In search of a novel scoring system for Tuberculous Meningitis Diagnosis: a Bayesian Latent Class Analysis Approach

Trinh Dong Huu Khanh1,2,✉, Joseph Donovan1,3, Guy Thwaites1,4, and Ronald Geskus1,4

1 Oxford University Clinical Research Unit, Centre for Tropical Medicine, Ho Chi Minh City, Vietnam  
2 King’s College London, London, United Kingdom  
3 London School of Hygiene and Tropical Medicine, London, United Kingdom  
4 Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, United Kingdom

✉ Correspondence: [Trinh Dong Huu Khanh <[trinhdhk@oucru.org](mailto:trinhdhk@oucru.org)>](mailto:trinhdhk@oucru.org)

# Introduction

*Tuberculous meningitis (TBM)* is the most severe form of tuberculosis. Diagnosing TBM is notoriously challenging, with microbiological confirmation requiring identification of *Mycobacterium tuberculosis* in paucibacillary cerebrospinal fluid (CSF). In addition to widely used confirmatory methods of CSF testing for *M. tuberculosis* such as Ziehl-Neelsen (ZN) staining, GeneXpert MTB/RIF (Xpert), and Mycobacterial culture, additional parameters may increase the likelihood of a diagnosis of TBM. Such characteristics are illustrated in the uniform case definition for TBM (Marais et al. 2010) where, in the absence of positive microbiological tests, an increased certainty of TBM is assigned in the presence of particular clinical, CSF, and imaging findings, or with evidence of non-neurological *M. tuberculosis*.

In the absence of a gold standard for TBM diagnosis, the uniform case definition has been widely adopted, particularly for the evaluation of new index diagnostic tests. However, the uniform case definition was developed as a standardised approach to reporting TBM research, rather than a gold standard for TBM diagnosis. When categorising cases of suspected TBM, the uniform case definition allocates cases to one of four levels; *Definite*, *Probable*, *Possible*, and *Not TBM*, with both *Probable* and *Possible* each encompassing a numerical score range where higher scores increase the probability of a diagnosis of TBM. All *Probable* cases do not therefore have the same probability of having TBM. *Definite TBM* incorporates microbiological tests, which, in the case of mycobacterial culture, may not return a positive result for several weeks. The best combination of these 4 levels that represents TBM, whilst including no cases of *non-TBM*, is not known and probably lies beyond the use of 4 levels. True TBM cases are likely represented by all *Definite* cases, most *Probable* cases, and some *Possible* cases.

Another risk may lay within the fact that scores are based on consensus where a mere change does need more time and efforts, its flexibility and adaptability to different settings is inferior to a statistical model, in which case several models can be made basing on one same dataset. *(in case this is ambiguous: this closely links to one of our objectives where we build a simplified score table that uses only clinical and demographic data, for resources-limited settings and quick screening)*

*Latent Class Analysis (LCA)* is a statistical technique which has been adopted in settings where a gold standard does not exist, in place of which are a set of mutually independent tests. This model assumes the whole population is separable into different unobserved classes; between them, the chance of positive test results are different. By combining information coming from the aforementioned tests – or manifest variables – we can recover prevalence of the hidden classes and the patients’ true allocation. Figure exemplifies a situation where there are 3 manifest variables representing for two latent classes. The technique is commonly used in social science and psychology, but has been gradually adopted in several diagnostic and prognosis models, especially in Tuberculosis (Stout et al. 2018; Adams et al. 2019; Schumacher et al. 2016; Lahuerta-Marin et al. 2018) given the limitation of current tools. Classical LCA imposes strong assumption of local independence, namely in each latent class, the distributions of all manifest variables are mutually independent, which might not hold in most cases hence hinders the applicability of such design. The introduction of random effects and Bayesian approach has brought more flexibility to the model (Qu, Tan, and Kutner 1996; Toft, Jørgensen, and Højsgaard 2005; Menten, Boelaert, and Lesaffre 2008).

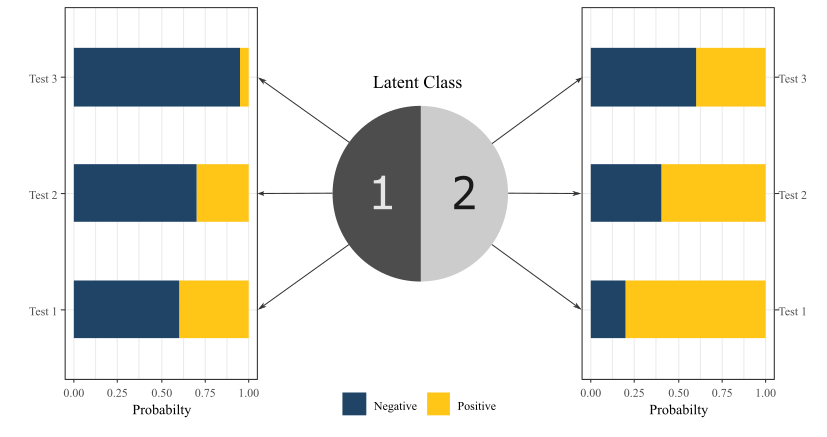


Figure : Design of classic Latent Class Analysis with two unobserved classes 1 and 2 and three iid. binary manifest variables (1, 2, and 3). Distributions of manifest variables in both classes are generated for demonstration only

This analysis’s main objective are to *(1)* re-estimate the latent performances, namely sensitivities and specificities, of current TBM confirmatory methods - ZN Smear, Xpert, and mycobacterial culture - against actual TBM status, taking into account the imperfection in current diagnosis; *(2)* make a re-adjustment to the the diagnosis scoring system, basing on statistical model, providing estimations of individual chance of having TBM.

As secondary objectives, our analysis also aim to build a simplified scoring system which only needs clinical and demographic information but has the capacity to approximate the full system’s output - so that TBM risk can be calculated at admission. We also estimate a latent representation of individual bacillary burden given that they get TBM which may impact the tests results.

# Methods

## Population to analyse

Data used for this analysis were extracted from an observational study of brain infection conducted at the Hospital for Tropical Diseases (HTD), Ho Chi Minh City, Vietnam, a large centre providing secondary and tertiary treatment for a wide range of tropical infections (G. Thwaites et al. 2002). Participants for this analysis were enrolled between 29th August 2017 and 22nd January 2021. Enrolled participants to the prospective observational study were at least 16 years old, with suspected brain infection, admitted to a neuro-infection ward, and undergoing lumbar puncture at baseline as a routine diagnostic procedure. Patients were ineligible for enrolment if performing a lumbar puncture was contraindicated, or if informed consent to join the study was not given (by the patient, or by a relative if the patient lacked capacity to consent). This study received ethical approvals from HTD and Oxford Tropical Research Ethics Committee (Donovan et al. 2020). Specifically for the analysis, we also excluded patients with contaminated MGIT results and with no subsequent culture in the first week since admission.

## Data collection and testing procedures

Demographic data and relevant medical history were recorded. HIV testing was performed on a case-by-case basis by the treating clinician. However, all patients that were to enrol in a subsequent Tuberculous meningitis randomised controlled trial also conducted at HTD were tested for HIV. Patients with either a known HIV positive status, or a positive HIV test result, were considered HIV positive. Patients underwent clinical examination and laboratorial investigations according to main study protocol, including: blood test, sputum, and lumbar puncture, unless contra-indicated. CSF was obtained by lumbar puncture, with CSF white blood cells and cellular differential, CSF protein, CSF glucose (with paired blood glucose), and CSF lactate routinely measured. Optimally 6mls CSF was used for mycobacterial testing unless less CSF was available (in which case mycobacterial testing was still performed). Confirmatory tests for TBM performed on CSF and used for this analysis were Ziehl–Neelsen staining and smear (ZN Smear), mycobacterial culture using Mycobacteria Growth Indicator Tube (MGIT), and Xpert or Xpert MTB/RIF Ultra (XpertUltra). Given Xpert and XpertUltra were considered diagnostically comparable in a recent comparative study from this centre using a subgroup of these data (Donovan et al. 2020), Xpert and XpertUltra were considered the same in this analysis. Confirmatory mycobacterial testing was performed if TBM was suspected by the treating clinician.

Methods of CSF processing have been described elsewhere Donovan et al. (2020), however, briefly; CSF samples were centrifuged at 3000g for 15 minutes, and most of the CSF supernatant was removed. The CSF deposit was resuspended in 500µL of remaining supernatant, with this resuspended pellet then used for ZN smear (100µL), MGIT (200µL), and either Xpert or XpertUltra (200 µL).

All patients received appropriate anti-TB chemotherapy regimens according to national and local guidelines, depending on the diagnosis, without any interference from the study. At the time of discharge or death, all patients received a final diagnosis. If at least one of ZN Smear, Xpert, Xpert Ultra, or MGIT, was positive at any time during the follow-up, the patient would be considered confirmed TBM; otherwise, if TBM was clinically suspected and treated, with confirmatory microbiological tests negative, the patient would be considered suspected TBM. If the patient recovered without anti-TB chemotherapy they would be reassigned to another diagnosis (i.e. not TBM).

In the instance of repeat CSF sampling (performed based on clinical need), only the first samples with at least 3mls of collect CSF collected not later than the first week since admission were used. Results of haematological and biochemical parameters, and of ZN Smear, Xpert, Xpert Ultra, and MGIT, were taken from one same CSF samples (in addition to the paired blood glucose).

## Statistical Model

In the model, three aforementioned TBM confirmation tests (ZN Smear, MGIT, and Xpert) were used as manifest variables. Linear predictors for the latent class prevalence were *history and demographic information* (HIV status, Age, Past noticeable TB contact), *clinical signs and symptoms* (Days from onset to admission, Systemic symptoms suggestive of tuberculosis, Focal neurological deficit, Cranial nerve palsy, Glasgow Coma Score - GCS), *Imaging chest data* (Pulmonary TB, Miliary TB), *CSF criteria* (lymphocyte count, neutrophil count, oeosinophil count, erythrocyte count, glucose - with corresponding blood glucose, protein, lactate), and *Cryptococcus Antigen/Indian Ink* in combination.

### Data pre-processing

Most continuous variables were transformed to logarithmic scale and subsequently centred and scaled by their standard deviations. Glasgow Coma Score (GCS) and its components (Voice - GCSV, Eyes - GCSE, and Muscle - GCSM) were translated to *Loss of GCS* (LoGCS = LoGCSV + LoGCSE + LoGCSM) so that a would be equivalent to , while would be translated to (table ).

Table : Conversion table from clasic Glasgow Coma Score (GCS) to Loss of GCS (LoGCS)

| Feature | Response | GCS | LoGCS |
| --- | --- | --- | --- |
| Eye response | Open sponatenously | 4 | 0 |
| Open to voice command | 3 | 1 |
| Open to pain | 2 | 2 |
| No eye open | 1 | 3 |
| Verbal response | Orientated | 5 | 0 |
| Confused | 4 | 1 |
| Inappopriate words | 3 | 2 |
| Incomprehensible sounds | 2 | 3 |
| No verbal response | 1 | 4 |
| Motor response | Obey command | 6 | 0 |
| Localising pain | 5 | 1 |
| Withdrawal from pain | 4 | 2 |
| Flexing | 3 | 3 |
| Extending | 2 | 4 |
| No motor response | 1 | 5 |

Compared to the uniform case definition, we also made a change to **Past TB contact**, in which we hypothesise that by editing the original question into “Past noticeable contact with TB patients within the past recent year,” we would be able to imply all “Unknown” answer as a “No.” In the main study, no brain imaging was taken, hence not included.

Binary variables were dummy coded into 0 for “Negative”/“No” and 1 for “Positive”/“Yes.”

### Latent class regression model

We created two-level model consisting of:

1. *Prevalence model*: A logistic regression model estimating the prevalence of TBM amongst the study population
2. *Latent class analysis*: estimating the probabilities of having positive results from three tests in each class. Similar to previous applications (Qu, Tan, and Kutner 1996; Hadgu and Qu 2002; Schumacher et al. 2016), we also corrected for individual bacillary burden and noise coming from the procedures themselves. *Latent bacillary burden* is regressed by a Gaussian random variable and fixed effects coming from relevant bio-markers. We hypothesised that amongst TBM-positive patients, those who have lower bacillary burden would be less likely to be tested positive.

Table : Contribution of different variables in each model

| **Predictor** | **Risk of TBM** | **Bacillary Burden** |
| --- | --- | --- |
| Age | + | 0 |
| HIV Status | ++ | + |
| Past TB contact | + | 0 |
| TB-suggested symptoms | + | 0 |
| Local motor deficit | + | 0 |
| Cranial nerve palsy | + | 0 |
| Days from onset | + | 0 |
| PTB/X-Ray | + | 0 |
| MTB/X-Ray | ++ | 0 |
| GCS | + | 0 |
| Cryptococcus Antigen/Indian Ink | - | 0 |
| Blood Glucose | - | 0 |
| CSF Glucose | - | ? |
| CSF Lymphocyte Count | + | ? |
| CSF Neutrophil Count | +/- | ? |
| CSF Protein | + | ? |
| CSF Lactate | + | ? |
| CSF Oeosinophil Count | - | 0 |
| CSF RBC Count | - | 0 |
| *"+", "+/-", and "-" reflects prior believes, but no difference was imposed in the model* | | |
| *"++" predictors were strongly believed to be positive risk factors* | | |
| *"0" predictors were not included in any model* | | |
| *"?" predictors were included in some models* | | |

The inclusion of covariates was done according to prior knowledge of potential risk factors (Marais et al. 2010) that share association with the infection and test results, summarised in table ). CSF oeosinophil count was additionally included as it is a strong bio-marker for oeosinophilic meningitis, a condition usually caused by parasitic helminths. CSF erythrocyte count was added as an independent marker for traumatic lumbar puncture, in which case we also made correction to white cell counts and biochemical features, as suggest in (Greenberg et al. 2008; Nigrovic, Shah, and Neuman 2011; Mehl 1986).

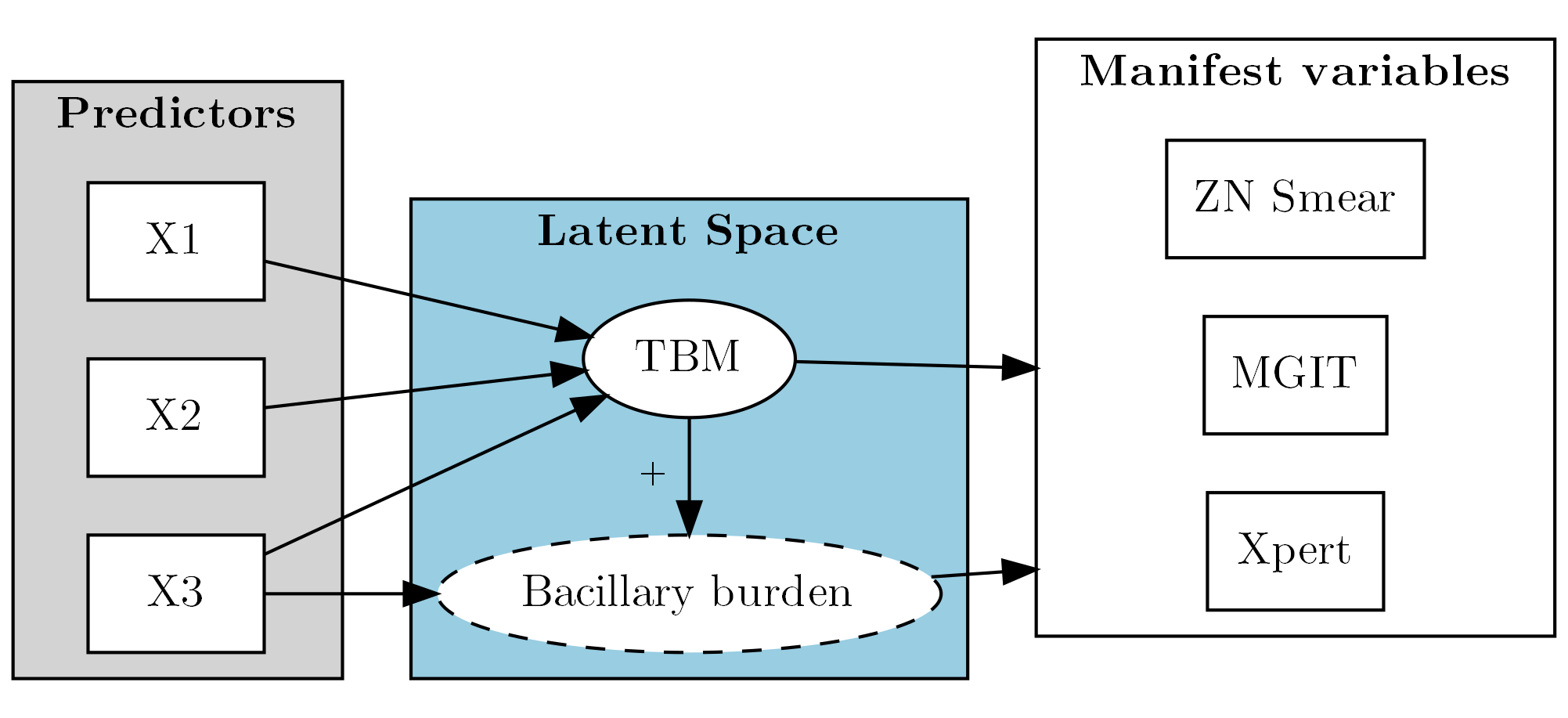


Figure : Model skeleton. Bacillary Burden is only available in model 2+

Basic designs for all models is shown in figure . We use a stepwise approach where we incrementally added up more flexibility and lifted more constraints. We also include several extensions to explore many possibilities than can improve performance. In the main analysis, only CSF neutrophil count was model via a quadratic effect. A summary of all architectures an extensions are shown in table , whereas technical formulations are detailed in appendix .

Table : Model architectures and extensions

| **Model** | **Base definition (compared to the lower one)** |
| --- | --- |
| 1 | No bacillary burden; everyone in the same class has equal risk of tested positive |
| 2 | Added individual bacillary burden; impacts of bacillary burden on test results are the same |
| 3 | Impacts of bacillary burden are different for different tests |
| 4 | Added technical fluctuation as a second random effect; fixed effects only contributes to bacillary burden |
| 5 | Added fixed effects for technical fluctuation |

### Missing data handling

#### Manifest variables

By design, most patients with very high chance of and/or evidently diagnosed with different diseases were not tested with TBM confirmation assays (namely ZN Smear, MGIT, and Xpert), unless there were excessive amount of CSF samples. However, as confirmation tests’ were all firmly believed to be almost specific Heemskerk et al. (2018), we assumed that patients who had no TBM confirmation tests are all negative. Apart from one premature death, all patients were tested or diagnosed with a different disease.

#### Predictors

In the analysis, most missingness were assumed to be Missing At Random (MAR) and imputed in one step within the main model, unless explicitly stated (table ). Composite predictors such as *TB-suggested symptoms* and *Glasgow Coma Score (GCS)* were imputed by corresponding compartments; and potentially correlated variables were grouped and imputed together (as in *CSF biomakers*’ case). Due to Stan not supporting Multivariate Logistic Distribution, in favour of methodological consistency, all binary predictors were imputed using (Multivariate) Probit models, which facilitate a (Multivariate) Normal Distribution instead (Albert and Chib 1993). Continuous predictors are imputed using Multivariate Linear Regression.

As HIV tests were not mandatory in the main study, HIV status itself was also incomplete, but their chance of missingness were mostly dependent on whether or not the patients were to enrol in a TBM study. Hence, it is safely to assume that HIV status are also MAR. Accordingly, we imputed HIV using probit regression, with Blood Lymphocyte and Blood Neutrophil counts as donors.

In the case where missing values were imputed, figure depicts how those were sampled, together with their potential donors. Note that due to technical limitation, imputed LoGCS were treated as-is in the sampling step, but were rounded when estimating model performance.

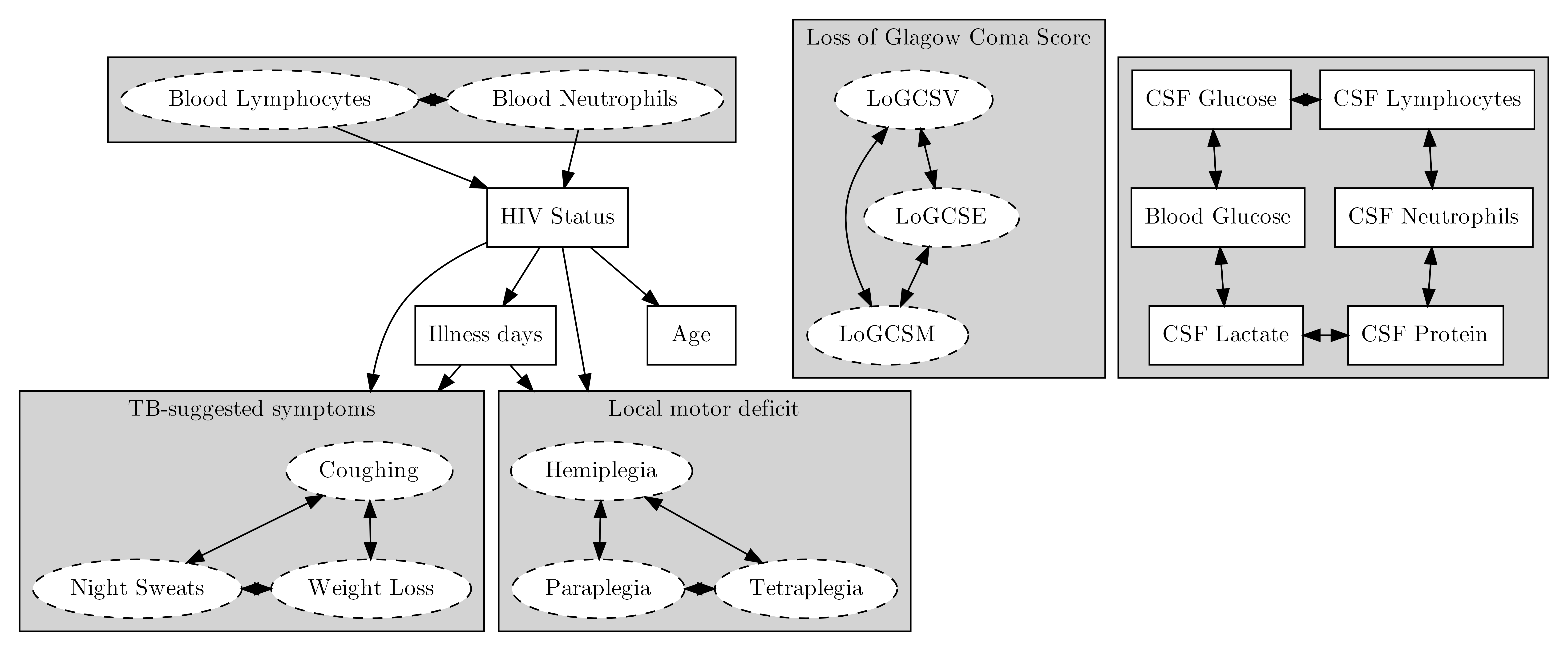


Figure : Imputation strategy for predictors. Variables in rectangular solid boxes were used in the model, in oval dashed were only contributed in the imputation model as donors and were not directly included in the main model. Clustered covariables were imputed together in a multivariate regression. Arrows demonstrate a donor response correlation.

### Prior choices

For imputation, was chosen as prior distributions for all intercepts and coefficients (Stan Development Team n.d.), except for LoGCS (E, V, and M) as their supports were two-sided constained (section ) - in which case, suitable Uniform distributions were chosen for the means, and Normal for standard deviations (appendix ).

Cholesky decomposition of the covariance matrix was sampled from a *LKJ Correlation Cholesky* prior with (Stan Development Team 2021b).

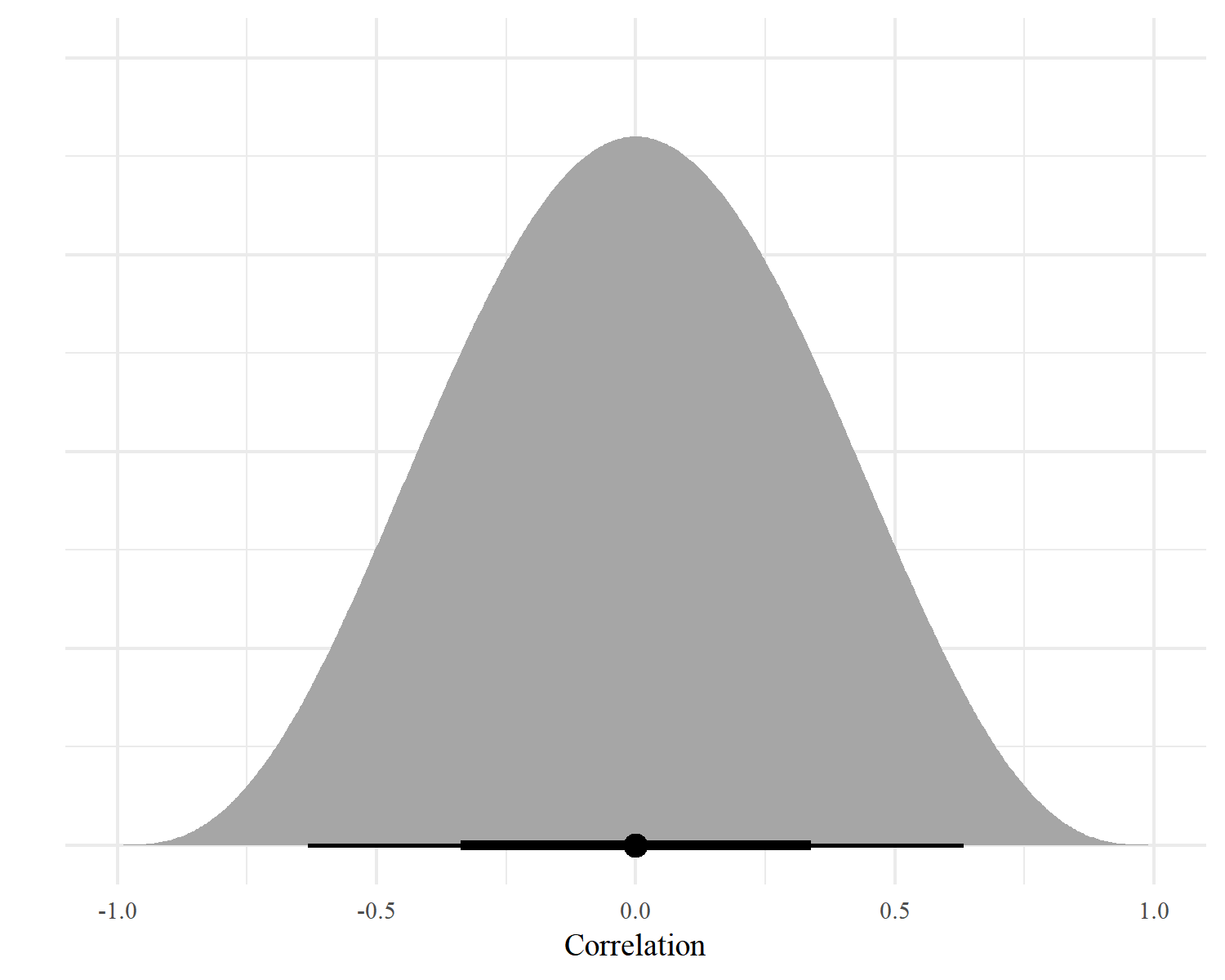


Figure : Example density plot for in bivariate case; 1 is perfect positive correlation, 0 is no correlation, and -1 is perfect negative correlation; black dot, thick, and thin line are the median, IQR and 95% credible interval

In LCA part, we imposed Student T with 4 degree of freedom () for intercept, in which we fixed scale to 5 as with belief that its norm would not be higher than 10 (Boonstra, Barbaro, and Sen 2019). For coefficients, we use Normal distribution and multiplied sd by data standard variation for continuous variables, and 2 for binary variables, so that relatively similar penalisation was formed between them Andrew Gelman (2007).

s.t: and = Standard Deviation of corresponding covariates after the imputation

For strongly positive risk factors (marked *++* in table ), priors were further left-truncated at 0. Individual random effects was sampled from distribution.

For manifest variables, our choices of prior are based on information collected from several previous studies involving Vietnamese cohort (Nhu et al. (2013), G. E. Thwaites, Chau, and Farrar (2004), Heemskerk et al. (2018)). A summarisation of these choices are shown in figure ), on Logistic and Linear scale, and their relations to corresponding results derived from different studies. As suggested by the prior knowledge, we used highly informative priors for specificity (Spc) and weakly informative priors for sensitivity (Sen) on the logit scale (formula (1)), given the discrepancies between different research.

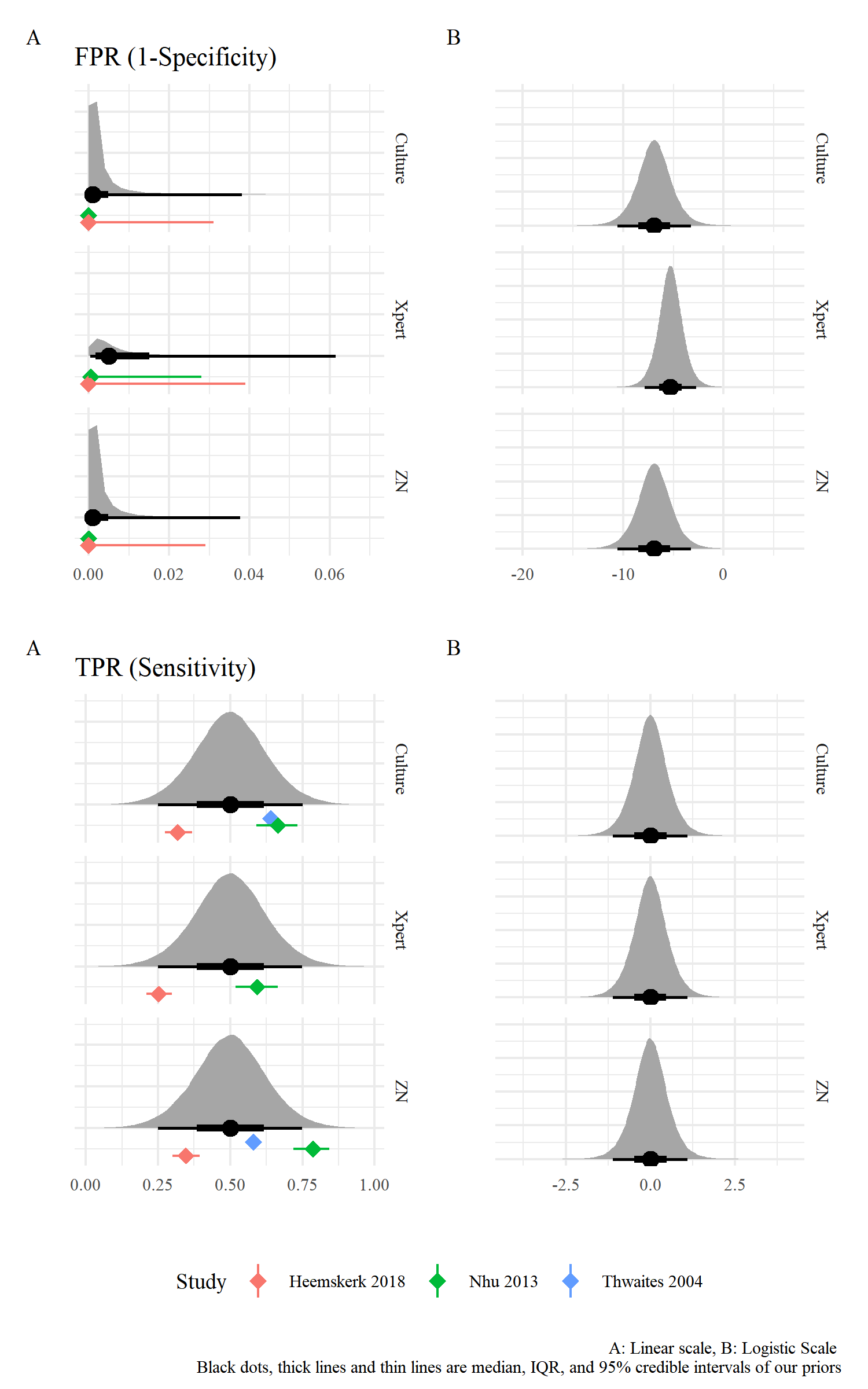


Figure : Density plots for prior distributions and their adherence to prior knowledge of sensitvity and specifity for TBM confirmation tests, against then-made clinical diagnosis. Note that Thwaites 2004 was descriptive only while ZN and Culture in Nhu 2013 were references hence no Confidence Interval

### Model performances

Models performances were estimated and compared by three metrics:

* Expected log point-wise predictive density (*elpd*) of held-out observations: Models with higher *elpd* are supposed to have more predictive values for untrained observations (Vehtari, Gelman, and Gabry 2016).
* Visualised model calibration between estimated predicted and observed probabilities of positive confirmation tests, using a non-parametric loess fit (Harrell Jr 2021). A model with good calibration would predict the risk equivalent to observation.
* Receiver Operating Characteristic (ROC) curve and corresponding Areas Under the Curve (AUC) . Confidence interval for AUC was estimated by a 2000-time bootstrapped samples. A model with good discriminative value would be better to distinguish between two class.

For a good insight, we also visualised class-wise predicted probability density plots to visualise how much separable the classes are based on the models. This demonstrates how separated predicted values are between two classes. We also used the diagnosis at discharge as a pseudo-gold standard to visualise the correlation of model-based and clinical diagnosis.

All metrics are combined from 5 repetitions of 20-fold cross validated datasets (i.e. 100 fits, as suggested by (Harrell Jr 2021)). Accordingly, we selected three best-performed models and re-estimated their parameters using the full dataset.

We also test for local independence assumption. .

### Exploratory and sensitivity analysis

#### Complete-case analysis and imputation under MNAR assumption

To check the level of impact from our imputation method, we did a complete-case analysis. Suspected MNAR variables as listed in table were also tested for MNAR where we randomly allocated values based on experts’ opinions. Under MNAR assumption, we did a pattern-mixture method (Mason et al. 2017; White et al. 2007), where we inquired prediction offsets . represents the difference between unobserved part and observed part of each variables, after correction for all hyper-predictors listed in . Offsets were collected from interviews with experts working at the Viet Anh Ward at HTD, where they were supposed to provide an estimation and 95% confidence intervals (95% CI), based on which s were then sampled from.

*where* *and* *are estimation and* *of 95% CI provided by interviewed experts*.

The main model underwent a relief of strict priors for test specificities in formula (1), the lifted model used the same prior for all three tests which cover specificity from at least 90%:

$$ 1-Spc\_{lifted} Logistic(, 1)

$$ *where* for ZN Smear and MGIT, for Xpert.

Lastly, as recent studies suggested a sup-optimal specificity of Xpert test on CSF samples(Nhu et al. 2013; Chen et al. 2020), our assumptions made in table might not completely valid. To tackle this, we considered a MAR scenario, where observation chance of confirmation tests depend on the unknown TBM status and locally independent to the value of confirmation tests. The observation status was then included in the model as a separated manifest variables. The validity of this method was depicted in a simulation study in .

In this analysis, we expected the chance in which at least one confirmation test was done are 95% for TBM-positive patients, and 50% for TBM-negative, hence led to two conservative priors:

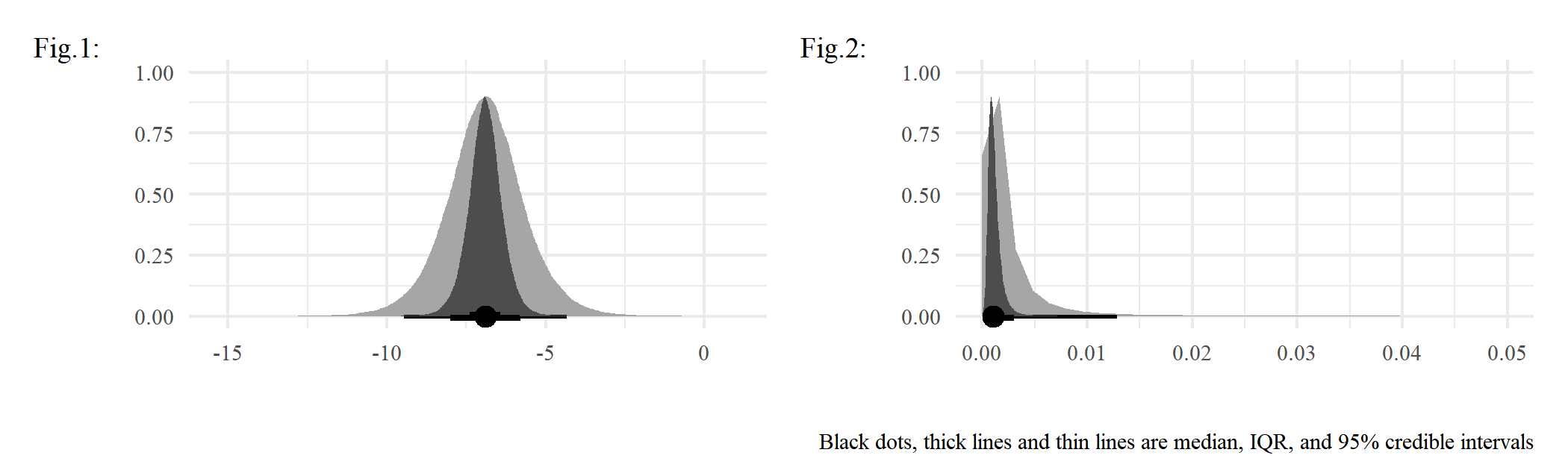


Figure : Lifted prior for ZN Smear and MGIT specificities, compared to the old one (dark grey). 1: logit scale, 2: linear scale

#### Prior families

Considering our prior family might have affected the estimation, we also consider different families of priors for coefficients in the best model. Scales were automatically selected and adjusted using the same technique mentioned in ..

* Weakly informative prior:
* LASSO-equivalent prior:

s.t: and = Standard Deviation of corresponding covariates after the imputation

#### Non-linearity and additional features

As part of the exploratory analysis, we considered two models:

* Non-linearity for all CSF bio-markers: We added quadratic effects for all CSF bio-markers in the prevalence model. LASSO-based variable selection was implemented for this analysis.
* Non-linearity for Glasgow coma scores: As GCS is not a continuous variable but rather an ordinal one de facto, it is possible that there is non-linear correlation between GCS and TBM risk. In this analysis, we employed a quadratic effect for GCS to capture this potential of non-linearity.
* Impact of CSF volume: By including the volume of sample collected, we can further investigate the effect size of CSF volumes on the sensitivity of each confirmation test.

### Simplified approximation of TBM risk

As the full model requires a plethora of predictors and measurements, it might not be pragmatic in some limited contexts. We hence developed a simplified version of the prevalence sub-model which exclude all laboratorial features. The aim of this exploratory analysis is to provide a decent approximation of TBM risk yet needs only a minimal amount of information. We took out the analytic forms of posterior distributions from the best-performed model, on the logit scale, fed them into subsequent model where the number of predictors were reduced. The error term follows Logistic distribution with .

*s.t.* is the full model’s output and is an approximation of .

All data preparation, cleaning, and processing were performed on statistical package R, version 4.1.0 (R Core Team 2021). The model was developed on the probabilistic language Stan via the interface Rstan , version 2.27 (Stan Development Team 2021a). Plotting was done using package bayesplot (Gabry et al. 2019), classifierplots (Defazio and Campbell 2020), and ggvenn (Yan 2021). Some other packages used include: R6 (Chang 2020), rms (Harrell Jr 2021).

# Results

Venn diagram for confirmation test results are demonstrated in

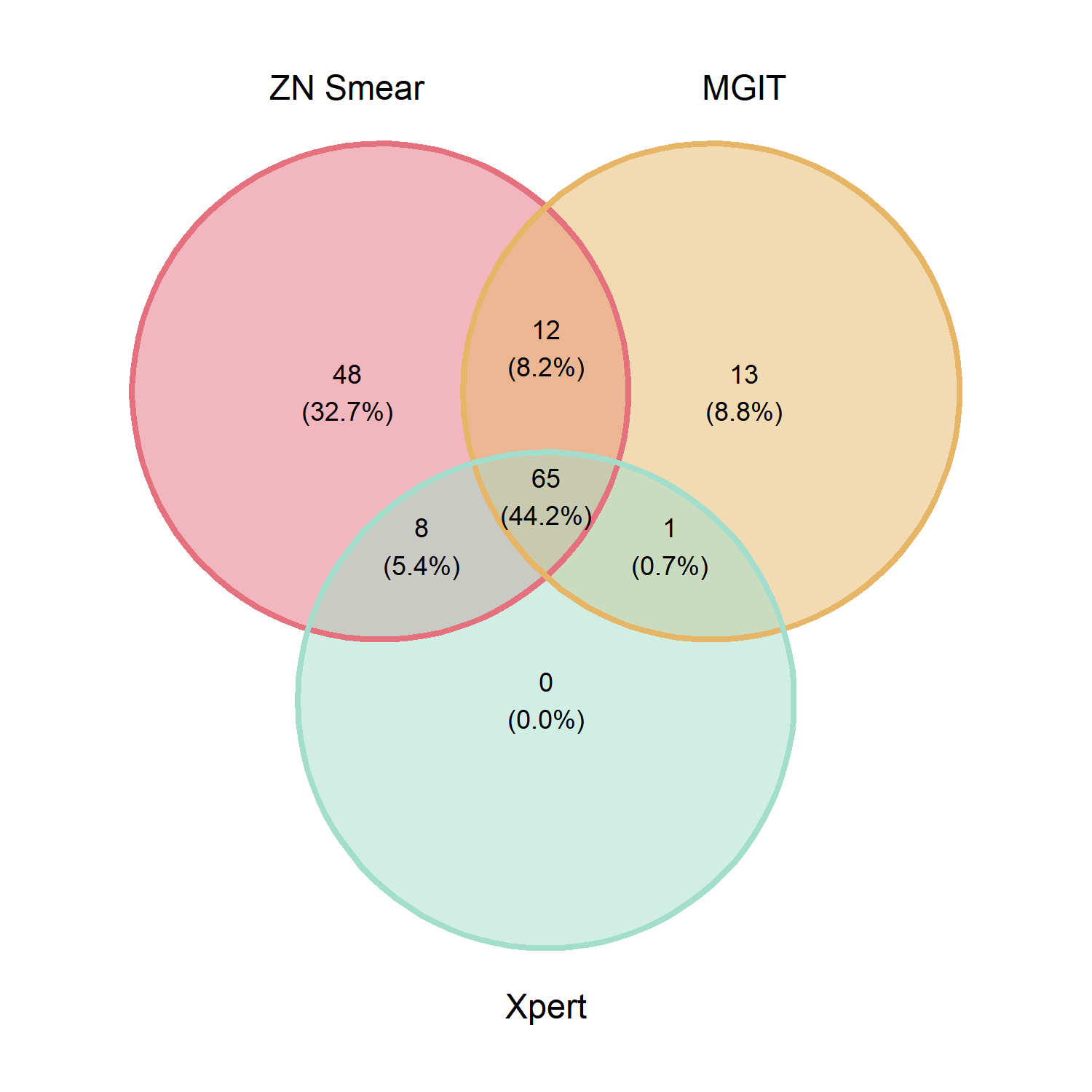


Figure : Venn diagram for ZN Smear, MGIT, and Xpert

Table : Rationales and measures to handle missing values

| **Variable** | **N  missing** | **Expected Reason of Missingness** | **Mechanism** | **Handling method** |
| --- | --- | --- | --- | --- |
| ZN Smear | 311 | Not suspected TBM | MNAR | Set = 0 |
| MGIT | 314 | Not suspected TBM | MNAR | Set = 0 |
| Xpert | 305 | Not suspected TBM | MNAR | Set = 0 |
| HIV Status | 215 | Test not mandatory | MAR/MCAR | Imputation |
| TB-suggested symptoms | 273 | Unmeasured / Unnoticed / Unconscious | MAR/MNAR | Imputation |
| Local neuro-deficit | 17 | Unconscious | MAR/MNAR | Imputation |
| Glasgow Coma Score | 21 | Unconscious (GCSV) / Blinded (GCSE) / Unmeasured | MAR | Imputation |
| Age | 1 | Input error | MCAR | Imputation |
| Illness days | 29 | Patients forget / Unconscious | MAR | Imputation |
| Blood Lymphocyte | 1 | Unmeasured (premature death) | MAR | Imputation |
| Blood Neutrophil | 1 | Unmeasured (premature death) | MAR | Imputation |
| Blood Glucose | 7 | Most likely input error / Unmeasured (premature death) | MAR/MCAR | Imputation |
| CSF glucose | 1 | Unmeasured (premature death) | MAR/MCAR | Imputation |
| CSF lymphocyte count1 | 2 | Very low or zero / Input error / Unmeasured (premature death) | MNAR/MAR | Manually set/Imputation |
| CSF neutrophil count1 | 21 | Very low or zero / Input error / Unmeasured (premature death) | MNAR/MAR | Manually set/Imputation |
| CSF protein | 2 | Data input error / Unmeasured (premature death) | MAR/MCAR | Imputation |
| CSF lactate | 1 | Unmeasured (premature death) | MAR/MCAR | Imputation |
| *1CSF Cell counts = CSF white-cell count x Pct of Cell Type / 100; if very low, then either lymphocytes or neutrophils had values, the other were left missing* | | | | |

Table : History and baseline information

|  | **MGIT** | | **Xpert (Ultra) MTB/RIF** | | **ZN Smear** | |
| --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **Negative, N = 5961** | **Positive, N = 911** | **Negative, N = 6131** | **Positive, N = 741** | **Negative, N = 5541** | **Positive, N = 1331** |
| Age | 42 (27, 56) | 41 (30, 50) | 42 (27, 56) | 40 (29, 48) | 42 (26, 56) | 42 (29, 53) |
| HIV Positive | 26 (6.7%) | 25 (29%) | 27 (6.7%) | 24 (34%) | 18 (5.1%) | 33 (27%) |
| - Missing | 208 | 6 | 211 | 3 | 202 | 12 |
| Diabetes | 37 (6.2%) | 2 (2.2%) | 39 (6.4%) | 0 (0%) | 35 (6.3%) | 4 (3.0%) |
| Day from onset | 10 (4, 11) | 19 (10, 20) | 10 (4, 11) | 20 (11, 22) | 10 (4, 11) | 16 (8, 17) |
| - Missing | 21 | 8 | 22 | 7 | 19 | 10 |
| TB-suggested symptoms | 82 (23%) | 39 (67%) | 86 (24%) | 35 (71%) | 66 (20%) | 55 (64%) |
| - Missing | 240 | 33 | 248 | 25 | 226 | 47 |
| Past noticeable contact TB within 12 months | 7 (1.2%) | 3 (3.3%) | 6 (1.0%) | 4 (5.4%) | 4 (0.7%) | 6 (4.5%) |
| Focal neurological deficit | 64 (11%) | 20 (22%) | 67 (11%) | 17 (23%) | 59 (11%) | 25 (19%) |
| - Missing | 15 | 2 | 16 | 1 | 13 | 4 |
| Crain nerve palsy | 61 (10%) | 26 (29%) | 64 (10%) | 23 (31%) | 58 (10%) | 29 (22%) |
| Glasgow Coma Score | 12 (10, 14) | 11 (10, 14) | 12 (10, 14) | 11 (9, 14) | 12 (10, 14) | 12 (10, 15) |
| - Missing | 19 | 1 | 19 | 1 | 19 | 1 |
| X-Ray Pulmonary TB | 18 (3.0%) | 19 (21%) | 18 (2.9%) | 19 (26%) | 13 (2.3%) | 24 (18%) |
| X-Ray Miliary TB | 2 (0.3%) | 3 (3.3%) | 2 (0.3%) | 3 (4.1%) | 2 (0.4%) | 3 (2.3%) |
| CSF Lymphocyte Count | 283 (9, 274) | 225 (77, 305) | 284 (10, 289) | 201 (54, 236) | 291 (8, 286) | 210 (65, 271) |
| - Missing | 2 | 0 | 2 | 0 | 2 | 0 |
| CSF Neutrophil Count | 1,108 (3, 201) | 241 (53, 223) | 1,080 (3, 171) | 270 (54, 301) | 1,170 (2, 173) | 256 (31, 231) |
| - Missing | 2 | 0 | 2 | 0 | 2 | 0 |
| CSF Oeosinophil Count | 8 (0, 0) | 0 (0, 0) | 8 (0, 0) | 0 (0, 0) | 8 (0, 0) | 0 (0, 0) |
| - Missing | 1 | 0 | 1 | 0 | 1 | 0 |
| CSF Red blood cell Count | 1,649 (0, 700) | 1,439 (0, 132) | 1,721 (0, 700) | 787 (0, 87) | 1,861 (0, 1,000) | 620 (2, 80) |
| CSF Protein | 1.74 (0.42, 1.97) | 2.39 (1.34, 2.43) | 1.69 (0.43, 1.99) | 3.00 (1.52, 2.35) | 1.65 (0.41, 1.90) | 2.57 (1.21, 2.43) |
| - Missing | 1 | 1 | 1 | 1 | 1 | 1 |
| CSF Lactate | 5.0 (2.3, 5.4) | 6.5 (4.5, 8.1) | 5.0 (2.3, 5.4) | 6.9 (5.2, 8.3) | 5.0 (2.2, 5.4) | 5.9 (3.9, 7.8) |
| - Missing | 1 | 0 | 1 | 0 | 1 | 0 |
| CSF Glucose | 3.43 (2.51, 4.30) | 2.04 (1.23, 2.46) | 3.42 (2.49, 4.30) | 1.77 (1.12, 2.16) | 3.51 (2.60, 4.42) | 2.13 (1.27, 2.81) |
| - Missing | 1 | 0 | 1 | 0 | 1 | 0 |
| Corresponding Blood Glucose | 7.01 (5.46, 7.72) | 7.01 (5.74, 7.47) | 7.03 (5.47, 7.80) | 6.87 (5.70, 7.10) | 7.06 (5.44, 7.91) | 6.81 (5.72, 7.26) |
| - Missing | 6 | 1 | 6 | 1 | 6 | 1 |
| Cryptococcal Antigen/Indian Ink | 19 (3.2%) | 1 (1.1%) | 19 (3.1%) | 1 (1.4%) | 19 (3.4%) | 1 (0.8%) |
| **1Mean (1st, 3rd quartiles) for numeric variables; n (%) for for categorical variables** | | | | | | |

# Discussion

# Conclusion

# References

Adams, Shahieda, Rodney Ehrlich, Roslynn Baatjies, Nandini Dendukuri, Zhuoyu Wang, and Keertan Dheda. 2019. “Evaluating Latent Tuberculosis Infection Test Performance Using Latent Class Analysis in a TB and HIV Endemic Setting.” *International Journal of Environmental Research and Public Health* 16 (16): 2912. <https://doi.org/10.3390/ijerph16162912>.

Albert, James H., and Siddhartha Chib. 1993. “Bayesian Analysis of Binary and Polychotomous Response Data.” *Journal of the American Statistical Association* 88 (422): 669–79. <https://doi.org/10.1080/01621459.1993.10476321>.

Andrew Gelman, Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st ed. Cambridge University Press.

Boonstra, Philip S., Ryan P. Barbaro, and Ananda Sen. 2019. “Default Priors for the Intercept Parameter in Logistic Regressions.” *Computational Statistics & Data Analysis* 133 (May): 245–56. <https://doi.org/10.1016/j.csda.2018.10.014>.

Chang, Winston. 2020. *R6: Encapsulated Classes with Reference Semantics*. <https://CRAN.R-project.org/package=R6>.

Chen, Yuan-Zhi, Li-Chang Sun, Yao-Hong Wen, Zhong-Wei Li, Shu-Jin Fan, Hong-Kun Tan, Min Qiu, et al. 2020. “Pooled Analysis of the Xpert MTB/RIF Assay for Diagnosing Tuberculous Meningitis.” *Bioscience Reports* 40 (1). <https://doi.org/10.1042/bsr20191312>.

Defazio, Aaron, and Huw Campbell. 2020. *Classifierplots: Generates a Visualization of Classifier Performance as a Grid of Diagnostic Plots*. <https://CRAN.R-project.org/package=classifierplots>.

Donovan, Joseph, Do Dang Anh Thu, Nguyen Hoan Phu, Vu Thi Mong Dung, Tran Phu Quang, Ho Dang Trung Nghia, Pham Kieu Nguyet Oanh, et al. 2020. “Xpert MTB/RIF Ultra Versus Xpert MTB/RIF for the Diagnosis of Tuberculous Meningitis: A Prospective, Randomised, Diagnostic Accuracy Study.” *The Lancet Infectious Diseases* 20 (3): 299–307. <https://doi.org/10.1016/s1473-3099(19)30649-8>.

Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. “Visualization in Bayesian Workflow.” *J. R. Stat. Soc. A* 182: 389–402. <https://doi.org/10.1111/rssa.12378>.

Greenberg, Rachel G., P Brian Smith, C Michael Cotten, M Anthony Moody, Reese H. Clark, and Daniel K. Benjamin. 2008. “Traumatic Lumbar Punctures in Neonates.” *Pediatric Infectious Disease Journal* 27 (12): 1047–51. <https://doi.org/10.1097/inf.0b013e31817e519b>.

Hadgu, A., and Y. Qu. 2002. “A Biomedical Application of Latent Class Models with Random Effects.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47 (4): 603–16. <https://doi.org/10.1111/1467-9876.00131>.

Harrell Jr, Frank E. 2021. *Rms: Regression Modeling Strategies*. <https://CRAN.R-project.org/package=rms>.

Heemskerk, A. Dorothee, Joseph Donovan, Do Dang Anh Thu, Suzaan Marais, Lidya Chaidir, Vu Thi Mong Dung, Chad M. Centner, et al. 2018. “Improving the Microbiological Diagnosis of Tuberculous Meningitis: A Prospective, International, Multicentre Comparison of Conventional and Modified ZiehlNeelsen Stain, GeneXpert, and Culture of Cerebrospinal Fluid.” *Journal of Infection* 77 (6): 509–15. <https://doi.org/10.1016/j.jinf.2018.09.003>.

Lahuerta-Marin, A., M.G. Milne, J. McNair, R.A. Skuce, S.H. McBride, F.D. Menzies, S.J.W. McDowell, A.W. Byrne, I.G. Handel, and B.M. de C. Bronsvoort. 2018. “Bayesian Latent Class Estimation of Sensitivity and Specificity Parameters of Diagnostic Tests for Bovine Tuberculosis in Chronically Infected Herds in Northern Ireland.” *The Veterinary Journal* 238 (August): 15–21. <https://doi.org/10.1016/j.tvjl.2018.04.019>.

Marais, Suzaan, Guy Thwaites, Johan F Schoeman, M Estée Török, Usha K Misra, Kameshwar Prasad, Peter R Donald, Robert J Wilkinson, and Ben J Marais. 2010. “Tuberculous Meningitis: A Uniform Case Definition for Use in Clinical Research.” *The Lancet Infectious Diseases* 10 (11): 803–12. <https://doi.org/10.1016/s1473-3099(10)70138-9>.

Mason, Alexina J, Manuel Gomes, Richard Grieve, Pinar Ulug, Janet T Powell, and James Carpenter. 2017. “Development of a Practical Approach to Expert Elicitation for Randomised Controlled Trials with Missing Health Outcomes: Application to the IMPROVE Trial.” *Clinical Trials* 14 (4): 357–67. <https://doi.org/10.1177/1740774517711442>.

Mehl, Albert L. 1986. “Interpretation of Traumatic Lumbar Puncture.” *Clinical Pediatrics* 25 (10): 523–26. <https://doi.org/10.1177/000992288602501008>.

Menten, Joris, Marleen Boelaert, and Emmanuel Lesaffre. 2008. “Bayesian Latent Class Models with Conditionally Dependent Diagnostic Tests: A Case Study.” *Statistics in Medicine* 27 (22): 4469–88. <https://doi.org/10.1002/sim.3317>.

Nhu, N. T. Q., D. Heemskerk, D. D. A. Thu, T. T. H. Chau, N. T. H. Mai, H. D. T. Nghia, P. P. Loc, et al. 2013. “Evaluation of GeneXpert MTB/RIF for Diagnosis of Tuberculous Meningitis.” *Journal of Clinical Microbiology* 52 (1): 226–33. <https://doi.org/10.1128/jcm.01834-13>.

Nigrovic, Lise E., Samir S. Shah, and Mark I. Neuman. 2011. “Correction of Cerebrospinal Fluid Protein for the Presence of Red Blood Cells in Children with a Traumatic Lumbar Puncture.” *The Journal of Pediatrics* 159 (1): 158–59. <https://doi.org/10.1016/j.jpeds.2011.02.038>.

Qu, Yinsheng, Ming Tan, and Michael H. Kutner. 1996. “Random Effects Models in Latent Class Analysis for Evaluating Accuracy of Diagnostic Tests.” *Biometrics* 52 (3): 797. <https://doi.org/10.2307/2533043>.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Schumacher, Samuel G., Maarten van Smeden, Nandini Dendukuri, Lawrence Joseph, Mark P. Nicol, Madhukar Pai, and Heather J. Zar. 2016. “Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis.” *American Journal of Epidemiology* 184 (9): 690–700. <https://doi.org/10.1093/aje/kww094>.

Stan Development Team. 2021a. “RStan: The R Interface to Stan.” <https://mc-stan.org/>.

———. 2021b. *Stan Modeling Language Users Guide and Reference Manual*. <https://mc-stan.org/>.

———. n.d. “Prior Choice Recommendations stan-dev/stan Wiki GitHub.” Accessed August 12, 2021. <https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations>.

Stout, Jason E, Yanjue Wu, Christine S Ho, April C Pettit, Pei-Jean Feng, Dolly J Katz, Smita Ghosh, Thara Venkatappa, and Ruiyan Luo. 2018. “Evaluating Latent Tuberculosis Infection Diagnostics Using Latent Class Analysis.” *Thorax* 73 (11): 1062–70. <https://doi.org/10.1136/thoraxjnl-2018-211715>.

Thwaites, GE, TTH Chau, K Stepniewska, NH Phu, LV Chuong, DX Sinh, NJ White, CM Parry, and JJ Farrar. 2002. “Diagnosis of Adult Tuberculous Meningitis by Use of Clinical and Laboratory Features.” *The Lancet* 360 (9342): 1287–92. <https://doi.org/10.1016/s0140-6736(02)11318-3>.

Thwaites, Guy E., Tran Thi Hong Chau, and Jeremy J. Farrar. 2004. “Improving the Bacteriological Diagnosis of Tuberculous Meningitis.” *Journal of Clinical Microbiology* 42 (1): 378–79. <https://doi.org/10.1128/jcm.42.1.378-379.2004>.

Toft, Nils, Erik Jørgensen, and Søren Højsgaard. 2005. “Diagnosing Diagnostic Tests: Evaluating the Assumptions Underlying the Estimation of Sensitivity and Specificity in the Absence of a Gold Standard.” *Preventive Veterinary Medicine* 68 (1): 19–33. <https://doi.org/10.1016/j.prevetmed.2005.01.006>.

van Erp, Sara, Daniel L. Oberski, and Joris Mulder. 2019. “Shrinkage Priors for Bayesian Penalized Regression.” *Journal of Mathematical Psychology* 89 (April): 31–50. <https://doi.org/10.1016/j.jmp.2018.12.004>.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2016. “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC.” *Statistics and Computing* 27 (5): 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.

White, Ian R, James Carpenter, Stephen Evans, and Sara Schroter. 2007. “Eliciting and Using Expert Opinions about Dropout Bias in Randomized Controlled Trials.” *Clinical Trials* 4 (2): 125–39. <https://doi.org/10.1177/1740774507077849>.

Yan, Linlin. 2021. *Ggvenn: Draw Venn Diagram by ’Ggplot2’*. <https://CRAN.R-project.org/package=ggvenn>.

# (APPENDIX) Appendix

# Model Specification Details

## Latent Class Model

For , {, } is the *latent class* and *individual risk of TBM*; , , and } are the *observed value* for ZN Smear, Culturing, and Xpert, respectively. The marginal joint distribution of observations is as follow:

Under the assumption of local independence (model 0 and 1), the probabilities of positive test within class only depend on the latent class itself, thus for .

For model and 2 and 3, we hypothesised that in the TBM positive class (C=1), test results are additionally dependent and only dependent on latent bacillary burden as in (Schumacher et al. 2016), shared between the three tests and in turn regressed on a set of covariables (denoted as ) using a mixed-effect logistic regression so that formula (2) becomes:

s.t.

where is the intercept, is the coefficient of bacillary burden on test results, is the vector of coefficients, and is the individual random effect.

In model 4, we additionally captured the sample-wise fluctuation that cannot be explained by the bacillary burden alone. This fluctuation is regressed by one extra random effect not shared between tests. For numerical stability, we combined the random effect in (4) with this, so that:

Lastly, for model 5, as an exploratory model, we explored all potential impacts of on the sample level.

was a linear combination of potential predictors, which implies a logistic regression for latent binary outcome.

*where X and A are vector of covariates and coefficients.*

with was used as a prior for as we believed that the intercept is bounded by . Let in which M is the number of covariates:

\*where is the penalised scale and one of is , , or .

RE was imposed a standard normal distribution as priors. To prevent sign-switching problem (i.e , was left-truncated at 0)

\*where is the penalised scale and one of is , , or .

Due to the set-up of model 3 and 4, penalisation on fixed effects was tricky as they were all subsequently multiply by *b\_RE*. The prior scales were hence divided by in model 3, and by the average of the three in model 4.

*in which where L is the number of features contributed to the bacillary burden model.*

We used separated penalty term (represented by separated penalised scale and ) for the prevalence model and bacillary burden model. was used as the prior for both, however.

# Simulation study

The purpose of this study is to estimate the perfomance an LCA model in the scenario where manifest variables were partially missing.

In this study, we created as three covariates for latent class . For simplicity, , and were three manifest variables whose data generation processes are locally independent w.r.t. . , however were not completely observed; its observation rate (represented by the binary variable obs\_rate) are different for and .

The data generation process was done in **R** version 4.1 (R Core Team 2021) as below:

# Misc function  
generate\_Y = \(C, probs){  
 Cs = sort(unique(C))  
 sapply(C, \(c) rbinom(1,1, probs[Cs==c]))  
}  
  
# Sample size  
N = 1000   
  
# Create data with 3 predictors, X, X2, and X3  
X = rnorm(N, 0, 1 )  
X2 = rnorm(N, 0, 1 )  
X3 = rbinom(N, 1, .3)  
  
# Create latent class  
probs = plogis(3\*X+X2+5\*X3-1)  
C = sapply(probs, \(p) rbinom(1,1,p))  
  
# Manifest variables  
Y1 = generate\_Y(C, c(.1 ,.3))  
Y2 = generate\_Y(C, c(.01,.5))  
Y3 = generate\_Y(C, c(.05,.8))  
  
# Simulate class-aware missing data for Y3. If obs3==1, Y3 is observed  
obs\_rate = c(.1, .95)  
obs3 = generate\_Y(C, obs\_rate)

The likelihood function of the model shall be

Consider an individual : If :

(as are conditionally independent on ).

Similarly, for those whose :

We considered five sets of prior distribution for , and corresponding for the level of prior knowledge. The formulation of priors follow the syntax in formula (1).

1. Weak prior:
2. Good but conservative knowledge for observation rate, weak prior for manifest variables:
3. Good but conservative knowledge for observation rate and manifest variables:
4. Very strong knowledge for observation rate, weak prior for manifest variables:
5. Very strong knowledge for observation rate and manifest variables:

All models were fitted under the probabilistic language version 2.26 (Stan Development Team 2021b) via inferface (Stan Development Team 2021a).