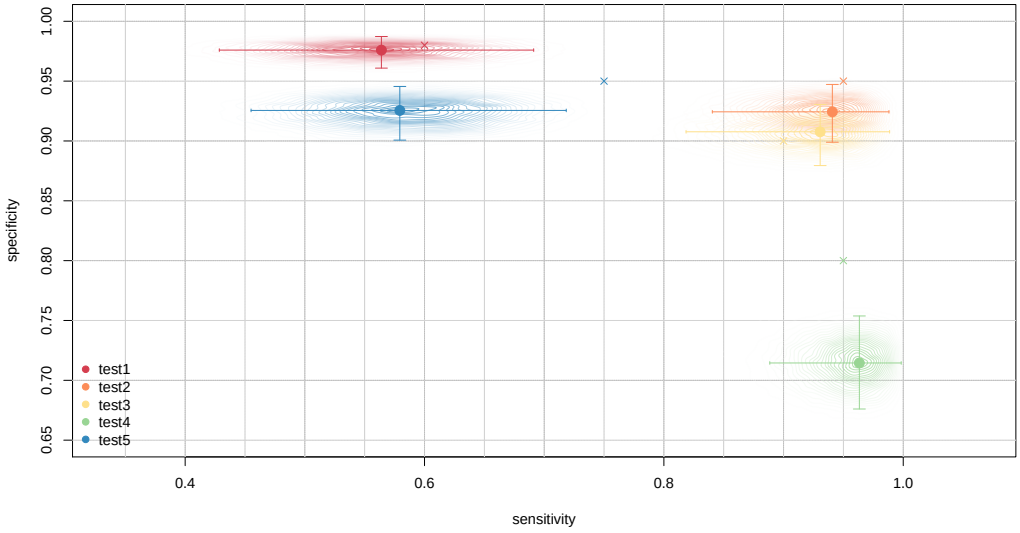
Having reassured ourselves that our implementations of the basic and mixed LMMs converge and give sensible results on simulated data, we can fit the LMMs from Table 1 to the Malawi *Salmonella* data described in section 3.2.

Both the basic time heterogeneous (model 2 from Table 1 and mixed time homogeneous (model 3) LMM show poor convergence. This is likely a result of the sparsity of the data. As described above, no participant has been observed at all $T = 13$ timepoints, and the overall missing data rate is 46%. Further, with infection being rare (78 of 421 observations have at least one positive test result [18.5%] and only 36 of 421 have two or more positive test results [8.6%]), this means the dataset is quite sparse. With models 2 and 3 not converging, we did not attempt to fit model 4, the mixed time heterogeneous LMM.

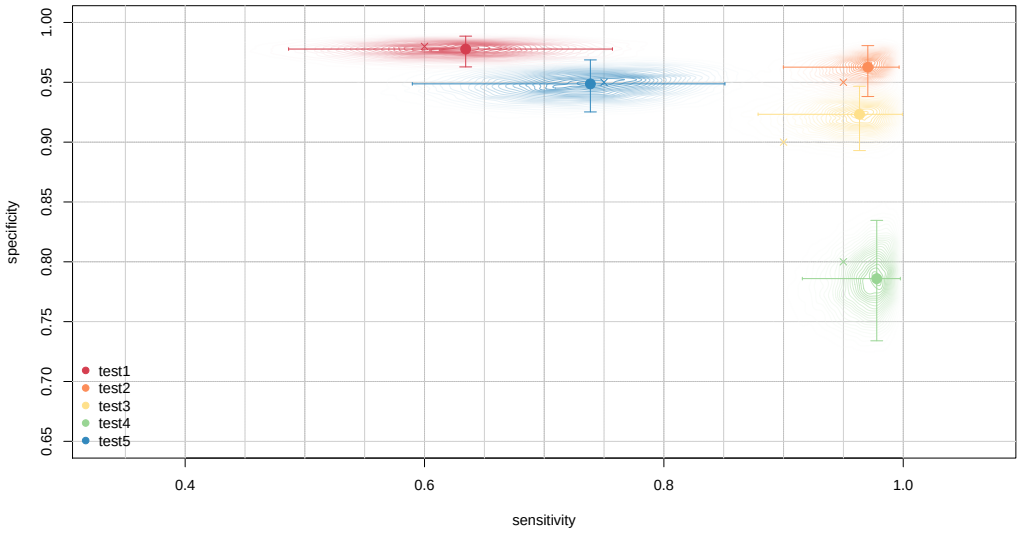
This illustrates a practical point: *Salmonella* infections should exhibit seasonal variations, yet the time heterogeneous model does not yield a better fit and similarly, while the Malawi *Salmonella* stool culture and PCR data motivated us to develop the mixed LMMs,

we are not able to fit these models to the data. The biologically more plausible models involve more parameters and hence unless there is enough data to fit these models well, one may have to stick with the simpler, more parsimonious model.

Figure 4 shows the posterior density estimates from the basic time homogeneous LMM for the sensitivities and specificities of the 5 diagnostic tests. These results agree with broad



(a) Basic LMM



(b) Mixed LMM

Figure 3: Posterior distributions of the conditional response probabilities (shown as sensitivities and specificities) for data simulated with subject-level random effects. The basic LMM cannot capture the conditional dependencies present in the data and the posterior distributions reflect the actual values (indicated by crosses) used to generate the data less well. Contours indicate the posterior distributions, the segments indicate the 95% highest posterior density credible intervals around the MAP estimates indicated by round dots.

domain knowledge that stool culture is near 100% specific but lacks sensitivity, whereas the molecular tests, designed to have much higher sensitivities, have lower specificity.

From the results obtained using the basic time homogeneous LMM, we conclude that the TTR PCR test achieves the overall best sensitivity & specificity trade-off (MAP estimates: sensitivity 99.5%, specificity 95.5%).

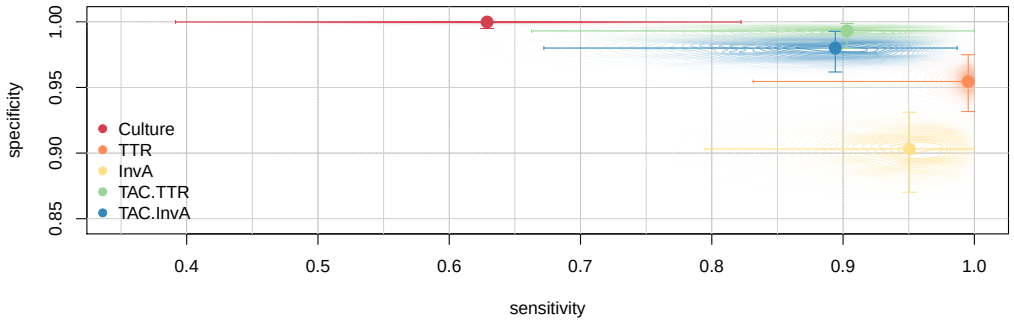


Figure 4: Posterior distributions of the conditional response probabilities (shown as sensitivities and specificities) for the Malawi *Salmonella* data using the basic LMM. Contours indicate the posterior distributions and the segments indicate the 95% highest posterior density credible intervals around the MAP estimates indicated by round dots.

5. Conclusion

Motivated by real Malawi *Salmonella* data we have developed mixed LMMs that relax the CIA and thereby widen the applicability of LMMs. We have presented the mathematical details of this extension, discussed our implementation thereof, evaluated and compared basic and mixed LMMs on both simulated data and the Malawi *Salmonella* dataset. While the simulations show the benefit of the mixed LMMs that we have developed, the Malawi data were too sparse to fit models of higher complexity (mixed LMMs and / or time heterogeneous LMMs) than the basic, time homogeneous LMM. We hope to apply the methodology we have developed here to future, similar — and hopefully less sparse — datasets.

REFERENCES

Altman, R. M. (2007), "Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting," *Journal of the American Statistical Association*, 102, 201–210.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013), *Latent Markov Models for Longitudinal Data*, Boca Raton, FL, USA: CRC Press.

Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017), "LMest : An *R* Package for Latent Markov Models for Longitudinal Categorical Data," *Journal of Statistical Software*, 81.

Chirambo, A. C., Nyirenda, T. S., Henrion, M. Y. R., Msefula, C., Jambo, N., Nyirenda, J., Kamngona, A., Iturizza-Gomara, M., and Gordon, M. A. (2018), "Evaluation of *Salmonella* detection methods in stool samples," *in preparation*.

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C. S., Sheeber, L. B., Allen, N. B., and Hamaker, E. L. (2017), "On the Use of Mixed Markov Models for Intensive Longitudinal Data," *Multivariate Behavioral Research*, 52, 747–767.

Gelman, A. and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Henrion, M. Y. R. (2018a), "Bayesian mixed latent Markov models for binary diagnostic data," http://marchenrion.com/?butSoft=1, accessed: 2018-09-10.

Henrion, M. Y. R. (2018b), "Bayesian mixed latent Markov models for binary diagnostic data," https://github.com/gitMarcH/Bayesian-mixed-latent-Markov-models-for-binary-diagnostic-data, accessed: 2018-09-10.

Jasra, A., Holmes, C. C., and Stephens, D. A. (2005), "Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling," *Statistical Science*, 20, 50–67.

Koukounari, A., Donnelly, C. A., Moustaki, I., Tukahebwa, E. M., Kabatereine, N. B., Wilson, S., Webster, J. P., Deelder, A. M., Vennervald, B. J., and van Dam, G. J. (2013a), "A Latent Markov Modelling Approach to the Evaluation of Circulating Cathodic Antigen Strips for Schistosomiasis Diagnosis Pre- and Post-Praziquantel Treatment in Uganda," *PLoS Computational Biology*, 9, e1003402.

Koukounari, A., Moustaki, I., Grassly, N. C., Blake, I. M., Basez, M.-G., Gambhir, M., Mabey, D. C. W., Bailey, R. L., Burton, M. J., Solomon, A. W., and Donnelly, C. A. (2013b), "Using a Nonparametric Multilevel Latent Markov Model to Evaluate Diagnostics for Trachoma," *American Journal of Epidemiology*, 177, 913–922.

Lazarsfeld, P. F. and Henry, N. W. (1968), *Latent Structure Analysis*, Boston, MA, USA: Houghton Mifflin.

Muthén, L. K. and Muthén, B. O. (1998-2017), *Mplus User's Guide*, Muthn and Muthn, Los Angeles, CA, USA, eight edition ed.

Nyirenda, T. S. (2015), "Natural immunity to *Salmonella* in humans," Ph.D. thesis, University of Liverpool.

Pan-ngum, W., Blacksell, S. D., Lubell, Y., Pukrittayakamee, S., Bailey, M. S., de Silva, H. J., Lalloo, D. G., Day, N. P. J., White, L. J., and Limmathurotsakul, D. (2013), "Estimating the True Accuracy of Diagnostic Tests for Dengue Infection Using Bayesian Latent Class Models," *PLoS ONE*, 8, e50765.

Plummer, M. (2003), "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," .

Plummer, M. (2008), "Penalized loss functions for Bayesian model comparison," *Biostatistics*, 9, 523–539.

Plummer, M. (2016), *rjags: Bayesian Graphical Models using MCMC*, r package version 4-6.

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.

van Smeden, M., Naaktgeboren, C. A., Reitsma, J. B., Moons, K. G. M., and de Groot, J. A. H. (2014), "Latent Class Models in Diagnostic Studies When there is No Reference Standard – A Systematic Review," *American Journal of Epidemiology*, 179, 423–431.