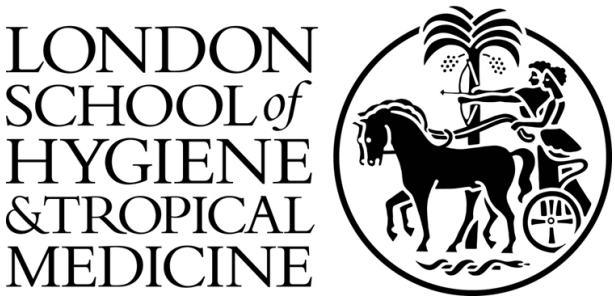


Latent class models for diagnostic test accuracy with application to fever aetiology

Suzanne H. Keddie

London School of Hygiene and Tropical Medicine

Department of Infectious Disease Epidemiology

2024

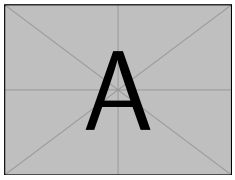
Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London.

Supervisors: **John Bradley, Ruth Keogh and Oliver Bärenbold**

Funded by MRC London Intercollegiate DTP Studentship

Declaration

I, **Suzanne H. Keddie**, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



Suzanne H. Keddie
London, United Kingdom
30th September 2024

Abstract

Febrile illness is a leading cause of health-care seeking and hospital admissions in many settings. Estimates of the extent to which febrile illness could be attributed to different fever causing pathogens are sparse and when they do exist, are limited by studies with small sample sizes, small numbers of diagnostic tests for a small number of potential causes, no controls, and a lack of statistical approaches to use the data gathered and estimate attribution.

The overall aim of this thesis is to further develop the application of Bayesian latent class models for investigating attribution of a syndrome to particular infections. Specifically, the aim is to provide estimates of the extents to which fever-related illness could be attributed to different fever-causing pathogens in four countries. The data used are diagnostic test results from fever cases and controls recruited in the Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) study.

To meet these aims, methodological and applied work using Bayesian latent class models in two different but linked applications is carried out. The first estimates the accuracy of a pre-specified list of diagnostic tests in the absence of a perfect reference standard. This involves a simulation study to investigate the impact of the conditional independence assumption on estimates of diagnostic test sensitivity and specificity. This is an assumption made in simple latent class models. Then, diagnostic test accuracy meta-analysis are applied to estimate the accuracy of each diagnostic test of interest.

The second application of latent class models uses the estimates of test accuracy from the meta-analyses as priors in a model with the observed multivariate imperfect binary diagnostic test data from cases and controls. The combined application of latent class models allows estimation of the fraction of fever cases attributed to different fever-causing pathogens from imperfect diagnostic tests.

Acknowledgements

I am very grateful for the support and guidance of my supervisors John Bradley, Ruth Keogh and Oliver Bärenbold throughout my PhD. This thesis would not have been possible without your encouragement and expertise and I am very grateful to have had all three of you as supervisors.

During this PhD, I have been fortunate to work with the Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) study. Thanks to Heidi Hopkins and David Mabey for welcoming me into the FIEBRE study with open arms and for all your advice along the way. It was also a pleasure to work with everyone involved with the FIEBRE study at LSHTM and at the study sites including (but not limited to) Christian Bottomley, Polycarp Mogeni, Elizabeth Fitchett, Ruth Lorimer, Michael Lawrence, Tegwen Marlais, John Crump, Elizabeth Ashley and Nicholas Feasey.

I would also like to thank Lauren Dalton for answering any and all of my queries related to studying at LSHTM or to my funding, we never met in person but Lauren made my PhD journey much easier.

I also want to thank the Medical Research Council for funding this PhD studentship and a placement at the University of Leicester.

Finally, thanks to friends and family who have supported me throughout this PhD. To Kate and Katherine for our weekly zoom coffee while we began our PhDs from home. Thanks to my parents for continuing to review my work for grammatical mistakes, to Matthew for your unwavering support and to Pablo for forcing me to take breaks and give him scratches.

Contents

Table of Abbreviations and Notation

Notation	Description
FIEBRE	Febrile Illness Evaluation in a Broad Range of Endemicities study.
PERCH	Pneumonia Etiology Research for Child Health study.
TP	True positive.
FN	False negative.
FP	False positive.
TN	True negative.
Se	Sensitivity.
Sp	Specificity.
LCM	Latent Class Models.
HIV	Human Immunodeficiency Virus.
AIC	Akaike Information Criterion.
DIC	Deviance Information Criterion.
LOO	Leave-One-Out cross validation.
elpd	Expected log predictive density.
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis.
QUADAS	Quality Assessment of Diagnostic Accuracy Studies.
SROC	Summary Receiver Operating Characteristic.
HSROC	Hierarchical Summary Receiver Operating Characteristic model.
ROC	Receiver Operating Characteristic.
DAT	Direct Agglutination test.
ELISA	Enzyme-linked immunosorbent assay.
qPCR	Quantitative polymerase chain reaction.

Notation	Description
NUTS	No-U-Turn sampler.
HMC	Hamiltonian Monte Carlo.
RSV	Respiratory Syncytial Virus.
CrAg	Cryptococcosis.
RPP	Respiratory Pathogen Panel.
MAT	Microscopic Agglutination Test.
PCR	Polymerase chain reaction.
RT-PCR	Reverse transcription polymerase chain reaction.
OR	Odds ratio.
PAF	Population attributable fraction.
VNT	Viral neutralization test.
GEMS	Global Enteric Multicenter Study.
RDT	Rapid diagnostic test.
JEV	Japanese Encephalitis Virus.
ONNV	Onyongnyong virus.
UTI	Urinary tract infection.
CrI	Bayesian 95% credible interval.
NOS	Not specified latent subgroup.
E.coli	Escherichia coli.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overall aim

The overall aim of this thesis was to provide estimates of the extents to which fever-related illness could be attributed to different fever-causing pathogens, in four countries, using binary diagnostic test data from fever cases and controls.

Three challenges identified included uncertainty about the accuracy of different diagnostic tests, multiple diagnostic tests used to detect the same infection and the possibility of co-infections. To address these challenges, I carried out methodological work on diagnostic test accuracy estimation in the absence of a perfect reference standard. This work was then applied in various diagnostic test accuracy meta-analyses. Finally, practical work on Bayesian estimation of the causes of fever from case-control data was needed that allowed for multiple imperfect diagnostic tests and the possibility of co-infections.

In this introductory chapter, I look at the context of this research, providing a brief introduction to febrile illness and diagnostic test accuracy before introducing latent class modelling which provides the framework for all statistical analyses conducted. The final sections of this chapter present the aims, objectives and outline of the rest of the thesis.

1.2 Febrile Illness

Febrile illness (henceforth referred to as fever), is a common symptom of infectious and non-infectious disease [1] and is a challenge for public health care systems globally. National Health Service England define a fever as a temperature of 38°C or more in both children and adults [2]. Not only is fever a leading cause of health-care seeking and hospital admissions in many settings, with an estimated 16 million hospital ad-

missions due to febrile-illness across sub-Saharan Africa in 2014 [3], it is clear from reports, like the Global Burden of Disease, that febrile illness leads to considerable morbidity and mortality [4].

A fever can be caused by many different pathogens including blood and tissue parasites, invasive bacterial infections, invasive fungal infections, bacterial zoonoses and viral infections [5]. The principal causes vary between locations [6, 7] and in many places are highly seasonal [8, 9]. The leading causes of fever in a place and time also differ depending on the patient population of interest; for example, the distribution was found to be different in adults compared to children in Tanzania [10, 11] and was found to be different in immunocompromised patients in Vietnam [12] compared to a study that excluded human immunodeficiency virus (HIV) infected patients [13].

Identifying the cause of fever is important for providing appropriate timely treatment at the individual level and ensuring appropriate resource allocation at the societal level. However, this identification is often very challenging because of limited point-of-care diagnostic capacity for many common causes of fever and the fact that it is difficult to determine the cause of a fever from clinical presentation. Examples of common infections that present with a fever not already mentioned include respiratory infections and urinary tract infections but fever can also be present with autoimmune or inflammatory disorders, cancer or even just due to use of antihistamines [14]. The negative outcomes of incorrect or missed diagnoses for individuals and society include excess morbidity and mortality, inefficient allocation of resources, and the growing spread of antimicrobial resistance by a "just-in-case" antibiotic prescription strategy [15]. Improving the epidemiological understanding of febrile illness is, therefore, an important research topic.

Historical investigations into the causes of fever in sub-Saharan Africa and South East Asia, overwhelmingly associated fever with malaria to the extent that the two words were almost synonymous; but as malaria case incidence has declined in these settings [16] and reliable point-of-care diagnostic tests for malaria are widely available [17], it is clear that assuming fever is caused by malaria is not tenable. As a result, more recent research has investigated the causes of non-malaria febrile illness [18, 19]. While these studies highlighted that malaria is not the predominant cause of febrile illnesses and that bacterial and arbovirus infections have been overlooked, they have been limited to single country studies and/or limited to a small number of diagnostic tests for a small number of potential causes of fever [18, 19]. These limited studies restrict

the generalizability of the research outputs and therefore, have limited ability to improve health outcomes for fever more generally.

The research presented in the current thesis uses data from the Febrile Illness in a Broad Range of Endemicities (FIEBRE) study [20], a large four-site case-control investigation into the causes of fever in children and adult outpatients and inpatients. The FIEBRE study used a large number of diagnostic tests at the point-of-care and at established reference laboratories across three continents, to look for over 50 different pathogens as potential causes of fever. The primary aim of FIEBRE is to improve the understanding of the causes of fever across the study sites. To utilise the large data set of observed results from diagnostic tests to estimate the leading causes of fever, it is important to adjust for imperfect diagnostic tests; those tests that incorrectly classify a subset of individuals as false positive or false negative. As a result, estimating the accuracy of the diagnostic tests used in FIEBRE is a fundamental part of this research.

1.3 Diagnostic test accuracy

Diagnostic tests aim to discriminate between individuals with and without diseases and have become a mainstay in medical decision-making. A perfect diagnostic test is assumed to correctly diagnose the true disease status of all individuals. That is, the test is 100% sensitive (defined as the ability of a test to correctly identify those individuals that have a disease) and, 100% specific (defined as the ability of a test to correctly identify those individuals who do not have a disease). However, perfect diagnostic tests are rare and do not exist for many infections. Where they do exist, they are often invasive procedures that take considerable time, money and resources to get results from, for example, a biopsy. As a result, the most commonly used diagnostic tests are not perfect but deliver quick results at the point-of-care. Therefore, it is important to remember that they are only a proxy for an individual's true disease status.

Given that diagnostic tests are frequently imperfect, their interpretation presents a challenge to health care workers. This is also a key statistical and epidemiological challenge when using the results of diagnostic tests to estimate the accuracy of a new test, or the prevalence of an infection from multiple imperfect diagnostic tests or, in the FIEBRE study, to estimate the proportion of fever cases attributable to a particular infection.

The traditional approach used to evaluate the accuracy of diagnostic tests, and the approach taught in most statistics classes, is to compare the results of the test of interest against a known reference standard, often called a gold-standard, which is assumed perfect (i.e. has 100% sensitivity and specificity). However, given that most tests are imperfect, this scenario is rare in practice and instead the accuracy of a new diagnostic test is usually evaluated against an imperfect test. This results in biased estimates of the sensitivity and specificity of the new test, with knock on effects for the quantity of interest in a study [21, 22, 23].

The challenge of imperfect diagnostic tests has been recognised for several decades and multiple methods have been proposed over the last thirty years to enable unbiased estimation of diagnostic test accuracy in the presence of imperfect reference tests. These methods include discrepancy resolution and composite reference standards [24]. Discrepancy resolution relies on a third 'resolver' test when the test under evaluation and the reference test disagree. Composite reference standards explicitly define a reference test which is a combination of imperfect reference tests. However, both methods ultimately cause their own biases [25]. Instead, one of the two the recommended methods in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [26]) for estimating diagnostic accuracy in the presence of imperfect reference tests stems from latent class models.

1.4 Latent class models

Latent class models are statistical models that are applied when the presence of subgroups within a population is assumed and cannot be directly observed. Latent class models also assume that membership to an unobserved subgroup is explained by patterns in responses to observed variables. Latent class models offer a flexible framework that has been used across medical health research including: understanding vaccine hesitancy [27], understanding risk of attrition in doctoral students [28], and exploring patterns in sleep problems [29]. It is often useful to think of latent class models hierarchically, where the basic principles are twofold. Firstly, there is a latent indicator that specifies the class, state, mixture or subgroup to which an individual belongs identified by a categorical latent variable [30]. For an individual i we denote this latent (i.e. true) subgroup as Z_i. Z_i takes a value $1, 2, .., C$ denoting C latent subgroups and:

$$\alpha_c = \Pr(Z = c) \tag{1.4.1}$$

where, α is a simplex with, $\sum_{c=1}^{C} \alpha_c = 1$. Secondly, the observed data Y are modelled conditional on the latent indicator Z. For example, assuming binary data:

$$Y|Z = c \sim Bernoulli(p_c) \tag{1.4.2}$$

where, p_c is a subgroup-specific parameter vector. Throughout the research in the current thesis, this parameter is how we incorporate the sensitivity and specificity of a diagnostic test. The output of a latent class model is a probability of class membership. This is unlike clustering algorithms where the outcome is a discrete class membership [31]. Estimating the parameters of interest in latent class models is inherently challenging due to issues of identifiability with many unknown parameters to estimate and few degrees of freedom. To get around this problem, additional information is often required. This could be in the form of additional diagnostic test results, or the same tests applied simultaneously in multiple populations or, by leveraging a Bayesian approach and incorporating informative prior distributions for some parameters [32].

First introduced by Hui and Walter in 1980 [33], latent class models are now commonly used to assess diagnostic accuracy in the absence of a gold-standard reference test across a wide range of diseases [34, 35, 36]. Latent class models have also more recently been used to estimate the causes of pneumonia in the Pneumonia Etiology Research for Child Health (PERCH) study [37, 38]. In this thesis, latent class models will be used in diagnostic test accuracy meta-analyses to estimate the sensitivity and specificity of diagnostic tests of interest using summary-level data from different studies. Latent class models will then also be used to estimate the causes of fever in the FIEBRE case-control study, combining the observed test results with the estimates from diagnostic test accuracy meta-analysis, adapted from the approach used in the PERCH study [37].

1.5 Aims and objectives

The overall aim of this thesis is to further develop the application of Bayesian latent class models for investigating attribution of a syndrome to particular infections. Specifically, the aim is to provide estimates of the extents to which fever-related illness could be attributed to different fever-causing pathogens in four countries using diagnostic test data from fever cases and controls. To achieve this goal, the specific objectives are listed below and a diagrammatic overview of the main investigations is provided in Figure 1.1:

1. Review how diagnostic test accuracy can be estimated in the presence of imperfect reference tests using Bayesian latent class analyses (Chapter 2)

2. Explore the impact of key latent class model assumptions on estimates of diagnostic test sensitivity and specificity through a simulation study (Chapter 3)

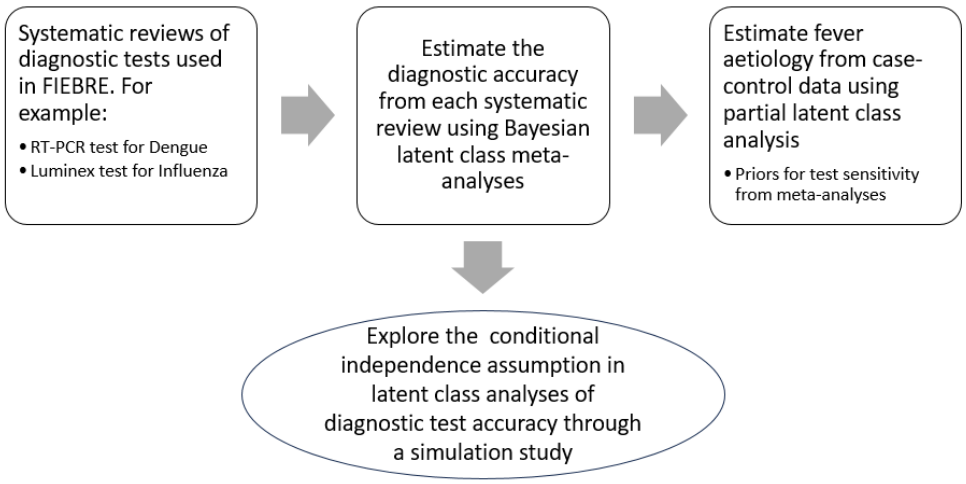
3. Estimate the sensitivity and specificity of diagnostic tests used in the FIEBRE study using Bayesian random-effect latent class meta-analyses (Chapters 4 and 5)

4. Review aetiology research methodologies for investigating the causes of syndromes (Chapter 6)

5. Estimate the aetiology of fever in the FIEBRE data set using Bayesian partial latent class analysis (Chapters 7 and 8)

6. Compare the strengths and limitations of two different statistical approaches, one a Bayesian partial latent class analysis and the second a more traditional Frequentist analysis to estimating fever aetiology (Chapter 9)

Figure 1.1: Diagrammatic overview of the main analyses within this thesis



1.6 Outline

The thesis is split into three parts. Part 1 (Chapters 2-5) focuses on latent class models for estimating the sensitivity and specificity of diagnostic

tests in the absence of gold-standard reference tests. Part 2 (Chapters 6-9) uses latent class models to estimate the causes of fever from a case-control study. Finally, Part 3 summarises the findings of the thesis. The rest of this thesis is organized as follows:

1.6.1 Part 1: Diagnostic Test Accuracy Estimation

Chapter 2 introduces the statistical concepts involved in diagnostic test accuracy estimation in the absence of a gold-standard reference test with an emphasis on latent class models.

Chapter 3 (published manuscript) presents a simulation study exploring how the fundamental assumption made in latent class models of statistical independence between tests conditional on the true disease status, or lack thereof, impacts the resulting estimates of diagnostic test sensitivity and specificity in the single-study setting.

Chapter 4 reviews latent class models for estimating sensitivity and specificity of diagnostic tests within meta-analyses, describing the statistical methods and outlining how the model results can be presented to aid their interpretation.

Chapter 5 (published manuscript) discusses the application of the meta-analyses methods presented in chapter 4 to five different systematic reviews for diagnostic test accuracy before presenting an application of the methods to estimate the accuracy of the Microscopic agglutination test in the diagnosis of Leptospirosis. I also presented the findings of the Dengue diagnostic accuracy review through a poster presentation at the Royal Society of Tropical Medicine and Hygiene annual meeting 2023.

1.6.2 Part 2: Fever Aetiology Estimation

In **Chapter 6**, I discuss what aetiology research is and the currently available statistical methods that can be used in aetiology research.

Chapter 7 introduces the FIEBRE case-control study including a basic description of the data and outlining my data-management and analysis role in this study.

In **Chapter 8** the statistical methods in chapter 7 are applied to the FIEBRE case-control study. I presented this work through an oral presentation at the 44^{th} annual conference of the International Society for Clinical Biostatistics in Milan, Italy, in person in 2023.

Chapter 9 compares the results of the aetiology estimation from the method used in chapter 7 with an alternative frequentist statistical approach and discusses the strengths and limitations of both.

1.6.3 Part 3: Thesis summary

Finally, **Chapter 10** summarizes the findings and contributions of this thesis and discusses future directions for latent class models in diagnostic test accuracy estimation and aetiology research.

Part I

Diagnostic Test Accuracy Estimation

Chapter 2

Statistical and epidemiological considerations

2.1 Preamble

This chapter reviews the literature on diagnostic test accuracy estimation in the single-study setting. That is, when multiple tests are applied to the same individuals from the same population. To estimate the sensitivity and specificity of a new or existing diagnostic test, this requires the comparison of test results from more than one test carried out on the same individuals, the number of tests required depends on the statistical method chosen and is discussed in the following chapter. Although ordinal and continuous diagnostic tests exist, the focus throughout this thesis is on binary diagnostic tests that are the most commonly relied upon tests for clinical decision making.

Overall, this chapter highlights the importance of diagnostic tests in health care delivery as well as the importance of understanding the accuracy of diagnostic tests. The statistical concepts that underpin diagnostic test accuracy estimation and the methods that are used in the proceeding chapters will also be introduced.

2.2 Diagnostic tests and their importance

Diagnostic tests aim to discriminate between individuals with and without disease. Not only are they a primary tool relied upon at the point of care for treatment prescriptions but also within national public health programs to estimate disease prevalence. This can be evidenced within the COVID-19 pandemic where diagnostic tests were used to detect outbreaks and determine which individuals must self-isolate [39, 40]. The result of a diagnostic test is typically a categorization, most commonly

dichotomous, of an underlying continuous variable, for example, parasite load or the presence of antibodies. With an underlying continuous variable, the designation of truly infected or not infected is not always clear cut. Diagnostic test results are therefore, only a proxy for an individual's true disease status and are frequently imperfect [41].

While it is easy to see the need for diagnostic tests, their imperfect nature can have negative consequences [42]. A false positive result can have negative health, financial and psychological effects on an individual and wider societal consequences, including the overestimation of disease resulting in misleading policy guidance and health care interventions. On the other hand, a false negative result at the societal level, may lead to underestimates of disease prevalence and failure to interrupt the chain of transmission (where relevant), while at the individual level, it may lead to a person not receiving treatment for a condition with serious health consequences. Although much time and effort is spent developing the most accurate diagnostic tests, the tests must combine accuracy with accessibility, ease of use, cost and speed of result. Therefore, the development of new diagnostic tests is always ongoing.

Diagnostic tests, like all private goods, are excludable and rival. As a result, not all diagnostic tests are available in all locations, to all individuals, for all infections. The COVID-19 pandemic highlighted inequities in access to diagnostic tests. It was reported in 2023 that only 20.4% of all COVID-19 tests administered worldwide have been used in low- and lower-middle- income countries, despite these countries comprising 50.6% of the global population [43], and this situation is not unique to COVID-19 [44]. Low-income and middle-income countries have a disproportionately large share of the global burden of disease but a disproportionately low share of global health resources [45] and a recent Lancet commission on diagnostics found that 47% of the global population has little to no access to diagnostics [46]. This is despite recognition at the highest levels that diagnostics are fundamental to delivering quality health care [47]. Increased recognition of the importance of diagnostic tests for health care has stimulated the development of new diagnostic tests and calls for enhanced diagnostic testing capacity across all regions. With this stimulation in diagnostic tests, robust methods to assess diagnostic test accuracy are needed.

2.3 Measuring diagnostic test accuracy

The traditional approach used to evaluate new diagnostic tests is to compare the results of a new test against a known reference standard, often called a gold-standard, which is assumed perfect. That is, the reference test is assumed to correctly diagnose the true disease status of all individuals. We denote the true disease status of an individual Z where $Z = 1$ denotes diseased and $Z = 0$ denotes not diseased and we introduce $T = 1$ to denote a positive result on a diagnostic test and $T = 0$ a negative result. The contingency table (See Table 2.1) for the outcomes of these two tests can then be used to estimate various estimands of interest (See Table 2.2 for some of these).

Table 2.1: Example of a $2x2$ contingency table for the cross-classification of results from a gold-standard reference test, that correctly classifies all individuals so that a positive result on the gold-standard reference test is equivalent to Z, and a new test performed on N individuals

New test	**Gold-Standard Reference Test**		**Total**
	Positive	Negative	
Positive	True Positive (TP)	False Positive (FP)	TP + FP
Negative	False Negative (FN)	True Negative (TN)	FN + TN
Total	TP + FN	FP + TN	TP+FP+FN+TN

Throughout this thesis, I focus on sensitivity and specificity, that are the most widely used measures of diagnostic accuracy. Sensitivity (Se) is the probability of a positive test result given the subject has the disease while specificity (Sp) is the probability of a negative test result given the subject does not have the disease. Using conditional probability notation we can write these as $Se = \Pr(T = 1|Z = 1)$ and $Sp = \Pr(T = 0|Z = 0)$ respectively. The true disease prevalence is then $\Pr(Z = 1)$. Throughout this thesis, and in general, false positivity rate is also referred to which is 1-minus the specificity (i.e. $\Pr(T = 1|Z = 0)$). Estimating sensitivity and specificity using conditional probabilities based on the reference test assumes that the reference test is perfect and requires that both tests have been applied to the same individuals. When the reference test is not perfect these estimates (shown in Table 2.2) are biased due to the misclassification of individuals [48]. Failure to recognise bias as a result of misclassification or other measurement error is a phenomenon across epidemiology research and not novel to diagnostic tests [49].

Table 2.2: Diagnostic accuracy estimands that can be estimated from a $2x2$ contingency table for the cross-classification of results from a reference test and a new test performed on N individuals

Estimand	**Definition**	**Definition in terms of conditional probability**	**Estimator**
Sensitivity (Se)	Probability that a subject who is truly diseased receives a positive test	$\Pr(T=1\|Z=1)$	$\frac{TP}{TP+FN}$
Specificity (Sp)	Probability that a subject who is truly not diseased receives a negative test	$\Pr(T=0\|Z=0)$	$\frac{TN}{TN+FP}$
Positive Predictive Value	Probability that a subject who tests positive actually has the disease	$\Pr(Z=1\|T=1)$	$\frac{TP}{TP+FP}$
Negative Predictive Value	Probability that a subject who tests negative is actually disease free	$\Pr(Z=0\|T=0)$	$\frac{TN}{TN+FN}$
Positive Likelihood Ratio	Ratio of the proportion that test positive among the diseased to those that test positive without disease	$\frac{\Pr(T=1\|Z=1)}{\Pr(T=1\|Z=0)}$	$\frac{TP/(TP+FN)}{1-TN/(TN+FP)}$
Negative Likelihood Ratio	Ratio of the proportion that test negative among the diseased to those that test negative without disease	$\frac{\Pr(T=0\|Z=1)}{\Pr(T=0\|Z=0)}$	$\frac{1-TP/(TP+FN)}{TN/(TN+FP)}$
Diagnostic Odds Ratio	Ratio of the odds of positivity among those diseased to those not diseased	$\frac{(\frac{\Pr(T=1\|Z=1)}{\Pr(T=1\|Z=0)})}{(\frac{\Pr(T=0\|Z=1)}{\Pr(T=0\|Z=0)})}$	$\frac{(\frac{TP/(TP+FN)}{1-TN/(TN+FP)})}{(\frac{1-TP/(TP+FN)}{TN/(TN+FP)})}$
Prevalence (π)	Proportion of truly diseased subjects in a population	$\Pr(Z=1)$	$\frac{TP+FN}{N}$

Note. The new test is denoted T where 1/0 indicates positive/negative and the true disease status indicated by the result of a perfect reference test is denoted Z where 1/0 indicates diseased/non-diseased. TP, FN, FP, TN denote true positive, false negative, false positive and true negative respectively. N denotes the total number of individuals tested and is the sum of TP, FN, FP and TN.

Assuming there is non-differential misclassification (that is, when the error in the classification of individuals as diseased or not diseased occurs at the same rate in both the truly diseased and truly not diseased individuals), when the reference test is incorrectly assumed perfect, the new test's accuracy is underestimated [50]. For example, suppose there are two diagnostic tests applied to a population of 200 individuals where the disease prevalence is known at 10% and the sensitivity and specificity of our reference test are $\hat{Se} = 65\%$ and $\hat{Sp} = 90\%$ and the true sensitivity and specificity of our new test are $\hat{Se} = 80\%$ and $\hat{Sp} = 70\%$. Given this information we can make a 2x2 table between each test and the truth (See Tables 2.3a and 2.3b). Then, under an assumption of conditional independence, the observed 2x2 table for the new test and the reference standard would be as per 2.3c. If we use both the reference test and the new test in this population and assume our reference test is perfect the estimate of our new test's sensitivity would be $Se = 15/31 = 48\%$ and specificity $Sp = 114/169 = 67\%$ when in truth our new test's sensitivity and specificity are 80% and 70% respectively. This example shows that the new test's accuracy is, and will always be, underestimated when using the traditional approach which incorrectly assumes our reference test is perfect and assuming there is non-differential misclassification.

Several methods have been proposed over the last thirty years to enable estimation of diagnostic accuracy in the presence of imperfect reference tests. A recent systematic review of diagnostic test evaluation methodology grouped these methods into three categories [51]:

- Correction methods
- Methods employed when using multiple imperfect reference standards
- Other methods

Correction methods were those applied when an imperfect reference test with known sensitivity and specificity is used which allows for adjustment of the estimated accuracy of the test under evaluation [52, 48]. However, just as it is unlikely that a diagnostic test is perfect, it is unlikely that the true sensitivity and specificity for a test is known across different population groups and different laboratories such that the correction applied would remove all bias. Further, while Brenner's [48] proposed correction method could account for two tests that were positively correlated the Staquet *et al.* [52] correction method could not and would result in biased estimates if the two tests were positively correlated.

Table 2.3: Numerical example of the impact of misclassification in a reference test on the accuracy of a new diagnostic test under the assumption of conditional independence

(a) Contingency table between a reference test (Se = 65% and Sp=90%) and the true disease status of 200 individuals ($\pi = 10\%$)

Reference test	**True status**		**Total**
	positive	negative	
Positive	13 (65%)	18 (10%)	31
Negative	7 (35%)	162 (90%)	169
Total	20 (100%)	180 (100%)	200

(b) Contingency table between a new test (Se = 80% and Sp=70%) and the true disease status of 200 individuals ($\pi = 10\%$)

New test	**True status**		**Total**
	positive	negative	
Positive	16 (80%)	54 (30%)	70
Negative	4 (20%)	126 (70%)	130
Total	20 (100%)	180 (100%)	200

(c) Contingency table between a new test (Se = 80% and Sp=70%) and a reference test (Se = 65% and Sp=90%)

New test	**Reference test**		**Total**
	positive	negative	
Positive	15 (48%)	55 (33%)	70
Negative	16 (52%)	114 (67%)	130
Total	31 (100%)	169 (100%)	200

Methods employed when using multiple imperfect reference standards include discrepancy resolution [53], composite reference standards [24], and latent class models [54]. Discrepancy resolution is where disagreement between two tests results in a third 'resolver' test. The third 'resolver' test is only applied to individuals who had discordant results on the first two tests. Composite reference standards are when multiple imperfect reference tests are combined with the goal of defining a better reference test [24]. Both discrepancy resolution and composite reference standards have been shown to create their own biases [25]. Latent class models offer a flexible framework that can account for imperfect reference tests without introducing the biases that discrepancy resolution and composite reference standards introduce. However, latent class models can still introduce bias if incorrect assumptions are made. Potential sources of bias from latent class models are discussed more in Section 2.4 and Chapter 3.

The "other methods" group includes considering alternative study designs like a validation or case-control study, a study of agreement which looks solely at concordance between two tests and, test positivity rate

which calculates the proportion of participants with a positive test result. Each of these avoids the challenge of imperfect reference tests rather than address it.

Throughout the rest of this thesis, I focus on estimating diagnostic accuracy in the presence of imperfect reference tests using latent class models. This approach is increasingly used when reference tests are known to be imperfect and is also the most suitable approach in this thesis as they can be extended to the meta-analysis setting to estimate the accuracy of a diagnostic test from multiple different studies and, to the eventual aetiology study.

2.4 Latent class models for estimating diagnostic test accuracy

Latent class models (LCM) were first applied to diagnostic test accuracy by Hui and Walter in 1980 [33]. LCM for diagnostic test accuracy have been compared to alternative approaches in multiple reviews [55, 51] and have been applied across a wide range of diseases [36, 35, 34]. Following on from the introduction chapter (See section 1.4), in these latent class models we define two latent subgroups, *diseased* and *not diseased*, hereafter referred to as disease states. Every individual belongs to exactly one disease state identified by the latent indicator Z. We then observe the results of two or more different diagnostic tests carried out on the same set of individuals. This could be designed to compare a new test against a current reference standard but equally this could be designed as repeat applications of the same test investigating the accuracy of different readers. The observed data are modelled conditional on the latent indicator and the probability of disease state membership depends on the probability of a positive result for each test carried out (1.4.2).

The methods described in this section will relate to the scenario where we have two binary diagnostic tests, however they are easily extended to more than two. Simple LCMs make two assumptions. First, that observed test results are imperfect measures of an underlying not directly observable (true disease) state (Z). While we focus on the most common application in which two disease states are considered, more than two states have been examined [56, 57]. The second assumption is that, conditional on the true disease state, the results from different tests performed on an individual are independent. For example, with two tests, the result of one test (T_1) has no bearing on the result of the second test (T_2) given the true disease state of an individual (i.e. the assump-

tion that was used to derive Table 2.3c, such that we can write the joint distribution as:

$$\Pr(T_1 = t_1, T_2 = t_2|Z) = \Pr(T_1 = t_1|Z) \Pr(T_2 = t_2|Z) \tag{2.4.1}$$

There are four possible combinations of test results when considering two diagnostic tests. Below I show how each of these can be written in terms of the sensitivity and specificity for each test, with test indicated by a subscript (i.e. Se_1 indicates the sensitivity of diagnostic test 1), and the true disease prevalence denoted (π). For now, I assume independence between the two tests, conditional on the true disease status (i.e. the result of one test provides no information about the result of the other test on the same individual conditional on the true disease status of that individual). The assumption of conditional independence and its plausibility in the diagnostic test scenario is discussed more in section 2.4.2.

$$\begin{aligned}
\Pr(T_1 = 1, T_2 = 1) =& \pi Se_1 Se_2 + (1-\pi)(1-Sp_1)(1-Sp_2) \\
\Pr(T_1 = 0, T_2 = 1) =& \pi(1-Se_1)Se_2 + (1-\pi)Sp_1(1-Sp_2) \\
\Pr(T_1 = 1, T_2 = 0) =& \pi Se_1(1-Se_2) + (1-\pi)(1-Sp_1)Sp_2 \\
\Pr(T_1 = 0, T_2 = 0) =& \pi(1-Se_1)(1-Se_2) + (1-\pi)Sp_1 Sp_2
\end{aligned} \tag{2.4.2}$$

Combining these equations, we can write the likelihood function for the observed data given our parameters as:

$$\begin{aligned}
&P(T_1, T_2|\pi, Se_1, Se_2, Sp_1, Sp_2) = \\
&\quad \prod_{i=1}^{N}((\pi Se_1^{t_{1i}} Se_2^{t_{2i}}(1-Se_1)^{(1-t_{1i})}(1-Se_2)^{(1-t_{2i})}) + \\
&\quad ((1-\pi)Sp_1^{(1-t_{1i})} Sp_2^{(1-t_{2i})}(1-Sp_1)^{t_{1i}}(1-Sp_2)^{t_{2i}}))
\end{aligned} \tag{2.4.3}$$

The above likelihood can be extended from two to R tests by introducing an index for test $r(r = 1, .., R)$ and re-defining Se and Sp as vectors:

$$\begin{aligned}
&P(T_1, .., T_R|\pi, Se, Sp) = \\
&\quad \prod_{i=1}^{N}((\pi \prod_{r=1}^{R} Se_r^{t_{ri}}(1-Se_r)^{(1-t_{ri})} + \\
&\quad ((1-\pi)\prod_{r=1}^{R} Sp_r^{(1-t_{ri})}(1-Sp_r)^{t_{ri}}))
\end{aligned} \tag{2.4.4}$$

2.4.1 Estimation and Identifiability

In section 2.4, the outlined methods show that with two diagnostic tests, we are seeking to estimate five unknown parameters (see equations 2.4.2, 2.4.3); sensitivity and specificity of each test as well as an overall prevalence. Since both tests are binary, the observed data can be expressed by 2^R cross-classifications and there are 2^R-1 degrees of freedom. The number of unknown parameters to be estimated can be expressed as $2R+1$. This means that when we evaluate just two binary diagnostic tests, the number of parameters to estimate ($2R+1=5$) is greater than the degrees of freedom ($2^R-1=3$). To estimate the unknown parameters in this scenario, LCMs have two options, the first is to use a frequentist approach and apply constraints to certain parameters. When two tests are evaluated, at least two parameters must be fixed to allow estimation of the remaining three parameters. Examples of imposed constraints include assuming a known sensitivity and specificity of one test or assuming a known prevalence. After sufficient constraints have been imposed (so that the number of parameters to estimate is equal to or less than the degrees of freedom), estimation via maximum likelihood can then proceed as normal [33].

The second option, is to use a Bayesian approach where the use of prior distributions combined with the likelihood can derive a posterior distribution for all unknown parameters using Bayes' theorem and a numerical approach such as iterative Markov-chain Monte Carlo [32] used for estimation. That is, if we use a Bayesian framework, identifiability (degrees of freedom $\geq$ the number of parameters to estimate) is not mandatory because of the addition of prior distributions [58].

The challenge described above, where there are fewer degrees of freedom than parameters, is defined as non-identifiability. Lack of identifiability, specifically, structural identifiability [59], is when multiple values of parameter estimates correspond to the same distribution of observed data [60] and even collection of infinite data is not able to reveal the unique values of the parameters [61]. When the number of parameters to estimate is greater than the degrees of freedom, with no additional information, a frequentist latent class model is not identified and constraints that assume certain parameters are fixed have to be imposed in order for the model to be identified. Alternatively, results from at least three tests [50], or application of the two tests in at least two different populations [33] is needed to ensure identifiability. However, if we use a Bayesian framework, identifiability (degrees of freedom $\geq$ the number of parameters to estimate) is not mandatory because of the addition of prior

distributions [58] and estimation of diagnostic test accuracy from latent class models that would not be identified in a frequentist framework, can be identified [32]. While the use of informative priors can ensure inferences can be made from a non-identifiable model, care must still be taken because incorrect prior information can still lead to flawed inferences.

The frequentist approach with constraints can be seen as nested within the Bayesian approach since the constrained parameters can be seen to have prior distributions with a probability mass equal to one in their constrained values [32]. Using the Bayesian approach with these same prior distributions will result in numerically identical point estimates, however, assuming a known value for the sensitivity and/or specificity of a test is unlikely to ever be correct and by doing so the interval estimates for our unconstrained parameters do not account for the uncertainty in these values. While the frequentist approach is less computationally expensive, it is rare that we know the exact sensitivity and specificity of a test. Instead, it is likely that we have some reliable information about the accuracy of a test for example, a high specificity for microbiological tests. In these cases, this information can and should be expressed as a prior distribution and included in our Bayesian model to improve estimates and interpretability of the sensitivity and specificity of a new test [62]. In latent class models for diagnostic test accuracy estimation, identifiability is a habitual challenge given the number of unknown parameters but it is also customary that some reliable information on a diagnostic test's accuracy exists. As a result, the use of a Bayesian approach in diagnostic test accuracy estimation is commonplace.

Bayesian inference

Bayesian inference is grounded in Bayes' rule and is encapsulated by the following example. For a given set of parameters θ and data Y we can write a joint probability distribution for θ and Y as the product of the prior distribution $p(\theta)$ and the sampling distribution $p(Y|\theta)$, where p indicates a probability distribution. Then, conditional on the known value of the data Y, using Bayes' rule and omitting the constant factor $p(Y)$, the unnormalised posterior density is:

$$p(\theta|Y) \propto p(\theta)p(Y|\theta) \tag{2.4.5}$$

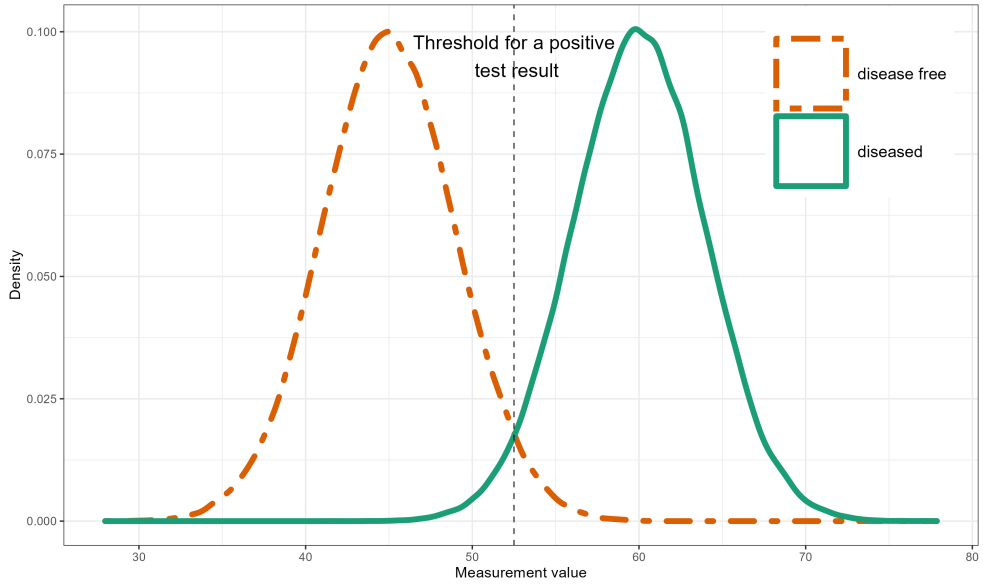
The most controversial part of Bayesian inference is the prior ($p(\theta)$). The prior is a probability distribution that represents our pre-data uncertainty for a parameter's true value and is required for all unknown parameters. The goal of these priors is to include information so that

only possible values are allowed while including all plausible values. Suppose we want to construct a prior distribution for the sensitivity of an imperfect reference test. The possible values set bounds so that the parameter values can only fall $\in [0, 1]$. Given these bounds a uniform distribution between 0 and 1 could be used. However, if we have prior knowledge that this is a commonly used diagnostic test that has been previously evaluated and been recommended for use for many years then a prior that gives equal weight to the test sensitivity being 0% as it does to the test's sensitivity being 50% makes less sense. In fact, the principle of insufficient reason states that a uniform distribution is only appropriate if nothing is known about a parameter which is often not the case [63]. Lack of any information about a parameter is rarely the case with a diagnostic test's accuracy and in this thesis, we choose to use prior distributions for test sensitivity and specificity that rely on the beta distribution, the conjugate prior probability distribution for the binomial distribution used in the likelihood component of our model.

2.4.2 The conditional independence assumption

The conditional independence assumption is the assumption that the result of one test from an individual provides no information about the result of another test on the same individual (conditional on the true disease status). An alternative definition, is that the misclassification errors of the tests are unrelated [23]. This is a strong assumption and it is unlikely to hold in practice [23, 64]. This is highlighted when we consider there is typically a spectrum of disease severity, particularly when the underlying variable used to determine severity is continuous. For a disease like malaria, this would be malaria parasitaemia, for human immunodeficiency virus (HIV) this could be a CD4 count and for cancers it could be tumor size. In these cases, you would expect the most severe cases to be detected with high probability by both tests if the factor that makes individuals experience more severe disease makes the disease easier to detect. If the disease is easier to detect in certain individuals this means that the test sensitivity is not constant in a population and is more sensitive in more severe cases. This idea of a test being more sensitive in more severe cases due to an underlying continuous factor is shown in Figure 2.1. When this happens, the conditional independence assumption fails [65] and failure to account for dependence results in biased estimations of the tests' accuracy [23].

Figure 2.1: Example of continuous test results from a simulated population of healthy or 'disease free' and 'diseased' individuals



Note. Test result is shown as measurement value on the x-axis with a high measurement value corresponding to a more severe case and a binary test result determined by a threshold or cut-off shown as a vertical dashed line

2.4.3 Accounting for conditional dependence

LCM have been criticized for the conditional independence assumption however, methods exist that allow us to relax this assumption. While early LCM for diagnostic test accuracy often assumed independence [54], more recent work has accounted for conditional dependence [66]. This is important because, while we showed in section 2.3 that a new test's accuracy is underestimated when the reference test is assumed perfect and the two tests are conditionally independent, when the two tests are conditionally dependent the new tests accuracy is biased in the opposite direction. When the two tests are not conditionally independent and there is a positive correlation in the errors that the two tests make between either the truly diseased and/or the truly not diseased individuals, the new test's accuracy may be overestimated due to conditional dependence [67].

The conditional independence assumption can be relaxed in several ways, two common methods are by incorporating fixed effects [68] or random effects [69]. Both methods link the results from two (or more) tests made on the same individual. These approaches were initially designed in a frequentist framework however relaxing the conditional independence assumption using both fixed and random effects have now been explained in a Bayesian framework [67].

Fixed effects to account for conditional dependence between tests are

modelled by introducing covariance terms between pairs of tests conditional on disease status. They are described as fixed effects because the sensitivity and specificity remain constant between individuals in comparison to random effects where the test sensitivity and specificity can vary between individuals. We denote the covariance between two tests among those truly diseased as $\rho_p = Cov[T_1, T_2|Z = 1]$ and among those not diseased as $\rho_n = Cov[T_1, T_2|Z = 0]$, then it has been shown [23] we can rewrite 2.4.2 (and subsequently 2.4.3 and 2.4.4) as:

$$\begin{aligned}
\Pr(T_1 = 1, T_2 = 1) =& \pi(Se_1Se_2 + \rho_p) + (1 - \pi)((1 - Sp_1)(1 - Sp_2) + \rho_n) \\
\Pr(T_1 = 0, T_2 = 1) =& \pi((1 - Se_1)Se_2 - \rho_p) + (1 - \pi)(Sp_1(1 - Sp_2) - \rho_n) \\
\Pr(T_1 = 1, T_2 = 0) =& \pi(Se_1(1 - Se_2) - \rho_p) + (1 - \pi)((1 - Sp_1)Sp_2 - \rho_n) \\
\Pr(T_1 = 0, T_2 = 0) =& \pi((1 - Se_1)(1 - Se_2) + \rho_p) + (1 - \pi)(Sp_1Sp_2 + \rho_n)
\end{aligned} \tag{2.4.6}$$

The result of this, is that when ρ_p is greater than 0, the occurrence of a positive or negative result on one test increases the probability that the other test will return the same result [23].

However, fixed effects may not always best reflect the correlation (dependence) structure for example, if we expect we have a simultaneous dependence between more than two tests [70] or when the most plausible cause of correlations between test results is through observed or unobserved individual-level characteristics [71]. A further issue with fixed effects are the increasing number of parameters to estimate as the number of tests increases and when higher-order correlations are included.

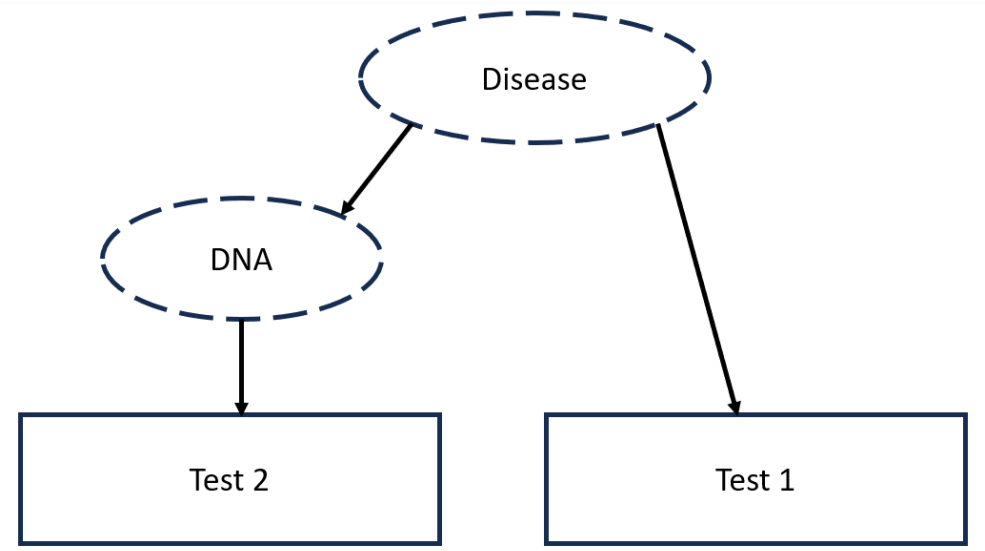
Under the random effects model, the sensitivity and specificity of tests are modelled as functions of a latent, subject-specific random variable (S_i). This variable represents some unobserved characteristic, for example infection intensity, that indirectly creates dependence between tests. Incorporating random effects means the probability that an individual has a positive result on a particular test is now conditional on disease status and the random variable infection intensity (i.e $\Pr(T_1 = t_{1i}|Z = z_i, S_i = s_i)$). The impact on estimates of sensitivity and specificity of relaxing the assumption of conditional independence or not and a full model specification for the random effect model in a Bayesian framework is presented in Chapter 3.

No matter which method is used to relax the conditional independence assumption, the number of parameters to be estimated in the model increases. In an application where identifiability is already a habitual challenge, this can present further considerations when deciding the structure of the latent class model to be used.

2.4.4 The assumption of two disease states

Up to this point I have only discussed a two-state latent class model; where individuals are either 'diseased' or 'not-diseased'. However multi-state latent class models may better reflect the latent subgroups within a population. For example, perhaps there exist three disease states: 'diseased-symptomatic', 'diseased-asymptomatic' and 'not-diseased'. Or, perhaps there are four states, with these three groups plus a 'pre-symptomatic' group. An alternative multi-state latent class model was introduced by Dendukuri *et al.* where multiple latent states reflected the scenario that different tests may measure different latent variables by design [56]. They hypothesized that of the four tests carried out in their study, two tests measured the latent variable 'diseased' or 'not diseased' but two other tests measured a different latent variable that was only a proxy for diseased or not diseased. A simplified version of this is shown in Figure 2.2.

Figure 2.2: Overview of a multi-state latent model adapted from [56]. Test 1 directly measured the latent variable of interest but test 2 indirectly measures the latent variable of interest through another latent variable (DNA)



Recent work [72], has shown that 2-state latent class models can lead to biased estimates of sensitivity and specificity if in fact more than 2-states exist. However, this analysis only considered the 2-state conditionally independent latent class model so further simulation work is needed to explore any potential bias that may exist from assuming a 2-state conditionally dependent latent class model when the truth is a 3-state conditionally dependent latent class model. Furthermore, a commentary on this work highlighted that while there are assumptions that have

to be made about conditional dependence as well as non-identifiability concerns, there remains no other satisfactory method for estimating diagnostic test accuracy in the presence of imperfect reference tests [21]. Until alternative methods are developed, latent class models that clearly report all assumptions made, remain the best method to estimate diagnostic test accuracy in the presence of imperfect reference tests.

Chapter 3

Estimating sensitivity and specificity of diagnostic tests using latent class models that account for conditional dependence between tests: a simulation study

3.1 Preamble

Chapter 2 introduced a central assumption of simple latent class models: the conditional independence assumption. I explained that this assumption is unrealistic in most scenarios and methods exist that allow this assumption to be relaxed. In both the fixed and random effect examples, I outlined how the assumption of conditional independence can be relaxed in both disease states. However, researchers can choose to account for conditional dependence in only one disease state or both and whether, in the case of more than two tests, all or just some tests expect to exhibit conditional dependence. This choice will depend on a) what makes most sense given the data/type of diagnostic tests and b) whether there is sufficient degrees of freedom or prior information to model the desired dependence structures.

The work presented in this chapter [22] explores the impact of not accounting for conditional dependence, in one or both disease states under different data generating mechanisms, on estimates of diagnostic test sensitivity and specificity through a simulation study. I also show the real world impact of this research by the re-analysis of data from a published

study.

Specifically, the aim of this chapter is to evaluate the impact of misspecifying the conditional dependence structure in latent class models on bias, coverage, and precision of estimates of diagnostic test sensitivity and specificity. To do this, I consider four data generating mechanisms with different conditional dependence structures that all stem from a scenario where five different binary diagnostic tests are carried out on 500 individuals with two possible latent states. Each of the 1000 simulated data sets for each data generating mechanism is then analysed using five different Bayesian latent class models.

In the following paper we estimate the bias in the median sensitivity and specificity of five diagnostic tests in scenarios that consider different assumptions regarding the conditional dependence of tests given the true disease status of an individual. The median was chosen so that we were estimating a parameter that was specified directly in the data generating mechanism. The mean could also have been estimated and this would incorporate the random effect however, this would have required some additional steps as the mean was not a parameter that was specified directly in the data generating mechanism.

3.2 Paper

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1702488	**Title**	Miss
First Name(s)	Suzanne		
Surname/Family Name	Keddie		
Thesis Title	Latent class models for diagnostic test accuracy with application to fever aetiology		
Primary Supervisor	John Bradley		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMC Medical Research Methodology		
When was the work published?	13/10/2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

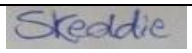
Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am the first author of the paper. I developed, coded, and carried out the statistical analysis for the simulation study and example application, created data visualizations and prepared the original draft of the manuscript. All authors conceived the study, contributed to the interpretations of the results and reviewed, edited and approved the final manuscript.

SECTION E

Student Signature	
Date	06/09/2023

Supervisor Signature	
Date	25/03/2024

Keddie *et al.*
BMC Medical Research Methodology (2023) 23:58
https://doi.org/10.1186/s12874-023-01873-0

BMC Medical Research Methodology

RESEARCH

Open Access

Estimating sensitivity and specificity of diagnostic tests using latent class models that account for conditional dependence between tests: a simulation study

Suzanne H. Keddie[1*], Oliver Baerenbold[1], Ruth H. Keogh[2] and John Bradley[1,3]

Abstract

Background Latent class models are increasingly used to estimate the sensitivity and specificity of diagnostic tests in the absence of a gold standard, and are commonly fitted using Bayesian methods. These models allow us to account for 'conditional dependence' between two or more diagnostic tests, meaning that the results from tests are correlated even after conditioning on the person's true disease status. The challenge is that it is not always clear to researchers whether conditional dependence exists between tests and whether it exists in all or just some latent classes. Despite the increasingly widespread use of latent class models to estimate diagnostic test accuracy, the impact of the conditional dependence structure chosen on the estimates of sensitivity and specificity remains poorly investigated.

Methods A simulation study and a reanalysis of a published case study are used to highlight the impact of the conditional dependence structure chosen on estimates of sensitivity and specificity. We describe and implement three latent class random-effect models with differing conditional dependence structures, as well as a conditional independence model and a model that assumes perfect test accuracy. We assess the bias and coverage of each model in estimating sensitivity and specificity across different data generating mechanisms.

Results The findings highlight that assuming conditional independence between tests within a latent class, where conditional dependence exists, results in biased estimates of sensitivity and specificity and poor coverage. The simulations also reiterate the substantial bias in estimates of sensitivity and specificity when incorrectly assuming a reference test is perfect. The motivating example of tests for Melioidosis highlights these biases in practice with important differences found in estimated test accuracy under different model choices.

Conclusions We have illustrated that misspecification of the conditional dependence structure leads to biased estimates of sensitivity and specificity when there is a correlation between tests. Due to the minimal loss in precision seen by using a more general model, we recommend accounting for conditional dependence even if researchers are unsure of its presence or it is only expected at minimal levels.

Keywords Diagnostic test accuracy, Latent class model, Bayesian inference, Conditional dependence

*Correspondence:
Suzanne H. Keddie
suzanne.keddie1@lshtm.ac.uk
[1]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.
[2]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK.
[3]MRC International Statistics and Epidemiology Group, London School of Hygiene and Tropical Medicine, London, UK.

Background

Diagnostic tests are used widely to discriminate between individuals with and without certain conditions and diseases. The results of these tests have important consequences, for both decision on treatment of individuals and for population health interventions. As a result, accurate characterization of diagnostic tests is paramount for optimal decision making. The usefulness of a diagnostic test is a combination of its accuracy, namely sensitivity and specificity, as well as practical considerations including cost, ease of use and speed of results. Because of this variety of factors to consider, which can involve difficult tradeoffs, new tests are continually being developed that aim to improve upon previous tests in any of these factors. To truly compare test effectiveness we must be able to assess the accuracy of a diagnostic test with minimal bias and high precision.

Standard methods estimate the sensitivity and specificity of a diagnostic test by comparing the results of a new test to the results of a 'gold standard' reference test. On the assumption that the reference test is indeed a 'gold-standard' or perfect test, with 100% sensitivity and 100% specificity, we can be certain of the true infection status of each individual tested and we can estimate the sensitivity and specificity of the new test directly. However, diagnostic tests are rarely perfect and in some instances there is no gold-standard test with which to compare. Examples of pathogens and infections where this is the case include Tuberculosis [1], Schistosomiasis [2] and Influenza [3]. In this situation, estimating the sensitivity and specificity of a diagnostic test is a difficult statistical problem and naively assuming the reference test is perfect will result in biased estimates of the new test's accuracy [4]. However, the accuracy of a given test can still be estimated by comparing the results of multiple imperfect tests applied to the same group of people. An increasingly popular method for making use of data on results from multiple imperfect diagnostic tests uses Bayesian latent class models (LCM) [5] and this approach has been applied across a wide range of pathogens [6–8].

To estimate diagnostic test accuracy with data from multiple imperfect tests using LCM requires making assumptions. Simple LCMs make the assumption that, conditional on the true infection status, results from multiple tests on an individual are independent. That is, the result of one test provides no information about the result of another test given the infection status of an individual. We refer to this situation as *conditional independence* throughout the rest of this paper. It has been highlighted by several researchers [9, 10] that the conditional independence assumption is unlikely to hold. For example, the assumption is unlikely to hold when there is a spectrum of disease severity. It is likely to be easier to detect disease in more severe cases for many pathogens and therefore, different tests on the same individual are more likely to return the same result. When disease severity, or some other factor associated with an individual, is associated with ease of detection, there remains a dependence between the tests even after conditioning for the true infection status of an individual. Tests that are based on the same underlying mechanism are also unlikely to be independent given the individual's disease status. When the assumption of conditional independence between tests is not valid, an analysis that assumes such independence is expected to result in biased estimates of sensitivity and specificity [11]. The assumption of conditional independence can be relaxed through incorporating either fixed [9] or random effects [12] into the LCM. The implementation of both in a Bayesian framework has been described elsewhere [11].

In a simulation study by Wang et al. [13] the authors showed that LCM with fixed effects, to account for conditional dependence among disease positive individuals, worked well both when tests were highly correlated (conditionally dependent) and when tests were truly conditionally independent. They also showed that the use of fixed effects or random effects has very little impact on the overall estimates of test accuracy. However, they only explored the possibility of conditional dependence in disease positive individuals as they assumed all diagnostic tests had a specificity of 99%. As a result, there could be no, or negligible, conditional dependence between these tests among disease negative individuals. However, the assumption of 99% specificity may not hold in many cases, so conditional dependence in non-infected individuals is also a possibility. When this is the case, researchers have a choice of conditional dependence structure in infected or non-infected individuals or both, and should be aware of the impact on estimates of sensitivity and specificity from choosing a particular structure, a situation highlighted in the case study by Menten et al. [14].

Much of the literature to date has focused on the importance of accounting for conditional dependence in disease positive individuals with much less discussion on the importance of conditional dependence among disease negative individuals. Above we discuss that disease severity or intensity may explain conditional dependence in disease positive individuals, In disease negative individuals, the presence of other parasites may work in a similar way, leading to a higher probability of false positive results on a range of tests, thus inducing a positive correlation among test results and a dependence between test results conditional on the true infection status [15].

There has been little research on the impact on estimates of sensitivity and specificity of choosing to account

for conditional dependence in disease positive or disease negative individuals only versus accounting for conditional dependence in both. Here we focus on LCM with random effects and carry out a simulation study investigating the size of bias and impact on precision of estimates of sensitivity and specificity and coverage of 95% credible intervals, when the conditional dependence structure is mis-specified. We also look at how the size of this bias changes depending on the level of dependence between tests. Finally, we extend the analysis from a motivating example that estimated the accuracy of five different diagnostic tests used in the identification of Melioidosis [16], to highlight the importance of the conditional dependence structure chosen in practice.

Methods

We begin this section with an overview of the latent class models used for diagnostic test accuracy before introducing our motivating example, followed by details of the simulation study utilizing the structured approach developed by Morris et al. [17].

Latent class models

We consider a sample of N individuals who all undergo R binary diagnostic tests. We have observed data $\boldsymbol{Y}=\{\boldsymbol{y_{ij}; i=1,\ldots,N, j=1,\ldots,R}\}$ where y_{ij} represents the test result (1 = positive, 0 = negative) of the jth test for the ith individual. We assume two disease classes, and we let d_i denote the true (but unobserved) infection status for individual i, with those who are truly infected having $d_i=1$ and those who are truly not infected $d_i=0$. The disease prevalence (i.e. the proportion for whom $d_i=1$) in the underlying population is denoted π.

For a given test, the probability that an individual who is truly infected will return a positive test result is defined as the sensitivity ($Se=\Pr(y=1|\,d=1)$) and the probability that an individual who is truly not infected will return a negative test result is defined as the specificity ($Sp=\Pr(y=0|\,d=0)$). Each test j has its own sensitivity and specificity, denoted Se_j and Sp_j. Under the assumption that the R diagnostic tests are conditionally independent, the likelihood of the observed data can be expressed as:

$$P(\boldsymbol{Y}\mid \pi, Se, Sp)=\prod_{i=1}^{N}\left(\left(\pi\prod_{j=1}^{R}Se_j^{y_{ij}}\left(1-Se_j\right)^{1-y_{ij}}\right)+\left((1-\pi)\prod_{j=1}^{R}Sp_j^{(1-y_{ij})}\left(1-Sp_j\right)^{y_{ij}}\right)\right) \quad (1)$$

To account for conditional dependence between tests in either or both truly infected individuals or truly not infected individuals, we allow the sensitivity and/or the specificity to vary by individual using a random effect. This reflects the situation where some subject-specific characteristic, besides the true disease status of the individual, affects the test result seen. The subject-specific value of the i^{th} individual in a disease class is denoted by s_{id}. Then, we can define the sensitivity of the j th test for the i th individual as $Se_{ij}=\Pr(y_{ij}=1\,|\,d_i=1,\ s_{i1})$ and similarly the specificity as $Sp_{ij}=\Pr(y_{ij}=0\,|\,d_i=0,\ s_{i0})$. The likelihood in [1] is then modified to include Se_{ij} and Sp_{ij} where we assume then that sensitivity takes the form:

$$Se_{ij}=\mathrm{g}^{-1}\left(\alpha_{\mathrm{j1}}+\beta_{\mathrm{j1}}s_{\mathrm{i1}}\right) \quad (2)$$

and, specificity:

$$Sp_{ij}=\mathrm{g}^{-1}\left(\alpha_{\mathrm{j0}}+\beta_{\mathrm{j0}}s_{\mathrm{i0}}\right) \quad (3)$$

where $g(\cdot)$ is a link function. In this study we use the inverse logit link, so $g^{-1}(x)=1/(1+e^{-x})$. α_{jd} and β_{jd} are unknown parameters to be estimated. β_{jd} describes the dependency of test j in disease class d on the random effects such that if all $\beta_{jd}=0$, there is no dependence on the random effect and all j tests among both disease classes are conditionally independent. We can estimate the mean or median sensitivity and specificity of a given test from the two parameters α_{jd} and β_{jd}. The random effect s_{id} is assumed to follow a standard normal distribution ($s_{id}\sim N(0,1)$) . For a more detailed description of random-effect latent class models see references [11, 12] and for details about how latent class models are implemented in this study see the model specification and implementation section below.

Motivating example

We illustrate the impact of different conditional dependence structures on estimates of sensitivity and specificity using data from a study that utilised LCM to estimate the sensitivity and specificity of five different diagnostic tests used in the diagnosis of Melioidosis [16]. Melioidosis is an infectious disease caused by the bacterium *Burkholderia pseudomallei*. The data are from a cohort of 320 febrile adult patients recruited over a 6 month period from a hospital in the northeast of Thailand in 2004 [18]. The five tests included four serological tests (indirect hemagglutination test (IHA), IgM immunochromogenic cassette test (ICT), IgG ICT, and ELISA) and culture test which was assumed 100% specific throughout all their analyses. For comparability we made the same assumption.

In the original analysis, Limmathurotsakul et al. implemented four different LCM with various conditional dependence structures as well as an analysis which assumed culture was a perfect gold standard. The LCM models varied from a model assuming conditional independence between all tests (Model 0) to those considering conditional dependence between a single pair of serological tests using fixed effects (Models 1 and 2) and finally those that use random effects to represent dependence between all serological tests within a disease class (Models 3 and 4) but they did not consider a model that simultaneously accounted for conditional dependence within both true positive and true negative individuals. See Table 1 for a summary of the models considered in the original paper. We extend their analysis to consider a 'Model 5' which allows dependence between all four serological tests among those individuals truly infected and those individuals truly not infected using random effects. Before reporting the results of this analysis we describe a simulation study used to explore the impact on estimates of sensitivity and specificity of using the wrong conditional dependence structure.

Simulation study

Aim

To evaluate the impact of mis-specifying the conditional dependence structure in latent class analysis on bias, coverage, and precision of estimates of sensitivity and specificity.

Data generating mechanism

Data are simulated on 500 individuals for five diagnostic tests. As in our motivating example, we imagine four tests ($j=2,3,4,5$), of a similar nature to the serological tests in the motivating example, that exhibit different conditional dependence structures among themselves, and one test ($j=1$), of a similar nature to a culture test, which is assumed independent of the serological tests. We consider four scenarios for the conditional dependence structure between serological tests described in Table 2. In all four conditional dependence scenarios, the underlying disease prevalence is 50% ($\pi=0.5$). All tests have a median sensitivity of 0.65 ($g^{-1}(\alpha_{1j})=0.65$) while the median specificity of the four serological type tests is 0.9 ($g^{-1}(\alpha_{0j})=0.9, j=2,..,5$) and the median specificity of our independent reference culture type test is 0.99 ($g^{-1}(\alpha_{01})=0.99$).

For the three scenarios in which there is conditional dependence, we set β_{jd} equal to 1. When the median sensitivity is 65%, the inclusion of this random effect means the interquartile range for sensitivity is 48–78% and with a median specificity of 90% the interquartile range is 82–94%. In a secondary simulation, we also compared this scenario with two additional scenarios under different values for β ($\beta=0.2,0.6$), where lower values of the

Table 1 Models and conditional dependence structures compared

Model	Dependence Structure	Effect Type Used	Included in this paper's simulation
Model 0	Conditional Independence between all tests	NA	Yes
Model 1	Dependence between IHA and IgM ICT in disease positive individuals	Fixed	No
Model 2	Dependence between IHA and IgG ICT in disease positive individuals	Fixed	No
Model 3	Dependence between all serological tests in disease positive individuals	Random	Yes
Model 4	Dependence between all serological tests in disease negative individuals	Random	Yes
MODEL 5	Dependence between all serological tests in disease positive and disease negative individuals	Random	Yes

Models 0–4 considered in Limmathurotsakul et al. [14]. Model 5 an extension not considered in the previous analyses. The last column highlights the scenarios that are considered in the simulation in this paper

Table 2 Data generating mechanisms considered

Data Generating Mechanism	Dependence in disease positive Individuals	Dependence in disease negative Individuals	Value of β_{jd} in models (2) and (3) for sensitivity and specificity
CIndep	No	No	$\beta_{jd}=0, d=0, 1$
CDP	Yes	No	$\beta_{j1}=1, \beta_{j0}=0$
CDN	No	Yes	$\beta_{j0}=1, \beta_{j1}=0$
CDPN	Yes	Yes	$\beta_{jd}=1, d=0, 1$

standard deviation correspond to a narrower interquartile range around the median sensitivity and specificity.

Estimand/target of the simulation

In each simulated data set we estimate the sensitivity and specificity of each diagnostic test ($j=1,..,5$) for the median individual (with random effect 0) and the associated 95% credible interval.

Methods

Each simulated dataset is analysed in the following five ways:

1) **GS Model:** A conditionally independent model where test 1 (culture) is assumed perfect, i.e. a gold standard model (GS) ($Se=Sp=1$)
2) **CIndep Model:** A conditionally independent (CIndep) model where test 1 (culture) is assumed imperfect (Eq. 1)
3) **CDP Model:** A model allowing conditional dependence in disease positive (CDP) individuals only (among serological tests, $j=2, \ldots, 5$) and all tests ($j=1,..,5$) are assumed imperfect (Eq. 2)
4) **CDN Model:** A model allowing conditional dependence in disease negative (CDN) individuals only (among serological tests, $j=2, \ldots, 5$) and all tests ($j=1,..,5$) are assumed imperfect (Eq. 3)
5) **CDPN Model:** A model allowing conditional dependence in both disease positive and disease negative (CDPN) individuals (among serological tests, $j=2, \ldots, 5$) and all tests ($j=1,..,5$) are assumed imperfect (Eqs. 2 and 3)

Performance measures

Under each scenario, we generated 1000 simulated data sets. We assess performance through bias in estimates of sensitivity and specificity (including the Monte Carlo standard errors), precision of those estimates measured by the empirical standard error, and the coverage of the 95% credible intervals. These measures are defined in Supplementary Table 1. Empirical diagnostics were recorded for all simulations to keep track of any simulations with inference validity concerns. Validity concerns occurred when either divergent transitions and/or the split $\hat{R}$ statistic values larger than 1.01 were recorded [19, 20]. Any simulations with validity concerns are removed from the presentation of results.

Model specification and implementation

All models are fitted using Bayesian methods and so prior distributions must be specified for all parameters. In all models, the prior distribution for prevalence is assumed uniform between 0 and 1. In models where culture is allowed to be imperfect (CIndep, CDP, CDN and CDPN Models) the sensitivity of all tests are assumed uniform between a lower limit of 1 minus the specificity $\text{Se}_{\text{jlower}}=1-\text{Sp}_j$, $j=1, \ldots, 5$, and 1. This ensures that the probability of a positive test is higher for somebody with disease than without. In these same models, the specificity of our independent test ($j=1$) is assumed to follow a $beta(10,1)$ prior distribution and the specificity of all other tests ($j=2,..,5$) is assumed to follow a $beta(5,1)$ prior distribution. Although in this simulation we are assuming test 1 is a culture test and therefore we could assume a much stronger prior distribution for specificity, for the purposes of a more generalizable simulation we have kept this relatively uninformative. Assuming a $beta(10,1)$ distribution corresponds to an assumption of 95% probability of the specificity being above 74% and a $beta(5,1)$ distribution corresponds to an assumption of 95% probability of the specificity being above 55%. Where we account for conditional dependence between tests using random effects, β_{jd} is assumed to follow a $gamma(1,1)$ prior distribution. In this paper we assume that β is the same between all serological type tests ($j=2,..,5$) but that that the culture type test ($j=1$) is independent, and for simplicity, we consider the case where $\beta_{jd}=\beta_d$, $j=2, \ldots, 5$. The effect of this is that a change in the random effect s_i will cause the sensitivity and of all serological type tests for the i^{th} individual to change by the same amount and similarly, the specificity of all serological type tests for the i^{th} individual to change by the same amount. We implement all models in R [21, 22] using stan [23] and all code can be found at: https://github.com/shk313/Evaluating-sensitivity-and-specificity-from-LCM-a-simulation-study.git.

Results

Simulation study

Bias

Throughout the presentation of the results tests 2–5 ($j=2,..,5$) are combined. Figure 1 shows the overall mean bias and associated 95% confidence interval which quantifies the uncertainty in the estimates of bias for median sensitivity and median specificity across all simulations (excluding those where either divergent transitions and/or the split $\hat{R}$ statistic values larger than 1.01 were recorded) for each model under each data generating mechanism. For all data generating mechanisms, use of the GS Model where test 1 (culture) is assumed perfect yields biased estimates. The sensitivity of test 1 (culture) is biased upwards because the test does not have a sensitivity of 100% as is assumed in the model and the specificity of the serological tests is underestimated by a minimum of 10% considering the upper limit of the 95%

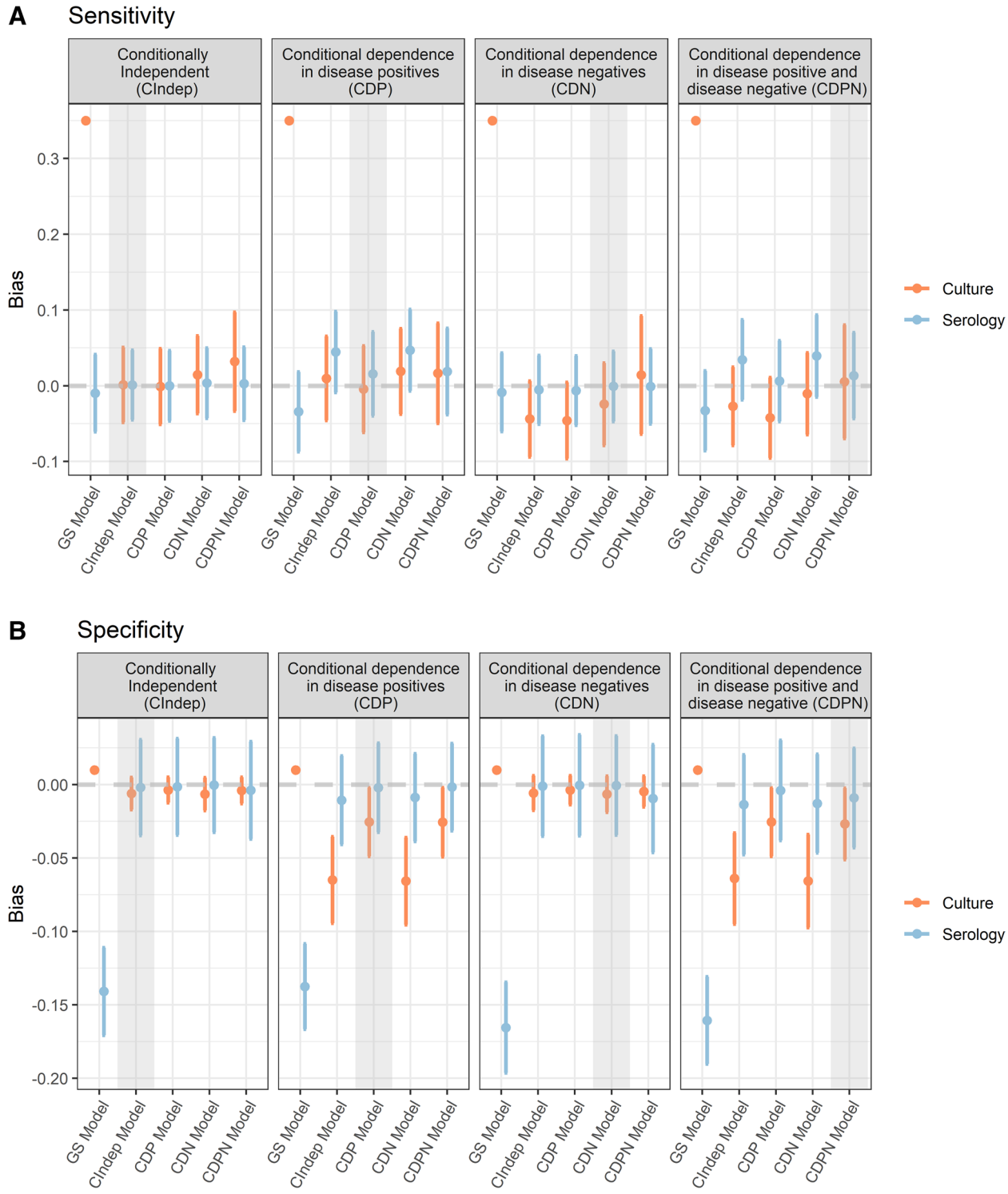


Fig. 1 Mean bias and 95% confidence interval in estimates of sensitivity and specificity under each scenario. Points show mean bias across all valid simulations and the bar extends to the lower and upper confidence intervals. Shaded areas highlight the model that corresponds to the data generating mechanism

confidence interval and a maximum of 20% using the lower limit of the 95% confidence interval across all data generating mechanisms. Under the CIndep data generating scenario all other models provide approximately unbiased estimates of sensitivity and specificity, with 0 being contained within all the 95% confidence intervals.

Considering the three data generating mechanisms where there is conditional dependence among serological tests within either or both disease positive and disease negative individuals, sensitivity estimates are approximately unbiased from all models with the exception of the GS Model. Each data generating mechanism and model combination for sensitivity estimates reported a Monte Carlo standard error less than 0.02 and 6 out of 40 scenarios had a Monte Carlo 95% confidence interval that contained zero while in the remaining scenarios the confidence interval did not contain 0, suggesting there is some small bias even after accounting for sampling variability. Bias in specificity estimates among data generating mechanisms with conditional dependence is minimised when the 'correct' model is used. This is most notable when there is conditional dependence between tests among positive individuals. For this data generating mechanism (CDP), when the model used assumes conditional independence between tests (CIndep Model) or conditional dependence between tests in negative individuals only (CDN Model), culture specificity is underestimated but serological tests specificity estimates are approximately unbiased. For specificity all scenarios had a Monte Carlo standard error less than 0.01 and, like with sensitivity, only six scenarios had a 95% Monte Carlo confidence interval for the estimate that contained zero.

Coverage

Tables 3 and 4 show the coverage probability, that is, the percentage of simulations where the 95% credible

Table 3 95% Coverage probabilities and 95% confidence intervals for sensitivity estimates across 1000 simulations

		Culture				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model[a]
Data generating mechanism	CIndep	**0(0–0.3)**	96.4(95.1–97.4)	97.4(96.3–98.3)	97.0(95.8–97.9)	96.8(95.6–97.8)
	CDP	**0(0–0.3)**	94.7(93.2–96.0)	94.8(93.3–96.0)	**93.5(91.8–94.9)**	96.6(95.3–97.6)
	CDN	**0(0–0.3)**	**81.0(78.5–83.3)**	**79.1(76.5–81.5)**	95.5(94.1–96.7)	99.2(98.5–99.6)
	CDPN	**0(0–0.3)**	**92.2(90.4–93.7)**	**84.2(81.8–86.4)**	96.7(95.5–97.7)	99.5(98.9–99.8)
		Serology				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model[a]
Data generating mechanism	CIndep	96.2(95.5–96.7)	97.0(96.4–97.4)	96.8(96.3–97.3)	96.6(96.0–97.1)	96.8(96.2–97.3)
	CDP	**89.1(88.1–90.0)**	**80.4(79.1–81.6)**	95.5(94.8–96.1)	**79.4(78.1–80.6)**	95.0(94.3–95.6)
	CDN	95.4(94.7–96.0)	96.5(95.9–97.1)	96.5(95.8–97.0)	96.9(96.3–97.4)	97.2(96.6–97.7)
	CDPN	**88.4(87.4–89.4)**	**87.5(86.5–88.5)**	97.4(96.8–97.8)	**85.2(84.1–86.3)**	97.0(96.4–97.5)

Values in bold show scenarios where the upper limit of the confidence interval is less than 95%. Confidence intervals for coverage calculated using Jeffreys prior

[a] Total number of simulations summarised is not equal to 1000 for the CDPN model due to a number of simulations with convergence problems

Table 4 95% Coverage probabilities and 95% confidence intervals for specificity estimates across 1000 simulations

		Culture				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model[a]
Data generating mechanism	CIndep	**0(0–0.3)**	99.8(99.4–100)	100(99.7–100)	99.8(99.4–100)	99.9(99.5–100)
	CDP	**0(0–0.3)**	**8.3(6.7–10.1)**	96.9(95.7–97.8)	**8.4(6.8–10.2)**	96.7(95.5–97.7)
	CDN	**0(0–0.3)**	99.0(98.2–99.5)	100(99.7–100)	98.7(97.9–99.3)	99.9(99.5–100)
	CDPN	**0(0–0.3)**	**12.5(10.6–14.7)**	97.8(96.7–98.6)	**12.1(10.2–14.2)**	97.3(96.2–98.2)
		Serology				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model[a]
Data generating mechanism	CIndep	**0(0–0.1)**	96.1(95.5–96.7)	96.0(95.4–96.6)	96.5(95.9–97.1)	97.5(96.9–97.9)
	CDP	**0(0–0.1)**	95.4(94.7–96.0)	97.8(97.3–98.2)	96.2(95.6–96.8)	98.4(98.0–98.7)
	CDN	**0(0–0.1)**	97.5(96.9–97.9)	97.6(97.1–98.0)	97.5(97.0–98.0)	97.6(97.1–98.0)
	CDPN	**0(0–0.1)**	94.7(93.9–95.3)	97.3(96.8–97.8)	95.3(94.6–95.9)	98.1(97.6–98.5)

Values in bold show scenarios where the upper limit of the confidence interval is less than 95%. Confidence intervals for coverage calculated using Jeffreys prior

[a] Total number of simulations summarised is not equal to 1000 for the CDPN model due to a number of simulations with convergence problems

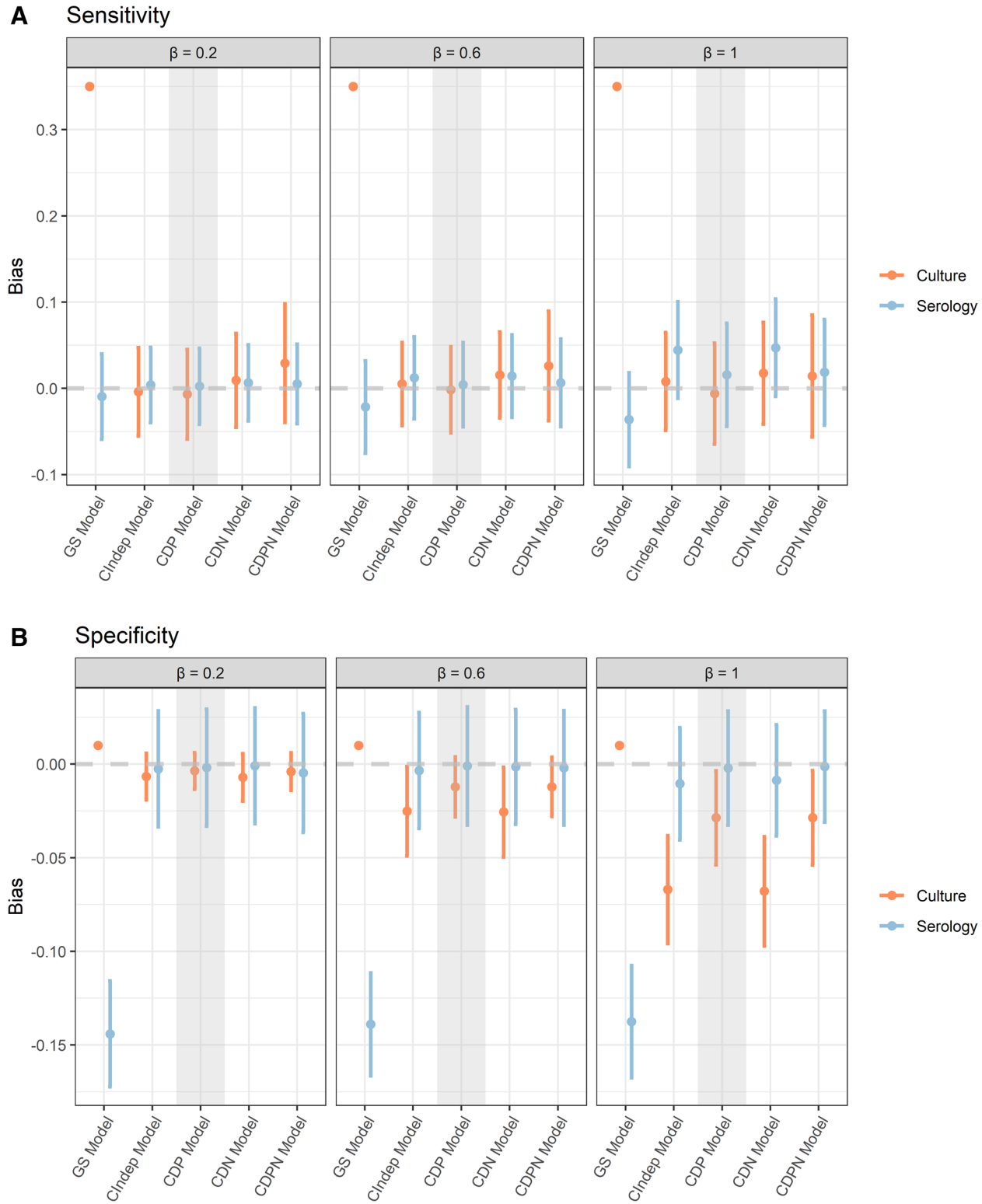


Fig. 2 Mean bias and 95% confidence interval in estimates of sensitivity and specificity with varying β 's. Points show mean bias across all valid simulations and the bar extends to the lower and upper confidence intervals. Shaded areas highlight the model that corresponds to the data generating mechanism

interval for the estimate (sensitivity or specificity) contains the true value. In the GS model where the sensitivity and specificity of culture is assumed perfect the 95% credible intervals for both estimates results in 0% coverage for the true sensitivity and specificity of culture and also 0% coverage for the specificity of serological tests across all data generating mechanisms. For sensitivity, all models except the GS Model show good coverage under the CIndep scenario. With conditional dependence among serology tests in disease positive individuals, models which do not account for this dependence have coverage around 80% for serology tests ($j=2,..,5$). On the other hand, when there exists conditional dependence between tests $j=2,...,5$ in disease negative individuals only, the coverage in those models that do not account for dependence remains close to 95% for serology tests but is below 85% for culture. The coverage of sensitivity for culture from the CDPN model is higher than the nominal 95% levels for both the CDN and CDPN data generating mechanisms with the upper limits of the confidence intervals approaching 100%.

Specificity estimates for all models, except the GS model, show good coverage for serology tests ($j=2,..,5$). For culture ($j=1$), coverage is higher than the nominal 95% level for all models except the GS model under CIndep and CDN data generating mechanisms. In the CDP and CDPN data generating mechanisms there is good coverage with models that account for the conditional dependence of tests in disease positive individuals (CDP and CDPN models) and poor coverage (<15%) with models that do not account for conditional dependence of tests in disease positive individuals (GS, CIndep and CDN models).

Precision

A complete table of precision estimates for each estimand within each scenario and for each model can be found in Supplementary Tables 2 and 3. Precision of estimates of sensitivity and specificity across all data generating mechanisms and models were similar for serological tests ($j=2,..,5$) but differed for our independent culture test ($j=1$). For culture, the empirical standard error of estimates for both estimands was larger using the most general model, CDPN model, similar across CIndep, CDP and CDN models, and 0 for the GS model which assumes the test was perfect. The loss of efficiency from using the most general model (CDPN model) was high for estimating the accuracy of culture but low for estimating the accuracy of serological tests. However, if we just consider using the CDP model (accounting for conditional dependence in disease positive individuals only) the loss of efficiency from using this model when the true data generating mechanism is CIndep was never more than 2% for either estimand and all tests.

Secondary simulation

All results so far considered the scenario where the standard deviation for the random effect is equal to one. We also considered, in a secondary simulation, the bias in estimates of sensitivity and specificity at two other levels of the standard deviation for the scenario where there exists conditional dependence in serological tests among infected individuals (CDP). These results are shown in

Table 5 Sensitivity and specificity (95% credible interval) estimated from each model for each diagnostic test

Model Name | Assumed dependence structure		Model 0 CIndep	Model 1 CDP among two tests	Model 2 CDP among two tests	Model 3 CDP	Model 4 CDN	Model 5 CDPN
Effect Type Used		**NA**	**Fixed**	**Fixed**	**Random**	**Random**	**Random**
Test	**Measure**						
Culture	Se[a]	61(53–69)	62(54–69)	62(54–69)	60(52–69)	74(59–97)	67(57–79)
	Sp[b]	100[c]	100[c]	100[c]	100[c]	100[c]	100[c]
Serology: IHA	Se[a]	73(66–80)	73(66–79)	73(67–78)	70(63–76)	72(65–79)	69(62–75)
	Sp[b]	87(79–93)	86(79–93)	86(79–92)	84(75–92)	75(61–88)	76(66–84)
Serology: Igm ICP	Se[a]	81(75–86)	80(74–85)	80(74–86)	77(71–83)	80(72–86)	76(69–82)
	Sp[b]	65(56–74)	64(55–74)	65(56–73)	62(53–72)	56(45–67)	55(46–64)
Serology: IgG ICT	Se[a]	91(86–95)	91(86–94)	90(86–94)	88(82–92)	89(84–94)	87(81–92)
	Sp[b]	76(67–85)	75(66–84)	75(66–84)	74(64–85)	62(48–77)	65(54–74)
Serology: ELISA	Se[a]	77(70–84)	78(70–84)	78(71–84)	75(68–78)	82(74–88)	80(72–86)
	Sp[b]	97(93–100)	98(94–100)	97(93–100)	97(92–100)	88(72–99)	95(82–100)

Values shown are mean estimates with 95% credible intervals

[a] Se = sensitivity, [b]Sp = specificity, [c]Specificity assumed perfect. Models 0–4 were considered in the original work of Limmathurotsakul et al. while Model 5 is the additional analysis considered in this paper

Table 6 Observed and predicted frequency of each response profile from each model

Response profile	Observed frequency	Expected frequency					
		Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
11111	69	49	53	53	63	49	65
11110	6	15	15	15	7	11	4
11101	0	5	6	1	2	6	2
11100	0	1	2	0	1	1	1
11011	9	12	8	13	6	13	7
11010	0	4	2	4	3	3	2
11001	0	1	1	0	1	2	1
11000	1	0	0	0	1	0	1
10111	14	18	14	14	11	19	12
10110	3	5	4	4	5	4	4
10101	0	2	1	6	1	3	1
10100	5	1	0	2	2	1	2
10011	3	4	9	3	4	5	6
10010	0	1	3	1	5	1	4
10001	3	0	1	2	1	1	2
10000	6	0	0	0	6	0	6
01111	35	31	33	33	42	31	35
01110	15	11	11	11	7	18	15
01101	0	3	4	1	1	3	2
01100	5	5	6	5	6	6	8
01011	5	8	5	8	4	4	4
01010	6	5	4	5	5	5	5
01001	0	1	1	0	1	1	1
01000	7	8	9	8	9	5	7
00111	5	12	9	9	8	9	7
00110	18	12	12	12	13	17	17
00101	0	2	2	5	2	2	1
00100	25	29	29	30	29	19	22
00011	7	3	6	3	3	3	3
00010	11	17	19	18	20	14	15
00001	2	1	2	2	2	2	2
00000	60	55	52	52	50	62	58

Observed frequency shown corresponds to five diagnostic test results from 320 patients with suspected melioidosis analysed in Limmathurotsakul et al. Models 0–4 were considered in the original analyses but model 5 is new to this paper

Fig. 2 and show that the size of bias increases as the value for the standard deviation increases when there exists conditional dependence but the model used assumes conditional independence among disease positive individuals. Among models where culture is assumed imperfect, this bias results in increasingly underestimated specificity estimates for culture.

Convergence

All simulations in GS, CIndep, CDP and CDN models passed our convergence checks and had a reported rank normalised split- $\hat{R}$ statistic as <1.01 and had no divergent transitions. The CDPN model reported convergence warnings in a number of simulations. 13% of simulations from the CDPN model under the CIndep data generating mechanism were removed along with 9, 3 and 10% in CDP, CDN and CDPN data generating mechanisms respectively. To run the CDPN and ensure there are no warnings, additional prior information may be required. In this analysis, all simulations with divergent transitions or a split- $\hat{R}$ statistic greater than 1.01 were removed from result summaries.

Motivating example

We re-analysed the data used by Limmathurotsakul et al. [16] and extended their work by considering a dependence structure not considered in the original paper. We fitted Models 0–5 as defined in Table 1 and as considered in the simulation study. Estimates of the sensitivity and specificity under each model are presented in Table 5. The point estimates and width of 95% credible intervals are similar across models 0–3 (model 0 being the model assuming conditional independence and models 1 and 2 being models that account for conditional dependence between two tests using fixed effects) however models 3, 4 and 5 (random effect models) do exhibit some important differences. Between models 4 and 5 the median sensitivity of culture and specificity of ELISA differs by 7% while between models 3 and 5 the median specificity of serological tests differs by 2–9%.

Table 6 shows the expected frequency of each possible response profile from the 5 tests under each model. Viewing the results in this way as opposed to looking at just estimates of sensitivity and specificity highlights a few key things. It highlights the importance of allowing conditional dependence as model 0 (assuming conditional independence) appears to fit the data least well, and also shows that the structure of the conditional dependence modelled affects the fit. We can see that models which only consider dependence between two of the four serology tests (models 1 and 2), underestimate the frequency of extreme response profiles (i.e. 0,0,0,0,0 and 1,1,1,1,1). Model 3 accounting for conditional dependence between all serological tests in those disease positive is able to capture those with all positive response profiles but unsurprisingly fails to capture those will all negative response profiles. On the other hand, Model 4 exhibits the same tendencies in reverse while our additional model accounting for conditional dependence in both disease positive and disease negative individuals (Model 5) is able to capture both extremes and appears to fit the data best. This is confirmed by comparing the models on the expected log predictive density [24] where Model 5 shows the best predictive performance closely followed by Models 3 and 4 (See Supplementary Table 4 for more details).

Discussion

We carried out a simulation study investigating the bias and coverage of sensitivity and specificity estimates arising from mis-specification of the conditional dependence structure in latent class models. We found that assuming conditional independence among tests within disease positive or disease negative individuals when conditional dependence exists leads to bias and poor coverage in estimates of test accuracy. Due to the minimal loss in precision seen by using a model which accounts for conditional dependence between serology type tests in disease positive individuals, our results suggest it makes sense to account for conditional dependence in positive individuals even if researchers are unsure of its presence or if it is only expected at minimal levels. And, if there is a suggestion that there is dependence in both disease positive and disease negative individuals we would recommend using the most general model, particularly if the specificity of diagnostic tests being investigated are less than perfect. The results from this simulation also reiterate findings from previous studies [5, 14, 16] that assuming conditional independence between imperfect tests is still much better than assuming an imperfect test is a gold-standard, even when the conditional independence assumption is not valid.

Our simulation study revealed that the size of bias in estimates of sensitivity and specificity was greatest when there existed conditional dependence among disease positive individuals and latent class models used assumed conditional independence among disease positive individuals. The size of this bias increased as the standard deviation of the random effects increased. Bias was larger when conditional dependence existed among disease positive individuals than conditional dependence in disease negative individuals. This reflects the fact that the true specificity was reasonably high in our simulation at 90% compared to a moderate sensitivity of 65%. In similar scenarios, where specificity is generally believed to be higher than sensitivity, these findings highlight that considering dependence among the disease positive individuals is most important to reduce the bias in accuracy estimates.

In the motivating example, accounting for conditional dependence in only disease positive or disease negative individuals may have resulted in biased estimates of the sensitivity and specificity of tests included in this analysis. Comparing the model that only considered dependence in positive individuals and the model that considered dependence in both positive and negative individuals, the median specificity of one test differed by 9 percentage points. Although dependence among disease negative individuals was thought to be negligible, examining the predicted frequencies for each profile highlighted shortfalls in the final selected model which assumed independence among tests in these individuals. This was confirmed with a relatively novel model comparison tool that addresses shortfalls of earlier estimates such as AIC and DIC [24]. This re-analysis highlights that examining predicted frequencies, when you have a truth to compare to, might be a useful addition in investigating

the most appropriate conditional dependence structure for a dataset.

There are limitations to this simulation study. Practically, we saw a limitation in implementing the CDPN model where some simulations exhibited divergent transitions and others a split $\hat{R}$ statistic greater than 1.01. In this case additional prior information may be necessary to ensure the model converges to the correct target distribution. Another limitation to this study is that we only considered a single prior distribution for the standard deviation of the random effect however estimates could be altered by a different choice of prior which has been investigated in a simulation study by Lee et al. [25]. We considered a single correlation among all serological tests in either disease positive or disease negative individuals. In practice you may have pairs or groups of tests that each require different random effect parameters with different standard deviations. However, if this is the case, this simulation still serves to highlight the potential biases that could be present in estimates of sensitivity and specificity if incorrect assumptions are made about the conditional dependence structure. Lastly, a key assumption of this simulation and our motivating example is that in the underlying population there exist only two disease classes; diseased and disease free. In some situations more than two classes may exist in a population, for example to distinguish between symptomatic and asymptomatic individuals. In cases where more than two diseases classes exist, recent work has shown that estimates of sensitivity and specificity from the two state LCM can be biased [26].

Conclusions

The impact of biased estimates of sensitivity and specificity is twofold. Firstly, a test whose accuracy is underestimated may not be used when it could be useful (more accurate, cheaper or easier to implement) and secondly, a test whose accuracy is overestimated may be used when more useful tests exist. Both outcomes ultimately result in negative consequences for individuals and societies, so minimizing the bias in our estimates of diagnostic test accuracy is paramount. This paper serves to highlight that not only should conditional dependence be taken account of but that the choice of conditional dependence structure is important and should be considered in any analysis of diagnostic test accuracy that utilizes latent class models.

Abbreviations

LCM	Latent Class Model
AIC	Akaike Information Criterion
DIC	Deviance Information Criterion

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-023-01873-0.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Acknowledgements

Not applicable.

Authors' contributions

All authors conceived the study, contributed to the interpretations of the results and reviewed, edited and approved the final manuscript. SK carried out the statistical analysis and prepared the original draft of the manuscript.

Funding

SK is supported by the Medical Research Council London Intercollegiate Doctoral Training Partnership Studentship (MR/N013638/1). The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Availability of data and materials

The dataset analysed during the current study are available from the supplementary material of the original article [16].
All code used in the current study to generate the simulated data sets and run each model are available from github (https://github.com/shk313/Evaluating-sensitivity-and-specificity-from-LCM-a-simulation-study.git).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 30 August 2022 Accepted: 20 February 2023
Published online: 10 March 2023

References

1. Pai M, Flores LL, Hubbard A, Riley LW, Colford JM. Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis. BMC Infect Dis. 2004;4(1):6.
2. Koukounari A, Webster JP, Donnelly CA, Bray BC, Naples J, Bosompem K, et al. Sensitivities and specificities of diagnostic tests and infection prevalence of Schistosoma haematobium estimated from data on adults in villages northwest of Accra, Ghana. Am J Trop Med Hyg. 2009;80(3):435–41.
3. Vemula SV, Zhao J, Liu J, Wang X, Biswas S, Hewlett I. Current approaches for diagnosis of influenza virus infections in humans. Viruses. 2016;8(4):96.
4. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol. 1988;41(9):923–37.
5. Joseph L, Gyorkos TW, 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(3):263–72.
6. Mweu MM, Murunga N, Otieno JW, Nokes DJ. Accuracy of diagnostic tests for respiratory syncytial virus infection within a paediatric hospital population in Kilifi County, Kenya. Wellcome Open Res. 2020;5:155.

7. Le HN, Sriplung H, Chongsuvivatwong V, Nguyen NV, Nguyen TH. The accuracy of tuberculous meningitis diagnostic tests using Bayesian latent class analysis. J Infect Dev Ctries. 2020;14(5):479–87.
8. Lindholz CG, Favero V, Verissimo CM, Candido RRF, de Souza RP, dos Santos RR, et al. Study of diagnostic accuracy of Helmintex, Kato-Katz, and POC-CCA methods for diagnosing intestinal schistosomiasis in Candeal, a low intensity transmission area in northeastern Brazil. PLoS Negl Trop Dis. 2018;12(3):e0006274.
9. Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests. Biometrics. 1985;41(4):959–68.
10. Torrance-Rynard VL, Walter SD. Effects of dependent errors in the assessment of diagnostic test performance. Stat Med. 1997;16(19):2157–75.
11. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57(1):158–67.
12. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics. 1996;52(3):797–810.
13. Wang Z, Dendukuri N, Zar HJ, Joseph L. Modeling conditional dependence among multiple diagnostic tests. Stat Med. 2017;36(30):4843–59.
14. Menten J, Boelaert M, Lesaffre E. Bayesian latent class models with conditionally dependent diagnostic tests: a case study. Stat Med. 2008;27(22):4469–88.
15. Dendukuri N. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Montreal: PhD thesis, McGill University; 1998.
16. Limmathurotsakul D, Jamsen K, Arayawichanont A, Simpson JA, White LJ, Lee SJ, et al. Defining the true sensitivity of culture for the diagnosis of Melioidosis using Bayesian latent class models. PLoS One. 2010;5(8):e12485.
17. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074–102.
18. Cheng AC, Peacock SJ, Limmathurotsakul D, Wongsuvan G, Chierakul W, Amornchai P, et al. Prospective evaluation of a rapid immunochromogenic cassette test for the diagnosis of melioidosis in Northeast Thailand. Trans R Soc Trop Med Hyg. 2006;100(1):64–7.
19. Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:170102434. 2017.
20. Vehtari A, Gelman A, Simpson DP, Carpenter B, Burkner P-C. Rank-normalization, folding, and localization: an improved R^ for assessing convergence of MCMC (with discussion). Bayesian Analysis. 2021;16(2):667–718.
21. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2022.
22. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
23. Gabry J, Češnovar R. cmdstanr: R Interface to 'CmdStan'. 2022. https://mc-stan.org/cmdstanr/,https://discourse.mc-stan.org.
24. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017;27(5):1413–32.
25. Lee J, Jung K, Park J. Detecting conditional dependence using flexible Bayesian latent class analysis. Front Psychol. 2020;11:1987.
26. Schofield MR, Maze MJ, Crump JA, Rubach MP, Galloway R, Sharples KJ. On the robustness of latent class models for diagnostic testing with no gold standard. Stat Med. 2021;40(22):4751–63.

Publisher's Note

Supplement

Supplementary Table 1. Performance measures: definitions, estimates and Montel Carlo standard errors where these are considered

Performance measure	Definition	Estimate	Monte Carlo standard error of estimate
Bias	$E[\hat{\theta}] - \theta$	$\frac{1}{n_i}\sum_{i=1}^{n_i}\widehat{\theta_i} - \theta$	$\sqrt{\frac{1}{n_{sim}(n_{sim}-1)}\sum_{i=1}^{n_{sim}}(\widehat{\theta_i} - \bar{\theta})^2}$
Empirical standard error	$\sqrt{Var(\hat{\theta})}$	$\sqrt{\frac{1}{n_i - 1}\sum_{i=1}^{n_i}(\widehat{\theta_i} - \theta)^2}$	
Coverage	$Pr(\widehat{\theta_{low}} \leq \theta \leq \widehat{\theta_{upp}})$	$\frac{1}{n_i}\sum_{i=1}^{n_i} 1(\widehat{\theta_{low,i}} \leq \theta \leq \widehat{\theta_{upp,i}})$	

θ represents the true value of an estimand, $\hat{\theta}$ the estimator, $\widehat{\theta_i}$ the estimate from the ith simulation, $\widehat{\theta_{low}}$ and $\widehat{\theta_{upp}}$ the estimate of the lower and upper 95% credible interval for the estimand respectively, $\bar{\theta}$ the mean of $\widehat{\theta_i}$ across simulations and n the number of simulations.n_{sim} is the number of simulations considered and $i = 1, \ldots, n_{sim}$, indexes a specific repetition of the simulations.

Supplementary Table 2: Empirical standard error of sensitivity estimates for each test type

		Serology ($j = 2,..,5$)				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model
Data generating mechanism	CIndep	0.037	0.033	0.033	0.033	0.035
	CDP	0.038	0.038	0.040	0.039	0.041
	CDN	0.037	0.033	0.033	0.033	0.036
	CDPN	0.038	0.038	0.039	0.039	0.041
		Culture ($j = 1$)				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model
Data generating mechanism	CIndep	0.000	0.036	0.036	0.037	0.155
	CDP	0.000	0.040	0.041	0.041	0.047
	CDN	0.000	0.036	0.036	0.039	0.057
	CDPN	0.000	0.037	0.038	0.039	0.056

Supplementary Table 3: Empirical standard error of specificity estimates for each test type

		Serology ($j = 2,..,5$)				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model
Data generating mechanism	CIndep	0.022	0.024	0.024	0.023	0.125
	CDP	0.021	0.022	0.022	0.022	0.120
	CDN	0.022	0.025	0.025	0.024	0.124
	CDPN	0.021	0.025	0.025	0.024	0.119
		Culture ($j = 1$)				
Model		GS Model	CIndep Model	CDP Model	CDN Model	CDPN Model
Data generating mechanism	CIndep	0.000	0.008	0.006	0.008	0.048
	CDP	0.000	0.021	0.017	0.021	0.152
	CDN	0.000	0.009	0.007	0.009	0.167
	CDPN	0.000	0.022	0.017	0.023	0.160

Supplementary Table 4: Motivating example model comparison on expected log predictive density

Model	ELPD[a] difference	SE[b] difference
Model 5	0.0	0.0
Model 3	-7.3	4.9
Model 4	-10.6	5.7
Model 2	-33.4	9.7
Model 1	-47.8	10.9
Model 0	-36670.0	19.1

[a] Expected log predictive density difference is the difference in Bayesian leave one out estimate of the expected log pointwise predictive density between two models (elpd_loo). Comparison is made between each model and the model with the largest expected log predictive density. [b] Standard error of component-wide differences in elpd_loo.

3.3 Additional details on model evaluation and selection

In the previous paper, I evaluated which model in the motivating example has the best predictive accuracy but do not provide many details about evaluation of predictive power. In this section, I introduce model evaluation and highlight the strengths of the approach used. When evaluating a model we are interested in measuring the performance of a particular model and to compare models. For LCM of diagnostic test accuracy there is no consensus on a recommended statistical criteria to evaluate models [73] but approaches that have been used before include Akaike Information Criterion (AIC) and Deviance Information Criterion (DIC) [74, 75].

In the previous paper, I evaluate how well the different models fit the data in two ways. The first, compared all possible response profiles of the considered diagnostic tests with those expected based on the model. An example of an observed response profile for an individual with three negative test results would be $(0, 0, 0)$. In the current analysis this was a very useful way of highlighting any discrepancies between the model and the observed data. The second method relied on a more traditional statistical approach using approximate leave-one-out cross-validation.

In this context of diagnostic test accuracy, the ideal measure of a model's fit would be its out-of-sample predictive performance for new data from the true data generating mechanism or, what is referred to as external validation [76]. However, out-of-sample data are often not available and instead we use methods that are approximations to external validation [77] of which, cross-validation is seen as the best alternative. Cross-validation is when the data are repeatedly partitioned into training and test data sets so that the model can be evaluated against the partition of the data set not used to fit the model. Leave-one-out cross validation (LOO) is the special case of cross-validation where each test data set represents a single data point. In this work, approximate leave-one-out cross-validation using Pareto Smoothed Importance Sampling to regularize the importance weights [78] is used to evaluate the predictive performance of models for the purpose of model comparison.

LOO-cross-validation uses a single data point to test the model's predictive power and the rest of the data to train the model, this process is then repeated by the number of data points available. For each data point, the measure of predictive accuracy is the leave-one-out predictive density given the data without the ith data point and uses the log score.

The overall estimate of out-of-sample predictive fit, referred to as the expected log predictive density (elpd) from LOO [79], is then:

$$elpd = \sum_{i=1}^{N} \log p(y_i|y_{-i}) \tag{3.3.1}$$

where, y_{-i} is the data minus the data point currently being evaluated and N is the total number of data points $i = 1, ..., N$. Exact cross-validation requires re-fitting the model with different training sets but with approximate LOO-cross-validation we can avoid this step using importance sampling. The posterior of the full model ($\sum_{i=1}^{N} \log p(y_i|y)$) provides the importance sampling proposal distribution but our target distribution are the posterior draws from the posterior conditional on all the data minus the observation that has been left out. The ratio between these is our importance ratio. The posterior of the full model is likely to have a smaller variance and thinner tails than the LOO distributions which induces instability when using raw importance sampling and why Pareto smoothed importance sampling was developed [79]. We can also estimate the standard error in the elpd, i.e. how much variation there is in the log scores.

$$SE = \sqrt{NVar(\log p(y_i|y_{-i}))} \tag{3.3.2}$$

To summarise, LOO using Pareto-smoothed importance sampling uses the entire posterior distribution to average the log predictive density to more completely take account of the uncertainty in our parameter estimates and its use is encouraged for Bayesian models [79].

A limitation of any cross-validation approach is that it can be computationally expensive. This explains why to date, few other diagnostic test accuracy studies have used this approach. Instead, previous diagnostic test accuracy studies have carried out model selection using Akaike Information Criterion (AIC) and Deviance Information Criterion (DIC) [74]. Both AIC and DIC were used by Limmathurotsakul *et al.* [75] in the published analysis explored in section 3.2. Both information criteria use a point estimate as the point at which to evaluate the log likelihood as a means to evaluate the model's predictive accuracy thereby ignoring the uncertainty in the parameter estimates. Each method requires a bias correction due to over-fitting from using the same data used to estimate the model [79] and with each criterion a smaller value indicates a better model fit. LOO-cross-validation is therefore advantageous over these two approaches as it does not rely on the same data used to fit the model in the evaluation and, a new R package called *loo* [80] has made implementing this method straightforward.

Comparing the model selection results from Limmathurotsakul *et al.* [75] that used AIC to select Model 3 as the best model, our re-analysis, that relied on psis-loo for model selection, also found that out of those models considered in Limmathurotsakul *et al.* [75] Model 3 was the best. However, our incorporation of an additional model (Model 5) with a different conditional dependence structure outperformed all other models using psis-loo for model selection.

3.4 Paper contributions

The presented paper describes the results of a simulation study investigating the impact of misspecification of the conditional dependence structure in latent class models on estimates of diagnostic test sensitivity and specificity. Bias and poor coverage were symptoms of model misspecification. Minimal loss in precision from using the most flexible model led to a recommendation that conditional dependence should be accounted for in both disease positive and disease negative states in the absence of further information to guide the choice.

Further recommendations from this research include: comparing where possible observed and modelled response profiles in tandem with more traditional statistical information criteria or cross-validation and further confirmation that assuming a reference test is perfect when it is not, leads to biased estimates of the tests' under evaluation sensitivity and specificity.

To the best of my knowledge, all diagnostic test accuracy meta-analyses published to date that present measures of predictive accuracy have been carried out using JAGS [81] or BUGS [82], these analyses have not taken advantage of tools like LOO using Pareto smoothed importance sampling. This paper is unique in its attempt to use novel statistical software aimed at improving measures of predictive accuracy.

The findings of this paper influenced the proceeding work presented in the following chapters of this thesis in two ways. Firstly, the importance of misspecification led to more in-depth discussions with topic experts to glean more information about whether or not conditional dependence between diagnostic tests might exist. Secondly, different conditional dependence structures were considered in all analyses as sensitivity analyses.

Chapter 4

Diagnostic test accuracy meta-analyses

4.1 Preamble

Chapters 2 and 3 have discussed latent class models for diagnostic test accuracy estimation in the single-study setting; when multiple tests are applied to the same set of individuals from the same population. However, when these studies compare only two or three tests, methods to estimate the accuracy of a given diagnostic test from such studies are limited by the degrees of freedom available (See section 2.4.1). One potential solution to this is meta-analyses, where data from several studies comparing only two tests can be utilized to estimate diagnostic test accuracy. Diagnostic test accuracy meta-analyses require that each study provide the cross-classification between the test of interest and another reference test. The studies need not be on the same individuals or have the same reference test but each study must include the same test of interest.

This chapter outlines how latent class models can be used to estimate diagnostic test accuracy in the meta-analysis setting. I provide a brief overview of the current state of diagnostic test accuracy meta-analyses research before specifying the Bayesian model applied in this thesis to estimate the accuracy of diagnostic tests used in FIEBRE. This chapter also discusses the presentation of meta-analysis results and the value of particular summary statistics as well as the implementation of this analysis method.

4.2 An introduction to diagnostic test accuracy meta-analyses

A meta-analysis can be defined as "statistical analysis that combines or integrates the results of several independent trials considered by the analyst to be combinable" [83]. Meta-analyses have the advantage of resolving conflicting results between studies, may be more generalizable and ultimately provide more reliable estimates of the outcome of interest. Meta-analyses as a tool for summarising diagnostic test accuracy data are not new with the first guidelines for meta-analyses evaluating diagnostic tests published in 1994 [84]. However, while there has been much development on the guidelines [85, 86] and methods [87, 88, 89] available for diagnostic test accuracy meta-analysis the majority of published diagnostic test accuracy meta-analyses still do not account for an imperfect reference test [90, 91]. A fundamental assumption that leads to biased estimated of test accuracy (See Section 2.3).

In this thesis, I focus on meta-analyses that use aggregate data as opposed to those based on individual participant data. Key to a successful meta-analysis is a high quality systematic review following a predefined protocol with carefully defined inclusion criteria and documenting the process, for example by using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [92]. The systematic review element of meta-analyses is not discussed in any more detail in this thesis, but detailed guidance can be found in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [86]. All meta-analyses presented in this thesis use data from a systematic review that followed PRISMA guidelines and Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) methodology [93].

As a requirement for the diagnostic test accuracy meta-analyses discussed here, each included study must provide the 2x2 classification of binary test results on a group of participants, between the test of interest and another 'reference' test. All included studies must use the same test of interest but different reference tests may be used. Currently, most diagnostic test meta-analyses aim to jointly estimate the sensitivity and specificity of a test of interest as summarising them separately fails to account for the trade-off between these two measures and leads to biased estimates [94]. The earliest method proposed for a diagnostic test accuracy meta-analysis that incorporates the trade-off between sensitivity and specificity was by Moses, Shapiro and Littenberg in 1993 [87]. Fundamental to their approach was that, with everything else being equal,

the 2x2 table arising from independent studies considering the same test of interest and reference test, different sensitivities and specifcities must be a result of different thresholds used to define a positive result [87]. This concept is applicable to all types of diagnostic tests. In the setting of a diagnostic test with a continuous outcome this is straightforward, a higher or lower cut-off value is used as the threshold to determine a positive result. In this section, I continue with the continuous outcome diagnostic test example but, more stringent or more lenient rules can be derived in almost all diagnostic test scenarios.

Moses, Shapiro and Littenberg proposed the Summary Receiver Operating Characteristic (SROC) curve to present joint estimates of sensitivity and specificity. Within a meta-analysis, the SROC curve represents the possible pairs of sensitivity and specificity values that can be achieved as the cut-off value for a positive test can vary across studies. Within this model, variation in sensitivity and specificity between studies was solely attributed to the cut-off value (or other threshold used to define a positive test). Advancements to this method have focused on more completely capturing the possible between- and within-study variability in sensitivity and specificity. Variability is common between studies for several reasons, including the chosen cut-off value, the study design, and the study population characteristics. This variation is most easily incorporated through the use of hierarchical models.

The two most common hierarchical methods for diagnostic test accuracy meta-analyses are the hierarchical summary receiver operating characteristic (HSROC) model [88], an extension of the binomial regression model [95] and, the bivariate model [96]. Both are currently recommended for use in diagnostic test accuracy meta-analyses in the Cochrane Handbook for Systematic reviews of diagnostic test accuracy [26]. The bivariate model has a different parameterisation to the HSROC model, but the models are mathematically equivalent when no covariates are included [97].

To account for within-study variability that arises due to dependence between the test of interest and reference test when the reference test is imperfect, Sadatsafavi *et al.* [98] extended the subject-specific random effect model developed by Qu *et al.* [69] for the diagnostic test accuracy single-study setting, to the diagnostic test accuracy meta-analysis setting. In 2012, Dendukuri *et al.* [89] implemented what they called the extended HSROC model with fixed effects to account for conditional dependence between the results of the test of interest and reference test within the same study conditional on the true disease status of individu-

als. They also discussed but did not implement the use of random effects. They highlighted how this model is extended to account for multiple different reference tests which would be expected within a meta-analysis. LCM, applied in the context of a meta-analysis, requires a minimum of four studies to estimate the model parameters without any informative priors [89].

Other important considerations relevant to this work, are the extension to polytomous diagnostic tests [99] and data with multiple cut-off values [100]. Throughout this thesis, I focus on binary diagnostic test results but these have often been derived from continuous or ordinal data and it is likely that we lose some information by discretizing this data into two categories: positive and negative. Furthermore, not only can test results be continuous, discrete or ordinal the tests can be implemented with different thresholds for a positive test [100]. Under the extended HSROC model presented in this chapter, neither of these considerations have been addressed.

4.3 Extended HSROC model

In this section I introduce the general specification for the extended HSROC model. This model and its formulation were first described in Dendukuri *et al.* [89]. This is then applied in Chapter 5 to estimate sensitivity and specificity of the microscopic agglutination test in detecting acute leptospirosis infection. Other studies where this model has been applied include estimating the accuracy of clinical illness in bovine respiratory disease by Timsit *et al.* [101], estimating the accuracy of real-time PCR for COVID-19 by Kostoulas *et al.* [102] and, estimating the accuracy of sputum gram stain for bacterial pathogens by Ogawa *et al.* [103].

4.3.1 Notation

Retaining the notation used in Chapter 2, I introduce the term *Index test* to refer to the test we seek to estimate the sensitivity and specificity of. Within a meta-analysis, each included study reports the results from an index test and one other reference test. T_{1ji}, denotes the test result on the index test for the ith individual ($i = 1, ..., N_j$) in the jth study ($j = 1, ..., J$) and T_{2ji} the reference test result for individual i in study j. Studies may have different or the same reference test. Let r_j denote a diagnostic test used in study j, taking possible values $1, 2, ..., R$ with 1_j always indicating the index test, and with $r_j (r = 2, .., R)$ indicating

the reference test in study j. In any given study only two tests are observed. Following this, the sensitivity and specificity of the tests used in each study are represented by Se_{r_j} and Sp_{r_j} respectively. The number of individuals in a study j is denoted n_j and it is assumed that results from both tests are available on all of these individuals.

4.3.2 Model specification

Under the extended HSROC model there are two levels of hierarchy corresponding to variation within and between studies, the first level captures within-study variability where the sensitivity and specificity of the Index test in study j is given by:

$$\begin{aligned} Se_{1_j} =& g^{-1}(-(\theta_j - \alpha_j/2)/\exp(\beta/2)) \\ Sp_{1_j} =& g^{-1}(\theta_j + \alpha_j/2)/\exp(-\beta/2) \end{aligned} \tag{4.3.1}$$

as in Dendukuri *et al.* [89], where g represents the logit link function but probit [89, 104, 105] or a probit approximation could alternatively be used [98, 99]. The parameter θ_j represents the threshold (or cut-off value) used to define a positive test result which models the dependence between the true positive fraction and false positive fraction in each study, α_j represents the diagnostic accuracy measuring the difference between true positives and false positives and β the scale parameter, allows differences in the variation of outcomes between disease positive and disease negative individuals (i.e. if different cut-off values were used in a study you would expect a different sensitivity and specificity at each of these cut-off values and would therefore want to allow asymmetry in the underlying receiver operating characteristic (ROC) curve to capture this). In the meta-analyses discussed in this thesis, each study uses only a single threshold so the shape of the underlying ROC curve in each study is the same and β is modelled as a fixed effect [106]. The impact of increasing values of θ_j is to shift the estimated specificity higher and the sensitivity lower while increasing values of α_j results in higher estimates for both sensitivity and specificity (See Figure 4.1).

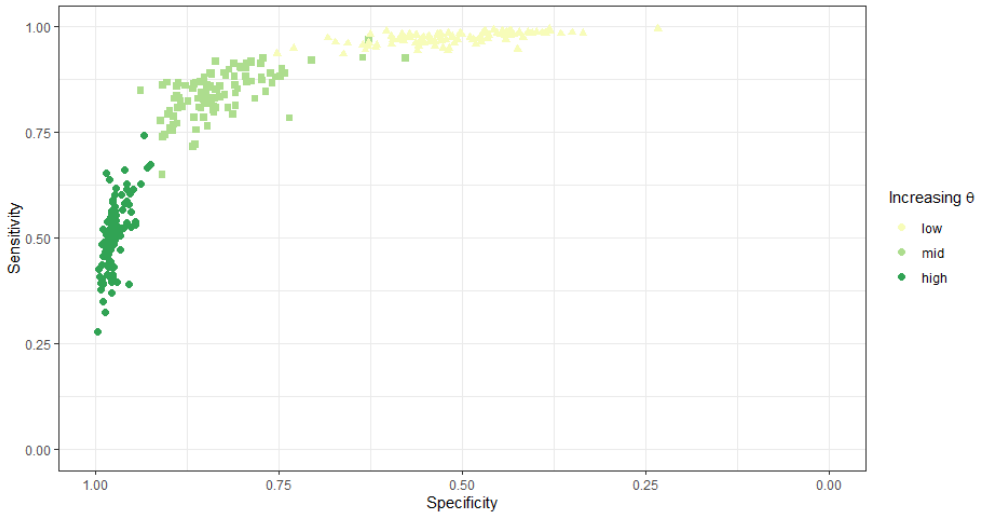
The second level accounts for between study variation, with both θ_j and α_j parameters modelled as random effects with independent normal distributions as in Dendukuri *et al.* [89] and Rutter and Gatsonis [88]:

$$\begin{aligned} \theta_j \sim& \text{Normal}(\Theta, \sigma_\theta) \\ \alpha_j \sim& \text{Normal}(\Lambda, \sigma_\alpha) \end{aligned} \tag{4.3.2}$$

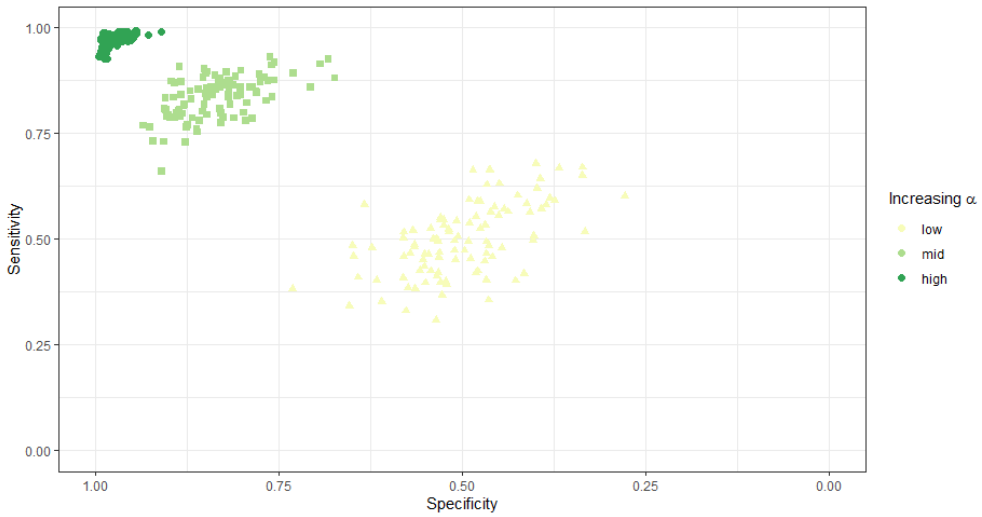
Lastly, we can consider a third level which contains the specification of

Figure 4.1: Example of how increasing values of θ and α can alter the estimated sensitivity and specificity using equation 4.3.1. A small simulation of 100 studies (shown as points) each with 1000 individuals and true disease prevalence drawn from a uniform distribution between 5 and 20%. Each simulation is repeated three times with increasing values of θ or α.

(a) Estimated sensitivity and specificity of an index test in 100 studies repeated for three different values of θ and holding all other parameters constant



(b) Estimated sensitivity and specificity of an index test in 100 studies repeated for three different values of α and holding all other parameters constant



the prior distributions (hyperpriors) for the mean parameters of θ_j and α_j (Θ and Λ) also known as hyperparameters.

Considerations about the dependence between test results from an individual given their true disease status are equally important in the meta-analysis setting. We can extend the above model to account for conditional dependence between the index test and reference test in a given study (j) in those individuals truly infected, by the addition of

a subject-specific random variable in each study S_{ij}, where S_{ij} is the random effect for person i in study j. Within study j, S_i are assumed to be normally distributed and I allow the standard deviation to differ by study (i.e. $S_{ij} \sim \text{Normal}(0, \tau_j)$). Now, the sensitivity of the index test on person i in study j is:

$$logit(Se_{1_{ji}}) = logit(Se_{1_j}) + S_{ij} \tag{4.3.3}$$

Similarly, the sensitivity of the reference test ($r = 2, .., R$) on person i in study j is:

$$logit(Se_{r_{ji}}) = logit(Se_{r_j}) + S_{ij} \tag{4.3.4}$$

We could also define $Sp_{r_{ji}}$ in a similar way with a separate random effect. However, in practice I did not allow a separate random effect for the specificity. The likelihood can be expressed similarly to the single-study setting, where t_{ji} is the test result for individual i in study j and π_j now denotes disease prevalence in the j^{th} study:

$$\begin{aligned} = \prod_{j=1}^{J}\prod_{i=1}^{N}(\pi_j \prod_{r=1}^{R}(Se_{r_{ji}}^{t_{ji}}(1 - Se_{r_{ji}})^{1-t_{ji}})I_{r_j=r}) + \\ ((1-\pi_j)\prod_{r=1}^{R}(Sp_{r_{ji}}^{(1-t_{ji})}(1 - Sp_{r_{ji}})^{t_{ji}})I_{r_j=r}) \end{aligned} \tag{4.3.5}$$

adapted from Dendukuri *et al.* [67] where, $I_{r_j=r}$ is an indicator of study j having the rth test.

To account for heterogeneity within this framework, for example by age of participants or location of study, we can introduce a study level covariate. We have to decide which model parameters (θ_j, α_j, β) we expect might differ by these covariates. For example, if we have an m-level covariate indicating a characteristic of interest (for example, adult versus child where $m = 2$), this is introduced as a m-level covariate, expressed via dummy variables for each level (m). And, if we believe that the diagnostic accuracy of the index test differs by levels of this covariate, then we replace the mean of the α_j parameter Λ, so that the mean is now determined by a linear function of the covariates:

$$\alpha_j \sim \text{Normal}(\lambda_m W_{mj} + \cdots \lambda_M W_{Mj}, \sigma_\alpha) \tag{4.3.6}$$

as in Rutter and Gatsonis [88], where λ_m represents the coefficient for each covariate level and W_{mj} is a dummy variable for each covariate level taking the value 1 if data from a given study correspond to that level of covariate and the value 0 otherwise. Here we have only introduced the

Table 4.1: First five rows of extracted data from the diagnostic test accuracy meta-analyses for the direct agglutination test (DAT)

ID	Author	Year	Region	DAT Type[1]	Comparison Test	I^+/C^{+2}	I^+/C^-	I^-/C^+	I^-/C^-
5	Abass, E.	2020	Northern Africa	LQ	Microscopy	141	17	14	180
8	Abass, E.	2015	Multiple	FD	Microscopy	100	21	13	97
10	Abass, E.	2007	Northern Africa	FD	Microscopy	10	7	0	31
11	Abass, E	2006	Northern Africa	LQ	ELISA	104	6	0	267
13	Abdalla, N	2011	Northern Africa	LQ	qPCR	0	3	32	297

[1] DAT type freeze-dried (FD) or liquid (LQ) antigen
[2] I^+/C^+ cell of the 2x2 table where both Index test (I) and Comparator (C) test are positive

covariate on the mean Λ, however, we could also introduce a covariate in the same way on θ_j or β parameters as well as the standard deviations; σ_θ and σ_α.

4.4 Presentation of results

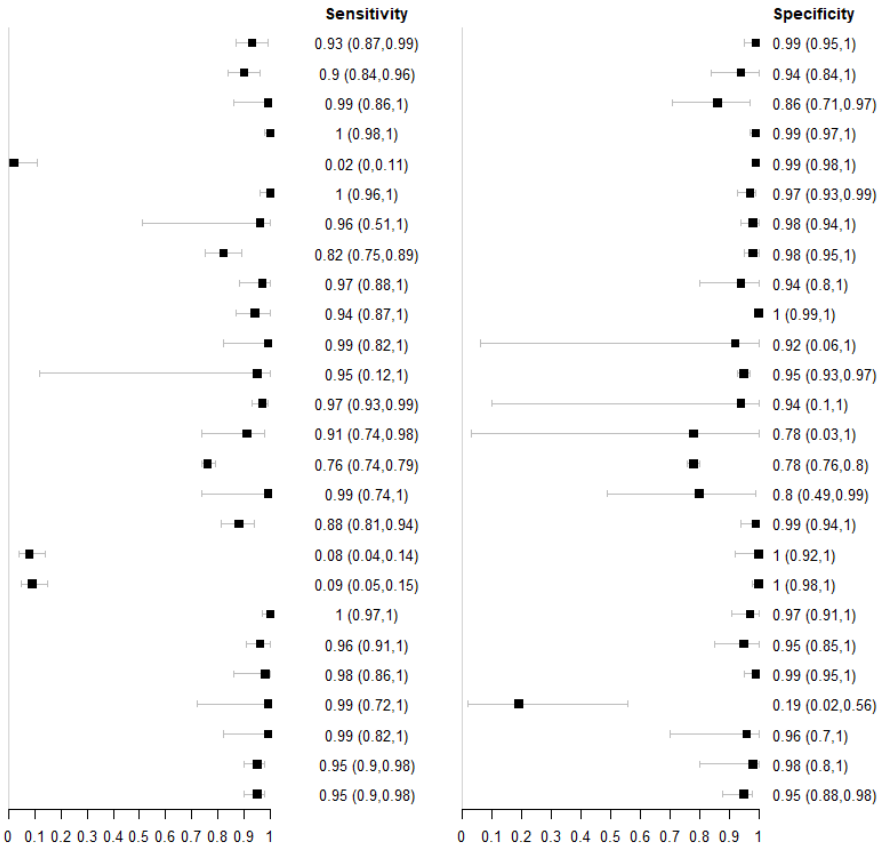
4.4.1 Primary results

The primary results of any meta-analysis include details on the search strategy as well as the key characteristics and methodological quality of the included studies [107] however, in this section I focus on the specific outputs of the extended HSROC model. For this purpose I will now introduce a motivating example taken from the review of the accuracy of the direct agglutination test for diagnosis of visceral leishmaniasis: [108] that will be further discussed in chapter 5. In this review, we sought to estimate the accuracy of the direct agglutination test (DAT) for the diagnosis of visceral leishmaniasis. From each of the 63 included studies a 2x2 table between DAT, the Index test (I) and a comparator test (C) was extracted as well as other variables including the region the study was carried out in and the type of DAT test used (freeze-dried or liquid antigen). Table 4.1 shows an excerpt of the extracted data. In this excerpt, while all studies included used the index test, DAT, three different comparator tests are used (Microscopy, Enzyme-linked immunosorbent assay (ELISA) and quantitative polymerase chain reaction (qPCR)), three studies used liquid antigen DAT and all but one study were from the Northern Africa region [109].

For each meta-analysis, I first present the individual study estimates of the Index test's sensitivity and specificity as coupled forest plots. For example, in the leishmaniasis example, the coupled forest plot from implementing the extended HSROC model and assuming conditional independence between index and comparator tests within a study given the true disease status of individuals is shown in Figure 4.2. For illustrative purposes, only the first 25 studies are shown. These plots highlight the overall trend and if there are any studies that appear to be outliers. In the example provided, three studies appear to have a much lower sensitivity than any others and one study has a much lower specificity. Next steps would be to review these papers and ensure they have been included appropriately.

Figure 4.2: Example of coupled forest plots showing individual study estimates of sensitivity and specificity for the index test as well as partially-pooled summary estimates (See equation 8.2.2) and the 2x2 table between Index (I) and reference (C) tests

Author	Year	Region	DAT Type	Comparator	I+/C+	I+/C-	I-/C+	I-/C-	Sensitivity	Specificity
Abass, E.	2020	Northern Africa	LQ	Microscopy	141	17	14	180	0.93 (0.87,0.99)	0.99 (0.95,1)
Abass, E.	2015	Multiple	FD	Microscopy	100	21	13	97	0.9 (0.84,0.96)	0.94 (0.84,1)
Abass, E.	2007	Northern Africa		Microscopy	10	7	0	31	0.99 (0.86,1)	0.86 (0.71,0.97)
Abass, E	2006	Northern Africa	LQ	ELISA	104	6	0	267	1 (0.98,1)	0.99 (0.97,1)
Abdalla, N	2011	Northern Africa	LQ	qPCR	0	3	32	297	0.02 (0,0.11)	0.99 (0.98,1)
Abdallah, K. A.	2004	Northern Africa	FD	KIT LQ-DAT	59	5	0	123	1 (0.96,1)	0.97 (0.93,0.99)
Abera, A.	2016	Eastern Africa	LQ	Leishmania skin test	3	14	2	255	0.96 (0.51,1)	0.98 (0.94,1)
Akhoundi, B.	2010	Southern Asia	LQ	FAST-DAT	105	5	25	193	0.82 (0.75,0.89)	0.98 (0.95,1)
Akhoundi, B.	2013	Southern Asia	LQ	A2-LAT	40	3	2	28	0.97 (0.88,1)	0.94 (0.8,1)
Al-Nahhas, S. A.	2008	Western Asia	FD	FAST-DAT	93	0	8	166	0.94 (0.87,1)	1 (0.99,1)
Asfaram, S.	2017	Southern Asia		ELISA	7	0	0	0	0.99 (0.82,1)	0.92 (0.06,1)
Ashkanifar, S.	2016	Southern Asia	LQ	ELISA	0	23	0	419	0.95 (0.12,1)	0.95 (0.93,0.97)
Ayelign, B.	2020	Eastern Africa	FD	AQ-DAT	107	0	3	0	0.97 (0.93,0.99)	0.94 (0.1,1)
Azazy, A. A.	2003	Western Asia	FD	ELISA IgG	14	2	2	0	0.91 (0.74,0.98)	0.78 (0.03,1)
Babiker, Z. O.	2007	Northern Africa	FD	Microscopy	2310	1241	782	3547	0.76 (0.74,0.79)	0.78 (0.76,0.8)
Barbosa Junior, W. L.	2015	South America	FD	ELISA	4	3	0	8	0.99 (0.74,1)	0.8 (0.49,0.99)
Basiye, F. L.	2010	Eastern Africa		PCR	81	3	11	87	0.88 (0.81,0.94)	0.99 (0.94,1)
Bejano, Shibabaw	2021	Eastern Africa	LQ	Leishmania skin test	7	8	60	178	0.08 (0.04,0.14)	1 (0.92,1)
Bekele, F.	2018	Eastern Africa	FD	Leishmania skin test	8	19	97	1096	0.09 (0.05,0.15)	1 (0.98,1)
Bern, C.	2000	Southern Asia	LQ	ELISA	85	11	0	139	1 (0.97,1)	0.97 (0.91,1)
Boelaert, M.	2004	Southern Asia		Microscopy	148	28	9	125	0.96 (0.91,1)	0.95 (0.85,1)
Canavate, C.	2011	Eastern Africa	FD	IFAT	32	1	3	66	0.98 (0.86,1)	0.99 (0.95,1)
Cavalcanti, A. T.	2012	South America		IFAT	3	6	0	1	0.99 (0.72,1)	0.19 (0.02,0.56)
Chakravarty, J.	2019	Southern Asia		ELISA	8	1	0	8	0.99 (0.82,1)	0.96 (0.7,1)
Chappuis, F.	2003	Southern Asia	LQ	Microscopy	130	15	8	30	0.95 (0.9,0.98)	0.98 (0.8,1)
Pooled									0.95 (0.9,0.98)	0.95 (0.88,0.98)



Pooled estimates are the partially-pooled estimate from the full model of 63 studies. DAT type is either liquid (LQ) or freeze-dried (FD). ELISA: Enzyme-Linked Immunosorbent Assay. IFAT: ImmunoFluorescence Antibody Test. Blank cells in the table represent missing data. Sensitivity and Specificity are presented as the median and 95% credible interval

One of the main reasons to perform a meta-analysis is to combine multiple results estimating the same estimand to retrieve a single summary measure that helps resolve conflicting results and serves to give a more precise estimate of the truth than a single study can provide. Throughout this thesis I use random effects models which rely on the partial pooling approach. The estimates are then referred to as partially pooled because, by the nature of the hierarchical design, the means are assumed to follow a common distribution despite each study having their own estimate. For example, Θ versus θ_j. The overall 'partially pooled' [110] sensitivity and specificity of the index test is presented alongside the individual study estimates as seen in Figure 4.2 and can be summarized from the extended HSROC model as:

$$\begin{aligned} Se_{pooled} &= g^{-1}(-(\Theta - \Lambda/2)/exp(\beta/2)) \\ Sp_{pooled} &= g^{-1}(\Theta + \Lambda/2)/exp(-\beta/2) \end{aligned} \tag{4.4.1}$$

where Θ and Λ are the means of the hierarchical priors for θ_j and α_j (assuming a normal distribution). Behind these summary statistics though, the results of individual studies may be heterogeneous or in fact may be very similar and this affects the interpretation of the summary measure and whether it can be generalized or not [111]. In the example presented in Figure 4.2 the pooled sensitivity and specificity are high 95% (95% CrI 90-98%) and 95% (95% CrI 88-98%) respectively however, by looking at the individual study estimates there is heterogeneity. Indeed, in meta-analyses of diagnostic test accuracy in particular, heterogeneity is the norm rather than the expectation [112]. As a result of heterogeneity it is recommended to present some assessment of heterogeneity alongside overall summary estimates.

4.4.2 Exploring heterogeneity

An early measure that tested for statistical heterogeneity in meta-analyses was called Cochran's Q test [113]. The test examined the null hypothesis that all studies are evaluating the same effect [113] however, it was noted that this test was susceptible to the number of studies included in the meta-analysis [112]. The I^2 statistic was developed to circumvent this issue and instead of measuring the degree of homogeneity, measures the degree of heterogeneity [114]. The I^2 statistic is quantified as:

$$I^2 = \frac{Q - df}{Q} \text{x} 100\% \tag{4.4.2}$$

where, Q is the chi-squared statistic and df is the degrees of freedom [112]. However, it has been shown that I^2 also exhibits bias when the number of studies is small [115]. In fact, the most recent Cochrane handbook for systematic reviews of diagnostic test accuracy recommends not using either of these univariate tests [106]. Instead, Macaskill *et al.* [106] recommend graphical displays such as summary ROC curves or prediction regions. Throughout this thesis all meta-analyses are presented with prediction intervals and displayed as prediction regions. The prediction region is defined as the region within which we have 95% confidence that the true sensitivity and specificity of any future study should lie [97] and has been recommended in recent work [116, 117].

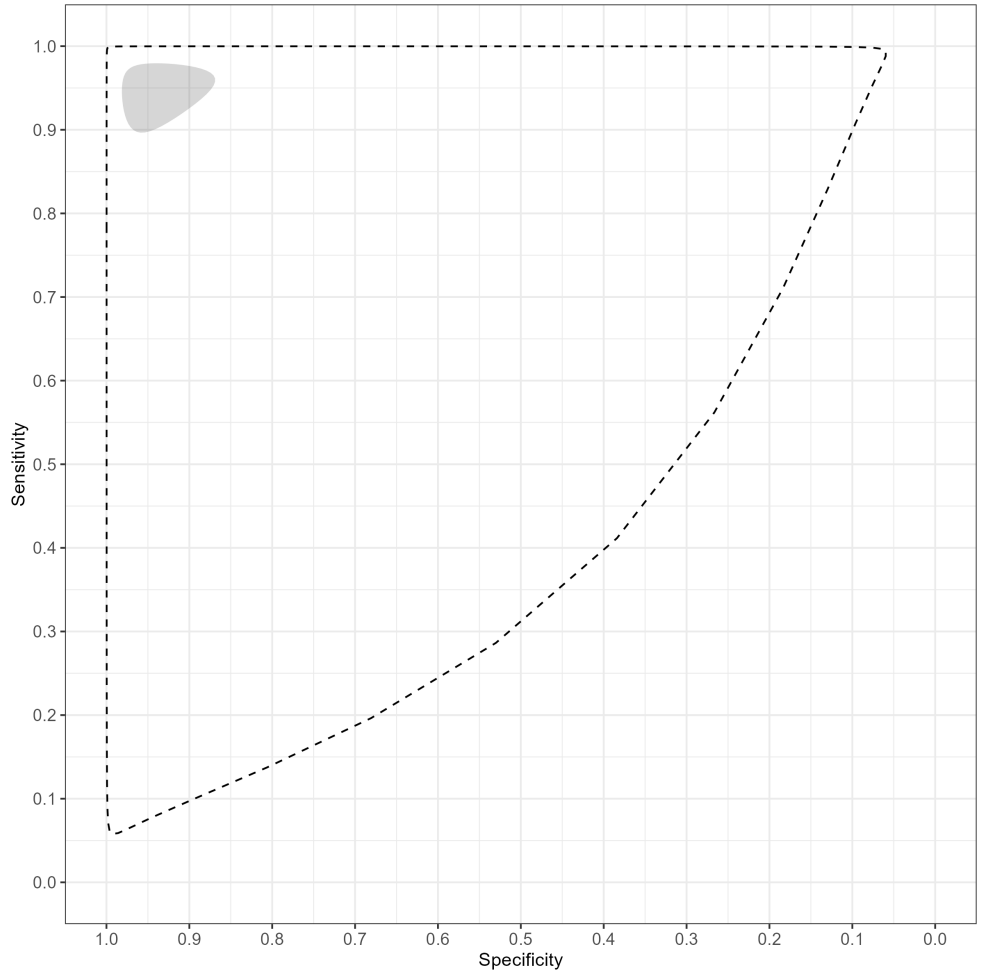
From the extended HSROC model we can estimate these prediction intervals, used to generate the prediction regions, for a future study ($j\prime$), where we assume that $\theta_{j\prime}$ and $\alpha_{j\prime}$ are drawn from a normal distribution with mean Θ and Λ respectively, as in equation 4.3.2:

$$\begin{aligned} Se_{1_{j\prime}} &= g^{-1}(-(\theta_{j\prime} - \alpha_{j\prime}/2)/exp(\beta/2)) \\ Sp_{1_{j\prime}} &= g^{-1}(\theta_{j\prime} + \alpha_{j\prime}/2)/exp(-\beta/2) \end{aligned} \tag{4.4.3}$$

In the meta-analysis of DAT for leishmaniasis the credible region (generated from the partially-pooled summary estimate) and prediction region for the model including all 63 studies is shown in Figure 4.3. The credible region, summarising the partially-pooled estimate is very small but the prediction region is very wide encompassing sensitivity and specificity of DAT in the full range of possible values. This shows that there is heterogeneity between the studies included in the current meta-analysis model.

In the motivating example, other study level characteristics were extracted from publications including the geographic region the study was carried out in and the DAT type. The DAT test is thought to have a different accuracy in different regions and so a covariate was introduced to allow additional heterogeneity in DAT accuracy by geographical region. The prediction regions presented in Figure 4.4 come from a model that includes a four level covariate, capturing the four unique geographical regions, on the diagnostic accuracy parameter (α_j). The prediction regions in this plot are smaller than the region seen in Figure 4.3 and highlight that while there is still heterogeneity in DAT test accuracy between geographical regions, the different geographical regions did explain some of the heterogeneity in the full model.

Figure 4.3: Example plot showing partially pooled 95% region (shaded) and 95% prediction region (dashed) from a meta-analysis for DAT used to diagnose leishmaniasis



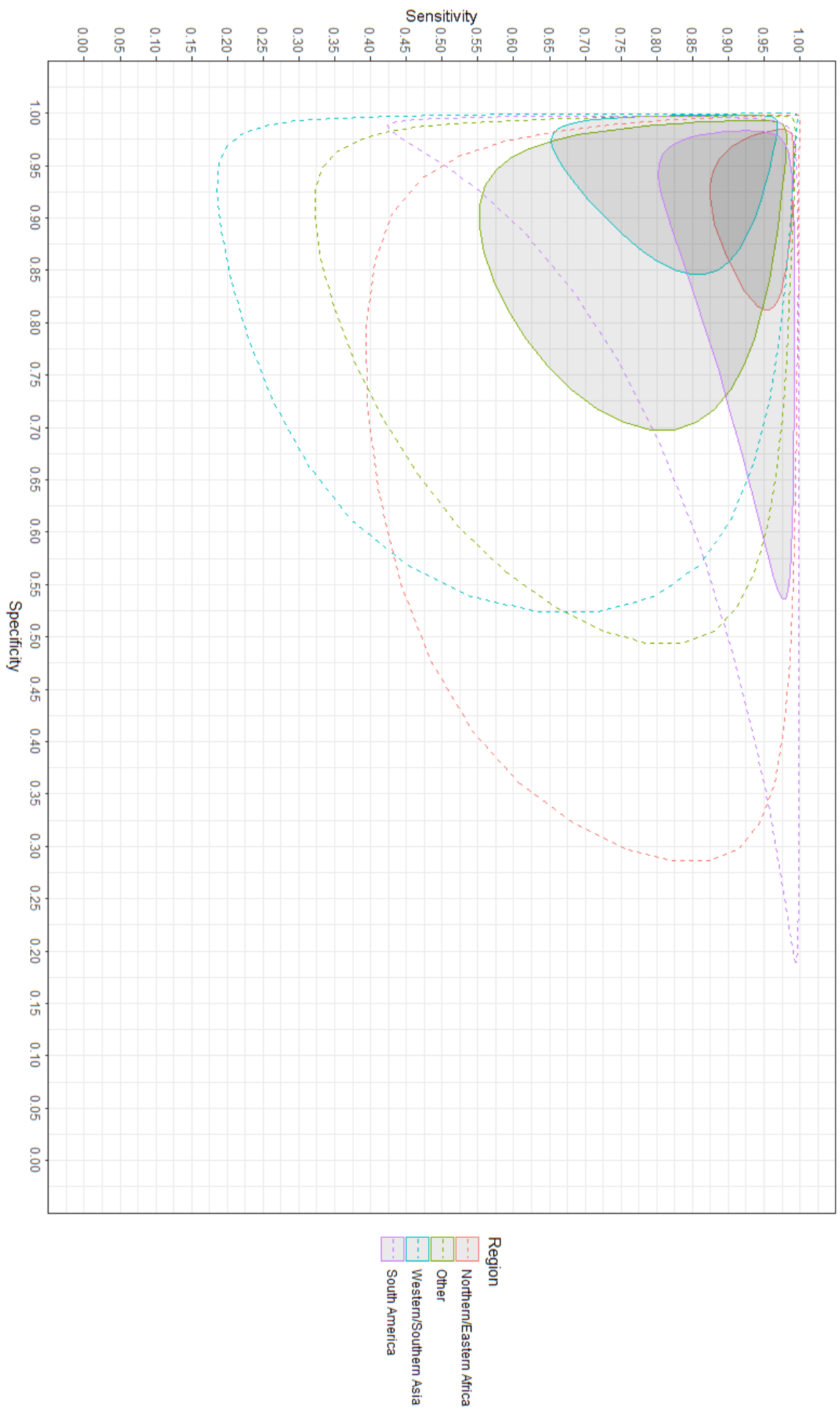


Figure 4.4: Example plot showing partially pooled 95% intervals (shaded) and 95% prediction intervals (dashed) from a meta-analysis that examined variability in estimates by geographical region. Partially-pooled and prediction regions are shown for each geographical region

4.5 Model Implementation and model checking

Meta-analysis models discussed here can be implemented in various software including JAGS, BUGS and Stan. Online tools have also been developed that allow users to carry out meta-analyses without having to write their own code [118].

As a Bayesian analysis, all parameters require prior distributions to be specified. I allow the prevalence in each study (π_j) to be uniform over 0 to 1. This was implemented as a Beta distribution with parameters α and β equal to 1 as in Dendukuri *et al.* [89]. Variance parameters (σ_α and σ_θ) follow zero-truncated standard normal distributions. This is similar but different to Dendukuri *et al.* [89] where they were assumed to follow a uniform distribution (Uniform(0,2)). The priors for Θ, Λ and β are chosen to ensure a uniform distribution over 0 to 1 for the partially pooled summary estimates as is advised in Dendukuri *et al.* [89]. In practice, this was achieved by carrying out prior predictive checks to investigate the prior model starting with the values used in Dendukuri *et al.* [89] and tuning these until an approximately uniform distribution on the pooled summary estimates was achieved. The standard deviation of the random effect τ_j was allowed to vary by study and was assumed to follow a gamma (1,1) distribution.

Published literature is used to help inform the prior distributions for each comparator test and I assume that for each comparator test, the probability of a positive test is greater in a disease positive individual than a disease negative individual (i.e. I assume that $Se_{2j} > 1 - Sp_{2j}$). The addition of this constraint helps with label-switching which is common to diagnostic testing models that use a latent class framework [119]. Label-switching replaces estimation of π, Se and Sp with $1 - \pi$, $1 - Se$ and $1 - Sp$ [120]. While this identifiability constraint can solve the label switching problem, in general it is not guaranteed [121].

All meta-analysis models described in this thesis are implemented using CmdStan in R [122]. Stan uses a No-U-Turn sampler (NUTS) to obtain estimates of the marginal posterior distribution for each parameter via Hamiltonian Monte Carlo (HMC). In comparison to the Gibbs sampler and Metropolis algorithm that sample the target distribution by random walk behaviour, HMC is more efficient and generally requires much fewer iterations to converge [76]. All models were run using 4 chains [123] and convergence was assessed visually with the *Bayesplot* R package [63]. I only report results where the rank normalised split-R-Hat

statistic is < 1.01, the total effective sample size is at least 400 and there are no warnings for energy fraction of missing information or divergent transitions [123]. The traditional R-hat statistic of Gelman and Rubin [124] which compares the variation between chains to the variation within chains, was demonstrated to fail to correctly diagnose non-convergence when the variance varies across chains, instead, the rank-normalized split-R-hat statistic is advised and used as the default in Stan [123]. In line with this, it is advised to use rank plots from multiple chains instead of trace plots for identifying convergence. These plots are histograms of the ranked posterior draws for each chain, if the chains are well mixed and are targeting the same posterior distribution we expect the ranks in each to be uniform. We add this diagnostic to the more traditional trace plots in our assessment of convergence. Figure 4.5 shows the trace and rank plots for the θ_j and α_j parameters of the first three studies ($j = 1, .., 3$) in the motivating leishmaniasis example. Both the trace plots, which show a good mixing of all chains over the course of sampling, and the rank plots, which appear in a uniform distribution, highlight the model has converged on a target distribution.

Bayesian hierarchical models often involve complex geometries which can be hard for the sampler to explore. To speed up sampling, I implement a non-centered parameterization [125]. Under this non-centered parameterization, equations 4.3.2 (i.e. $\theta_j \sim Normal(\Theta, \sigma_\theta)$) are implemented as:

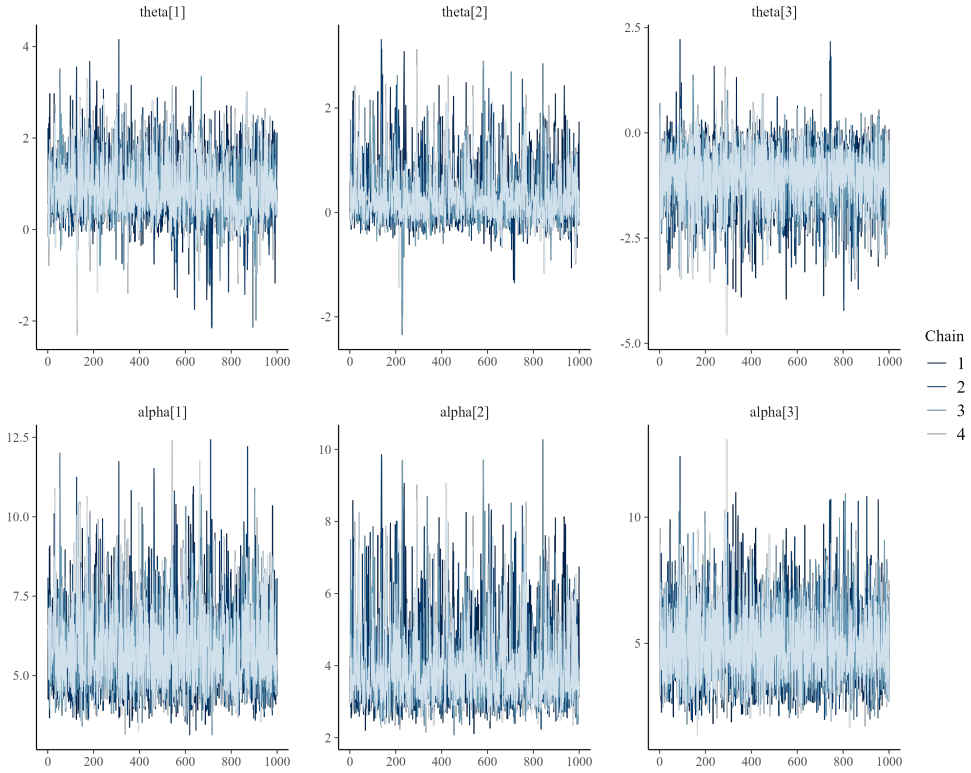
$$\theta_j = \Theta + \sigma_\theta \theta_{raw} \tag{4.5.1}$$

where, θ_{raw} is of length J and are sampled as independent standard normal distributions.

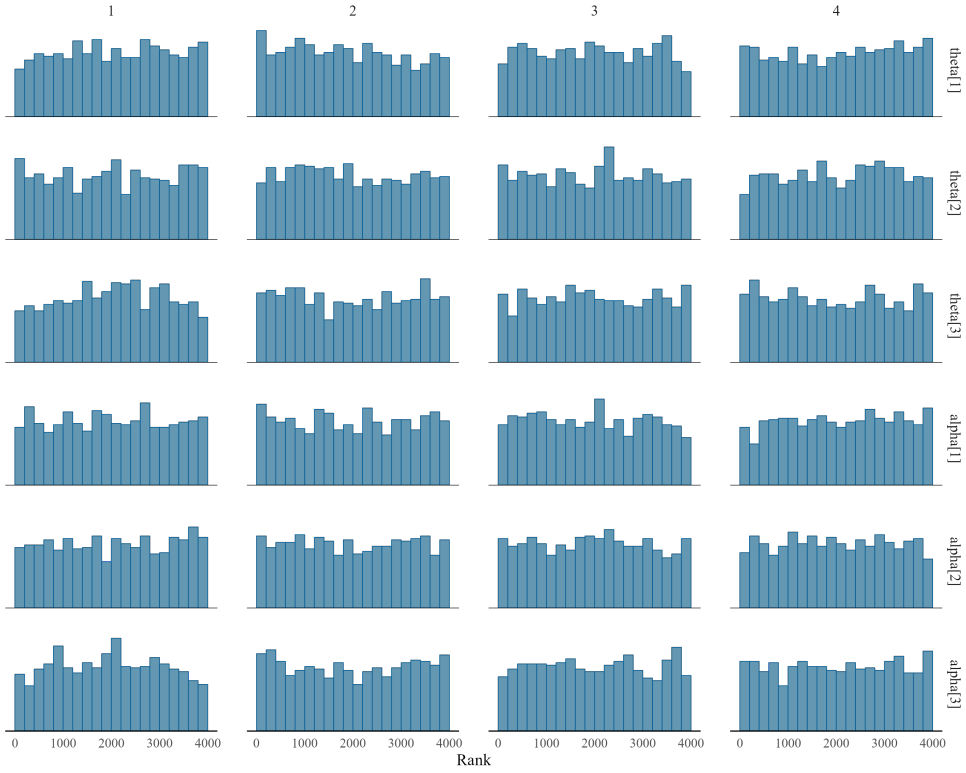
Model checking and evaluation, not already discussed in this thesis to date, included checking the posterior estimates are not contrary to expert knowledge from within the systematic review group (discussed more in the next chapter) and sensitivity analyses. All meta-analyses compared a model that relaxed the assumption of conditional independence with a model that assumed conditional independence and some went further with additional sensitivity analyses as described in sections 5.2.3, 5.2.4 and 5.2.5. While there is no general criteria for identifiability in these models [119], utilising a range of tools for model checking is pivotal.

Figure 4.5: Example of graphical convergence diagnostics a) trace plots and b) rank plots for θ_j and α_j of the first three studies from the meta-analysis model for the motivating leishmaniasis example including all 63 studies

(a) Trace plots for all four chains for 6 parameters



(b) Rank plots for each chain shown as columns for 6 parameters shown as rows



Chapter 5

Application of a Bayesian random-effect meta analysis

5.1 Preamble

Chapter 4 introduced how meta-analysis can be performed in the context of investigating diagnostic test accuracy and provided an overview of the meta-analysis model used to estimate the sensitivity and specificity of diagnostic tests used in the FIEBRE study (Objective 3, Section 1.5).

The FIEBRE study carried out a systematic review and meta-analysis to estimate the diagnostic test accuracy of 10 tests used in the diagnosis of 6 pathogens. These are:

1. Influenza [126]
2. Respiratory syncytial virus (RSV) [126]
3. Cryptococcosis (CrAg) [127]
4. Leishmaniasis [108]
5. Leptospirosis [128]
6. Dengue (Drafted for publication)

This chapter introduces these reviews that have all been, or are in the process of being, published. I first discuss how each of these reviews requires a slightly nuanced approach to the analysis before presenting the systematic review and meta-analysis for diagnostics used in FIBERE in the diagnosis of human leptospirosis. In this chapter, I include the full manuscript and supplementary material relevant to the statistical analysis.

Each diagnostic test accuracy systematic review was carried out by a small research team. I helped conceive and design the data extraction

tool for each review but for all reviews other than the dengue review, I had no role in data extraction. Once data were extracted, I would review the data with the review team to ensure all variables were being interpreted correctly. Next, I would formulate and code the meta-analysis model, summarise the results in figures and tables before interpreting the results once again with the review team. For each review's manuscript, I wrote the statistical analysis section of the methods as well as the meta-analysis results section. For leptospirosis, leishmanaisis and dengue, I also helped write the first draft of the manuscript. Given the nuances required for each review, it was paramount that these reviews were carried out by a multidisciplinary team of clinicians, diagnostic laboratory experts, statisticians and epidemiologists. As a result, I have had the pleasure of working with a large number of inspiring researchers, as each review was carried out by a different team, and I wish to thank them all for allowing me to be a part of these interesting reviews.

I would like to thank Heidi Hopkins (Scientific coordinator of the FIEBRE study) and my supervisors, John Bradley, Ruth Keogh and Oliver Baerenbold, for helping across all of these reviews. RSV/Influenza: Sophie Jullien, Felicity Fitzgerald, Colin Fink and Marie Voice. CrAg: Catriona Macrae, Jayne Ellis and Joseph Jarvis. Leishmaniasis: Tamalee Roberts, Sayaphet Rattanavong, Santiago Gomez, Petra Mens and Elizabeth Ashley. Leptospirosis: Marta Valente, Justina Bramugy, Quique Bassat, Paul Newton, Mathieu Picardeau and John Crump. Dengue: Kamla Pillay, Elizabeth Fitchett, Cassandra Akinde, Audrey Dubot-Peres, Zhia Lim, David Mabey, Behrouz Maldonado, Laura Maynard-Smith, Ellen Sugrue and Okuda Taylor. Last but not least thanks to Jane Falconer and the rest of the Library team at LSHTM.

5.2 FIEBRE meta-analyses practical considerations

The planned analysis for each diagnostic test was the same however, upon reviewing the data collected from each systematic review, practical considerations required nuances to the analysis plan for each. Here I summarise these nuances. The reviews are introduced in chronological order of when the analyses were carried out. Each review that has been published is available open access and all code to run the analyses are available on github.

5.2.1 Respiratory syncytial virus and Influenza

The index test for Respiratory syncytial virus (RSV) and Influenza was the Luminex Respiratory Pathogen panel (RPP). This systematic review and meta-analysis is presented in Appendix A [126]. We found variability in the included studies as to whether the index test was used to detect single sub-types (e.g. RSV A) or whether the results did not distinguish between sub-types (e.g. RSV A and/or B). Because we believed the diagnostic test may have a different accuracy depending on the sub-type, we introduced a covariate into the model for sub-type. A full model specification is provided in the supplementary material of the paper.

5.2.2 Cryptococcosis

The index test for Cryptococcosis (CrAg) was the IMMY CrAg lateral flow assay. This systematic review and meta-analysis has been published and is presented in Appendix B [127]. The review suffered from few studies and sparse data (zero cells) in those studies. As a result, we chose to fit a simpler hierarchical model than planned using fixed effects instead of random effects. An alternative approach could have been to assume a symmetric SROC curve rather than allow differences in the variation of outcomes between disease positive and disease negative individuals [129].

5.2.3 Leishmaniasis

The index test for Leishmaniasis was the direct agglutination test. This systematic review and meta-analysis has been published and is presented in Appendix C [108]. This was a large review (n=78) with many expected potential sources of variability in test accuracy across studies. This review featured a sensitivity analysis more extensive than the first two and considered variability by patient characteristics (symptomatic only or HIV-positive only), geographical region where test was performed and sample type used in the test (freeze-dried or liquid) as well as exploring the conditional independence assumption.

5.2.4 Leptospirosis

There were three index tests for Leptospirosis: Microscopic agglutination test (MAT), PCR and IgM ELISA. This systematic review and meta-analysis has been published and is presented in section 5.3 [128]. The reviews for PCR and IgM ELISA found less than four studies and so no meta-analysis was carried out. This review was different to those we had

already done because of the type of tests investigated. The microscopic agglutination test is a serological test that is ideally carried out at two time points. The outcome of the test is then only determined after the second test and by comparing the result at the first time point with the result at the second time point with a positive result often determined by a four-fold rise in titre between the tests or seroconversion. Guidance states the time between the two tests should be ten days [130]. However, as with many aspects of healthcare, loss to follow up means a substantial proportion of individuals only receive the test at the first time point and do not return for the second or 'paired' test. As a result of this common phenomena, it was of interest to estimate the accuracy of the test not only in the ideal scenario of 'paired' test results but also when only the first test result is available. Results found support the guidance that the first test should not be used in isolation to diagnose acute leptospirosis.

5.2.5 Dengue

The Dengue review was the largest of the systematic reviews and included 193 papers after applying selection criteria and had five index tests: Reverse Transcription-PCR (RT-PCR), IgM ELISA, IgG ELISA, NS1 ELISA and Viral neutralization. This systematic review and meta-analysis has been drafted for submission and is presented in Appendix D. The timing of the tests in relation to days post onset of fever is particularly important in this review as the index tests work by different mechanisms which specifically relate to different responses of the body to infection. For example, RT-PCR detects viral RNA which peaks between one and four days post onset of symptoms. In comparison, IgM ELISA detects antibodies which peak between days five and 14 depending on whether it is a primary or secondary infection [131]. These details are important to consider when incorporating prior information on test accuracy, for example, when IgM ELISA is used as a reference test against NS1 ELISA on samples taken between zero and four days post onset of symptoms our prior beliefs about the accuracy of IgM ELISA would be different than if those samples were taken between four and seven days post onset of fever.

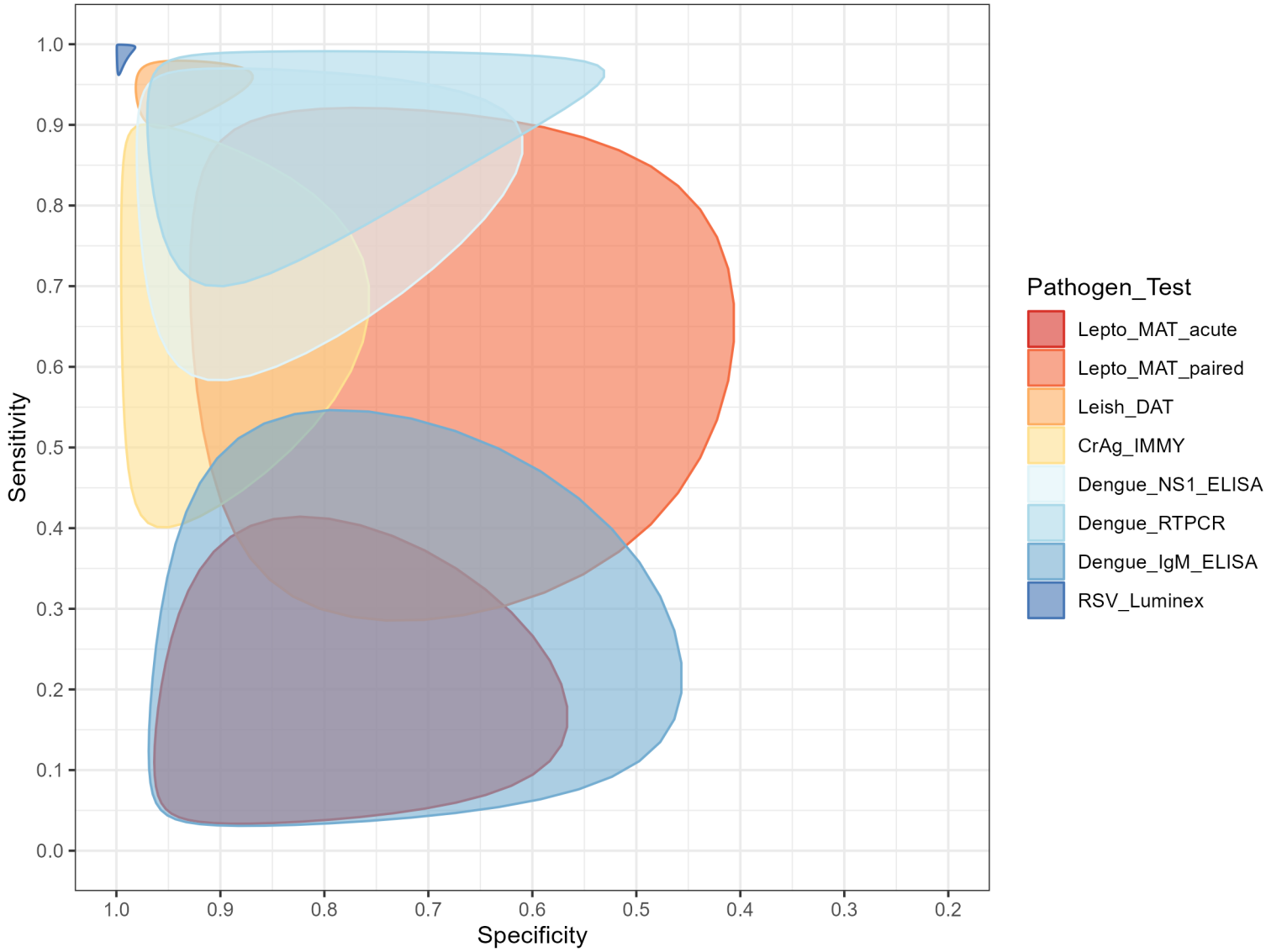
5.2.6 Overall

A summary of the key characteristics of each review are provided in Table 5.1 and, the partially-pooled estimates from selected pathogen/test combinations is presented in Figure 5.1.

Table 5.1: Characteristics of diagnostic test accuracy meta-analyses carried out in this thesis

Infection	**Number of diagnostic tests investigated**	**Number of studies included**	**Symptomatic (febrile) cohort?**	**Covariates considered**	**Sensitivity analysis details**	**Reference (if published)**
RSV	1	10	No	None	With and without conditional dependence between disease positive individuals via random effects and varied prior distributions for variance parameters	[126]
Influenza	1	9	No	Influenza A or Influenza B	With and without conditional dependence between disease positive individuals via random effects and varied prior distributions for variance parameters	[126]
CrAg	2	9	No	None	With and without conditional dependence between disease positive via fixed effects	[127]
Leishmaniasis	1	78	No	Patient group (Symptomatic, HIV positive), Geographic region of study, Freeze-dried or liquid samples	With and without conditional dependence between disease positive via random effects	[108]
Leptospirosis	4	15	Yes?	None	With and without conditional dependence between disease positive via random effects and with and without non-endemic countries	[128]
Dengue	5	193	Yes	Days post onset of fever	Only those studies at low risk of bias and low cause for applicability concerns	Na

Figure 5.1: Partially-pooled 95% credible regions for each diagnostic test used to diagnose each pathogen



Note. Results shown include: the MAT test for leptospirosis (lepto) in acute and paired samples, DAT test for leishmaniasis, IMMY rapid test for CrAg on serum, three tests for dengue (NS1 and IgM ELISA and RT-PCR) all carried out in samples collected between 0 and 4 days post onset of symptoms

5.3 Paper

Addendum: The notation used in the model specification published in the Supplementary material of the published paper differs in notation to that presented in Chapter 4 of this thesis, however the model in this published paper is the same as that presented in Chapter 4.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1702488	**Title**	Miss
First Name(s)	Suzanne		
Surname/Family Name	Keddie		
Thesis Title	Latent class models for diagnostic test accuracy with application to fever aetiology		
Primary Supervisor	John Bradley		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMC Infectious Diseases		
When was the work published?	07/02/2024		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

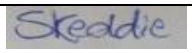
Where is the work intended to be published?	

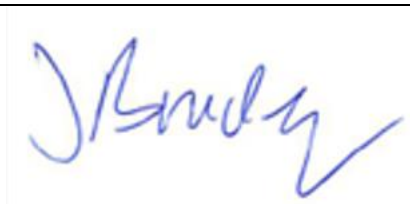
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am joint first author of the paper. I developed, coded and carried out the statistical analysis with support from my supervisors. I also created data visualizations of the results for the manuscript. Along with the other first authors, I prepared the original draft of the manuscript.

SECTION E

Student Signature	
Date	07/02/2024

Supervisor Signature	
Date	25/03/2024

BMC Infectious Diseases

RESEARCH

Open Access

Diagnosis of human leptospirosis: systematic review and meta-analysis of the diagnostic accuracy of the *Leptospira* microscopic agglutination test, PCR targeting *Lfb1*, and IgM ELISA to *Leptospira fainei* serovar Hurstbridge

Marta Valente[1,2*†], Justina Bramugy[2†], Suzanne H. Keddie[3†], Heidi Hopkins[3], Quique Bassat[1,2,4,5,6], Oliver Baerenbold[3], John Bradley[3,7], Jane Falconer[3], Ruth H. Keogh[3], Paul N. Newton[8,9], Mathieu Picardeau[10] and John A. Crump[11]

Abstract

Background Leptospirosis is an underdiagnosed infectious disease with non-specific clinical presentation that requires laboratory confirmation for diagnosis. The serologic reference standard remains the microscopic agglutination test (MAT) on paired serum samples. However, reported estimates of MAT's sensitivity vary. We evaluated the accuracy of four index tests, MAT on paired samples as well as alternative standards for leptospirosis diagnosis: MAT on single acute-phase samples, polymerase chain reaction (PCR) with the target gene *Lfb1*, and ELISA IgM with *Leptospira fainei* serovar Hurstbridge as an antigen.

Methods We performed a systematic review of studies reporting results of leptospirosis diagnostic tests. We searched eight electronic databases and selected studies that tested human blood samples and compared index tests with blood culture and/or PCR and/or MAT (comparator tests). For MAT selection criteria we defined a threshold for single acute-phase samples according to a national classification of leptospirosis endemicity. We used a Bayesian random-effect meta-analysis to estimate the sensitivity and specificity of MAT in single acute-phase and paired samples separately, and assessed risk of bias using the Quality Assessment of Studies of Diagnostic Accuracy Approach- 2 (QUADAS-2) tool.

Results For the MAT accuracy evaluation, 15 studies were included, 11 with single acute-phase serum, and 12 with paired sera. Two included studies used PCR targeting the *Lfb1* gene, and one included study used IgM ELISA with *Leptospira fainei* serovar Hurstbridge as antigen. For MAT in single acute-phase samples, the pooled sensitivity and specificity were 14% (95% credible interval [CrI] 3–38%) and 86% (95% CrI 59–96%), respectively, and the predicted sensitivity and specificity were 14% (95% CrI 0–90%) and 86% (95% CrI 9–100%). Among paired MAT samples,

†Marta Valente, Justina Bramugy and Suzanne H. Keddie contributed equally and share primary authorship (joint first authors) on this work.

*Correspondence:
Marta Valente
anamarta.valente@gmail.com
Full list of author information is available at the end of the article

the pooled sensitivity and specificity were 68% (95% CrI 32–92%) and 75% (95% CrI 45–93%) respectively, and the predicted sensitivity and specificity were 69% (95% CrI 2–100%) and 75% (2–100%).

Conclusions Based on our analysis, the accuracy of MAT in paired samples was not high, but it remains the reference standard until a more accurate diagnostic test is developed. Future studies that include larger numbers of participants with paired samples will improve the certainty of accuracy estimates.

Keywords Leptospirosis, Meta-analysis, Agglutinations tests, Polymerase chain reaction, Enzyme-linked immunosorbent assay, Systematic review, Sensitivity and specificity

Background

Leptospirosis is an underdiagnosed infectious disease, with an estimated global annual number of illnesses of more than one million per year from 1970 to 2008 [1], 60,000 estimated annual deaths [1], and a mortality ratio ranging from 2% through to 60%, among older patients with icteric disease or renal failure [2]. Although tropical regions have the highest incidence of disease, with climate change and massive urbanization of frequently flooded areas in low-income countries, the epidemiology of this zoonosis is changing and it is a growing global public health problem [3–5]. In tropical and subtropical settings, the symptoms and signs of leptospirosis overlap with those of many other acute febrile illnesses including malaria, arboviral, and rickettsial diseases, and thus require laboratory confirmation for diagnosis [6–8].

Numerous diagnostic tests based on nucleic acid or antibody detection have been developed for early diagnosis of leptospirosis [9], but the serologic reference standard remains the microscopic agglutination test (MAT) on paired samples with a four-fold or greater rise, or seroconversion, confirming the diagnosis [10, 11]. Nevertheless, reported estimates of sensitivity vary [12, 13]. The clinical characteristics of the populations studied, including days post-onset of symptoms and prior use of antibacterials, the serovars included in the MAT panel in relation to the epidemiology of the disease in the geographic region studied, as well as the laboratory performance, contribute to heterogeneous estimates of MAT sensitivity in paired samples [11–13].

Because MAT is an imperfect reference test, accuracy evaluations that do not account for the imperfect nature of the test are biased [13, 14]. To explore this, Bayesian latent class analysis can be used to estimate the accuracy of a test, without assuming that any test is 100% accurate [15]. To our knowledge there is no published systematic review regarding MAT diagnostic accuracy using latent class analysis.

The Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) study is a prospective observational study of the infectious causes of fever at four sites in Africa and Asia, collecting data and samples from adult and paediatric outpatients, inpatients, and community controls [16]. FIEBRE tests for preventable and treatable infections, including leptospirosis, using reference standard diagnostic tests performed at specialised laboratory centres of excellence. The approach for the diagnosis of leptospirosis used in FIEBRE was an initial IgM ELISA screen using *Leptospira fainei* serovar Hurstbridge antigen on participants' convalescent sera, or for participants who did not provide convalescent serum, screening of acute serum from the day of clinical presentation. For IgM ELISA positive samples, MAT using a globally representative panel of *Leptospira* serovars enriched when possible with local strains was performed on acute and, when available, convalescent sera. MAT was also performed on all acute plasma samples positive by SYBR Green based real-time polymerase chain reaction (PCR) assay targeting the *Lfb1* gene [17, 18].

We conducted a systematic review and meta-analysis to assess the accuracy of the index tests: MAT, PCR with the pathogenic *Leptospira* target gene *Lfb*1, and ELISA IgM with the target antigen *Leptospira fainei* serovar Hurstbridge. We compared the index tests with reference standard diagnostic tests for lepstospirosis diagnosis [10]: blood culture and/or PCR and/or MAT (comparator tests). We used a Bayesian latent class model to evaluate the sensitivity and specificity of MAT on single acute-phase samples and MAT on paired samples.

Methods

PROSPERO protocol

The protocol of our systematic review was developed prior to conducting the review, and was registered in the International Prospective Register of Systematic Reviews (PROSPERO) at https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=285773, registration number CRD42021285773.

Search strategy

The original searches were conducted by a library information specialist (JF) on 9 September 2020 for PCR, 10 September 2020 for MAT, and 30 November 2020 for IgM ELISA, and all searches were updated on 16 August 2022. Databases searched included OvidSP Medline, OvidSP Embase, OvidSP Global Health, Wiley Cochrane

Central Register of Controlled Trials, Clarivate Analytics Web of Science (Science Citation Index Expanded and Social Sciences Citation Index only), Elsevier Scopus, Ebsco Africa-Wide Information, World Health Organization (WHO) Latin American and Caribbean Health Sciences Literature, and WHO Global Index Medicus.

The search included strings of terms, synonyms, and controlled vocabulary terms to reflect two concepts: leptospirosis, and either MAT, PCR, or IgM ELISA, hereafter referred to as the index test of each search. The exact search terms used for each search are shown in the Supplementary material (Appendix S1). Animal studies were excluded, and the search was limited by date of publication from 1950 when MAT protocols were initially published [19] through 16 August 2022. Duplicates were removed. Additional eligible studies were found by manually searching the reference lists of relevant manuscripts and by contacting authors.

Selection criteria

The selection criteria applied to all studies found in the search are detailed in Table 1.

For the MAT systematic review, we included the threshold of single acute-phase sample in the selection criteria. Since leptospirosis case definitions for single acute-phase samples vary according to background seroprevalence [10], we sub-classified the study settings considering where leptospirosis is endemic and non-endemic based on national level assessments. In line with Costa et al. [1] we considered non-endemic settings to be countries with 10 or fewer leptospirosis cases per 100,000 population per year, and endemic settings to be countries with more than 10 cases per 100,000 population per year. Costa's review [1] identified 80 studies from 34 countries that fulfilled the selection and quality criteria for a disease incidence study with a defined study period of leptospirosis endemic transmission, and developed a multivariable regression model to estimate leptospirosis incidence for each country and territory.

Following this rationale, we set as selection criteria the titre cut-off for a positive MAT in a single acute-phase sample of $\geq$ 1:400 for endemic settings, and $\geq$ 1:100 for non-endemic settings. For all settings, the criteria for a serologically confirmed case of leptospirosis was defined as seroconversion or a four-fold or greater rise in MAT antibody titre between paired samples from a person with a history of measured or reported fever, or with suspected leptospirosis [10].

Study selection and data extraction

Two reviewers (JB, MV) screened and selected all studies independently and in duplicate, using two separate Excel spreadsheets (Authors, Title, Abstract, Journal, Year, Volume, Issue, Pages, DOI) for MAT and PCR studies, and for IgM ELISA studies using the online tool Cadima (https://www.cadima.info/) [20].

The initial eligibility assessment of all titles and abstracts identified by the search strategy was performed using the predetermined selection criteria (Table 1). Full-text copies of all potentially eligible reports were retrieved and reviewed, independently and in duplicate by JB and MV. Any disagreements about eligibility were resolved through discussion between JB and MV, leading to the inclusion of reports meeting all selection criteria and exclusion of those not meeting criteria. For each included report, JB and MV independently abstracted data using a standardized data abstraction sheet that was first piloted on fifteen studies (see Supplementary material, Table S1). We contacted study investigators when a report appeared to meet selection criteria, but data reported were unclear or insufficient to abstract a 2×2 contingency table comparing one or more index with another test. If sufficient data were not available or there was no reply from the authors, the study was excluded.

Table 1 Selection criteria applied to studies found in the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022

Selection criteria
1) Studies performed using human blood samples
2) Observational and interventional studies among patients with fever history or suspected leptospirosis
3) Article in English, Spanish or Portuguese
4) Test of interest (MAT, PCR targeting the *Lfb1* gene or IgM ELISA with the target antigen Hurstbridge) and at least one comparator test (MAT, PCR with any target gene or Culture) performed on the same samples
5) Data for extraction of a 2×2 contingency table
6) For studies for MAT accuracy evaluation, results of testing single acute samples presented separately from results of testing paired samples (i.e. acute and convalescent samples)
7) For studies for MAT accuracy evaluation, threshold for single acute-phase samples in endemic settings ≥ 1:400 and in non-endemic setting ≥ 1:100

Bias assessment

We assessed study quality using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) criteria, which assesses both the risk of bias and applicability to the review question for four domains: participant selection, index test, reference standard, and flow and timing of participants [21]. Each included article was graded as 'low risk' or 'high risk.' Each category was defined according to the criteria included in the manuscript, as shown in Tables 2 and 3.

Data analysis

For analysis we required data from each study in the form of a 2×2 contingency table showing results of the index test and a comparator test. The index test was any of the tests of interest for each systematic review: single acute-phase MAT, paired MAT, PCR with target gene *Lfb1,* or ELISA IgM with target antigen Hurstbridge. The comparator tests were pre-determined before beginning the review according to the reference standard diagnostic tests for lepstospirosis diagnosis [10]. When MAT (on either a single sample or paired sera) was the index test, the comparator tests were blood culture and/or PCR to any target gene; when PCR with target gene *Lfb1* was the index test, the comparator test was MAT (on either a single sample or paired sera) and/or blood culture and/or PCR (with other target genes); when ELISA IgM was the index test, the comparator test was MAT (on either a single sample or paired sera) and/or PCR (with any target gene) and/or blood culture.

Regarding MAT (on either a single sample or paired sera) meta-analysis, when a study reported data on multiple comparator tests, we created separate 2×2 contingency tables comparing the index test with each comparator test. In these cases, without individual level data we were unable to include all data in the meta-analyses without introducing bias. To systematically ensure only one 2×2 table from each study was included in the meta-analyses, we chose to include the 2×2 table where the comparator test was blood culture. This choice was made because more accuracy data on the specificity of blood culture are available than data on the sensitivity or specificity of PCR [22].

We implemented a Bayesian random-effect latent class meta-analysis, which is an extension to the Hierarchical Summary Receiver Operating Characteristic (HSROC) Model [18] to estimate the sensitivity and specificity of index tests. This framework took into account the imperfect nature of all tests included, as well as accounting for within- and between-study variability.

We fitted separate meta-analyses for MAT single acute-phase and paired sera, and for each analysis calculated the median and 95% credible interval (CrI) for the estimated sensitivity and specificity of the index test in each study. Importantly, we also calculated both the estimated median and 95% CrI for sensitivity and specificity across studies, known as pooled accuracy, as well as the predicted sensitivity and specificity. These predicted values estimate the sensitivity and specificity that would be expected if the test were to be used in a hypothetical future study. These pooled and predicted estimates of accuracy are presented through summary Receiver Operating Characteristic (ROC) curves which represent the 95% credible region for the joint estimate of the index tests sensitivity and specificity. If a meta-analysis could not be performed due to scarcity of data, as was the case with PCR and ELISA reviews, we estimated accuracy of the index test in individual studies using latent class analysis [23].

All analyses were carried out in R using stan [24]. A full model specification including sensitivity analysis investigating the impact on estimates of accounting for conditional dependence between tests within a disease class, as well as results where non-endemic studies are excluded, can be found in Supplementary material (Appendix 2). All code can be found at: https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/Leptospirosis.

Results

Study selection

Single acute-phase and paired MAT

Our systematic review of MAT performed on single acute-phase and paired samples identified 691 reports. Of these, 58 (8.4%) were identified as potentially relevant on the basis of the title and abstract and underwent full-text review. Of these, 15 (25.9%) met our selection criteria and were included [25–39]; 12 (80%) [25–36] tested samples from endemic countries and three (20%) [37–39] from non-endemic countries. Of the 12 studies in endemic countries, nine studies (75%) [25–30, 35, 36] reported data from single acute-phase samples and ten studies (83,3%) [25–29, 31–34] reported data from paired samples. Of the three studies in non-endemic countries, two (66.6%) [37, 38] reported data from single acute-phase samples and two (66.6%) [38, 39] from paired samples. We excluded results of single acute-phase samples from three studies [32, 33, 39] because the threshold of detection used was different from our national leptospirosis endemicity-based selection criteria (Fig. 1).

The studies that were not included due having insufficient data available to create a 2×2 contingent table for single acute-phase samples and/or paired samples are detailed in Appendix S3.

Table 2 Criteria for assessing bias in the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022

Domain	Grade	Criteria
A. Criteria for assessing bias in studies selected for MAT accuracy evaluation		
Patient selection	Low risk	Prospective studies and case-control studies in the same population
	High risk	Case-control studies in different populations or healthy controls; eligibility other than suspected leptospirosis
Index test (MAT)	Low risk	MAT performed in paired samples with a positivity criteria of ≥ 4-fold rise or seroconversion
	High risk	MAT performed in single acute-phase samples; any other positivity criteria for paired samples different than ≥ 4-fold rise or seroconversion
Comparator test (culture and/or PCR)	Low risk	Performed in recruitment samples; performed according to standard methodology
	High risk	Performed in convalescent samples; not performed according to standard methodology
Flow and timing	Low risk	All patients subject to the same comparator tests; comparator tests and index test performed on samples taken at the same time for acute phase
	High risk	Not all participants performed the same comparator test; use of samples collected on different days for acute phase
B. Criteria for assessing bias in studies selected for PCR accuracy evaluation		
Patient selection	Low risk	Prospective studies and case-control studies in the same population
	High risk	Case-control studies in different populations or healthy controls; eligibility other than suspected leptospirosis
Index test (PCR)	Low risk	Performed in recruitment samples; performed according to standard methodology
	High risk	Performed in convalescent samples; not performed according to standard methodology
Comparator test (MAT and/or culture and/or PCR)	Low risk	Use of MAT on paired samples in at least 75% of participants; cases defined with ≥ 4-fold rise in antibody titers or with a positive culture of Leptospira; tests performed according to standard methodology
	High risk	Use MAT on less than 75% of paired samples, any other positivity criteria for paired samples different than ≥ 4-fold rise or seroconversion; tests not performed according to described methodology
Flow and timing	Low risk	All patients subject to the same comparator tests; comparator tests and index tests performed on samples collected at the same time for acute phase
	High risk	Not all participants performed the same comparator test; use of samples collected on different days for acute phase
C. Criteria for assessing bias in studies selected for IgM ELISA accuracy evaluation		
Patient selection	Low risk	Prospective studies and case-control studies in the same population
	High risk	Case-control studies in different populations or healthy controls; eligibility other than suspected leptospirosis
Index test (IgM ELISA)	Low risk	Threshold for positivity defined a priori; test performed according to manufacturer's recommendations
	High risk	Threshold for positivity not defined a priori; test not performed according to manufacturer's recommendations
Comparator test (MAT, culture and/or PCR)	Low risk	Use of MAT on paired samples in at least 75% of participants, cases defined as a positive PCR, MAT with ≥ 4-fold rise in antibody titers or a positive culture of *Leptospira*; tests performed according to described methodology
	High risk	Use MAT on less than 75% of paired samples; culture and PCR performed in convalescent samples, any other positivity criteria for MAT than ≥ 4-fold rise or seroconversion between paired samples; tests not performed according to standard methodology

Table 2 (continued)

Domain	Grade	Criteria
Flow and timing	Low risk	All patients subject to the same comparator tests; comparator tests and index test performed on samples collected at the same time for acute phase
	High risk	Not all participants performed the same comparator test; use of samples collected on different days for acute phase

Table 3 Criteria for assessing applicability in the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022

Domain	Grade	Criteria
A. Criteria for assessing applicability in studies selected for MAT accuracy evaluation		
Patient selection	Low risk	Patients with a febrile illness, symptoms of leptospirosis or fever of unspecified duration
	High risk	Patients without febrile illness or without clinical suspicious of leptospirosis
Index test (MAT)	Low risk	Panel of local known circulating serovars; where local serovars are unknown, a globally representative serovar panel is used; MAT performed according to described methodology
	High risk	Panel without local circulating serovars; MAT not performed according to described methodology
Comparator test (Culture and/or PCR)	Low risk	PCR and/or culture performed according to standard methodology
	High risk	PCR and/or culture not performed according to standard methodology
B. Criteria for assessing applicability in studies selected for PCR accuracy evaluation		
Patient selection	Low risk	Patients with febrile illness, symptoms of leptospirosis or fever of unspecified duration
	High risk	Patients without febrile illness or without clinical suspicious of leptospirosis
Index test (PCR)	Low risk	PCR performed according to standard methodology
	High risk	PCR not performed according to standard methodology
Comparator test (MAT and/or culture and/or PCR)	Low risk	Panel of local known circulating serovars; where local serovars are unknown, a globally representative serovar panel is used; tests performed according to standard methodology
	High risk	Panel without local circulating serovars; tests not performed according to standard methodology
C. Criteria for assessing applicability in studies selected for IgM ELISA accuracy evaluation		
Patient selection	Low risk	Patients with febrile illness, symptoms of leptospirosis or fever of unspecified duration
	High risk	Patients without febrile illness or without clinical suspicious of leptospirosis
Index test (IgM ELISA)	Low risk	IgM ELISA performed according to standard methodology
	High risk	IgM ELISA not performed according to standard methodology
Comparator test (MAT, culture and/or PCR)	Low risk	Panel of local known circulating serovars; where local serovars are unknown, a globally representative serovar panel is used; MAT, PCR and/or culture performed according to standard methodology
	High risk	Panel without local circulating serovars; MAT, PCR and/or culture not performed according to standard methodology

PCR target gene lfb1

Our PCR review identified 1,094 reports. Of these, 18 (1.6%) were identified as potentially relevant on the basis of the title and abstract and underwent full-text review. Of these 18 reports, two (11.1%) articles [27, 40] met our selection criteria and were included (Fig. 1).

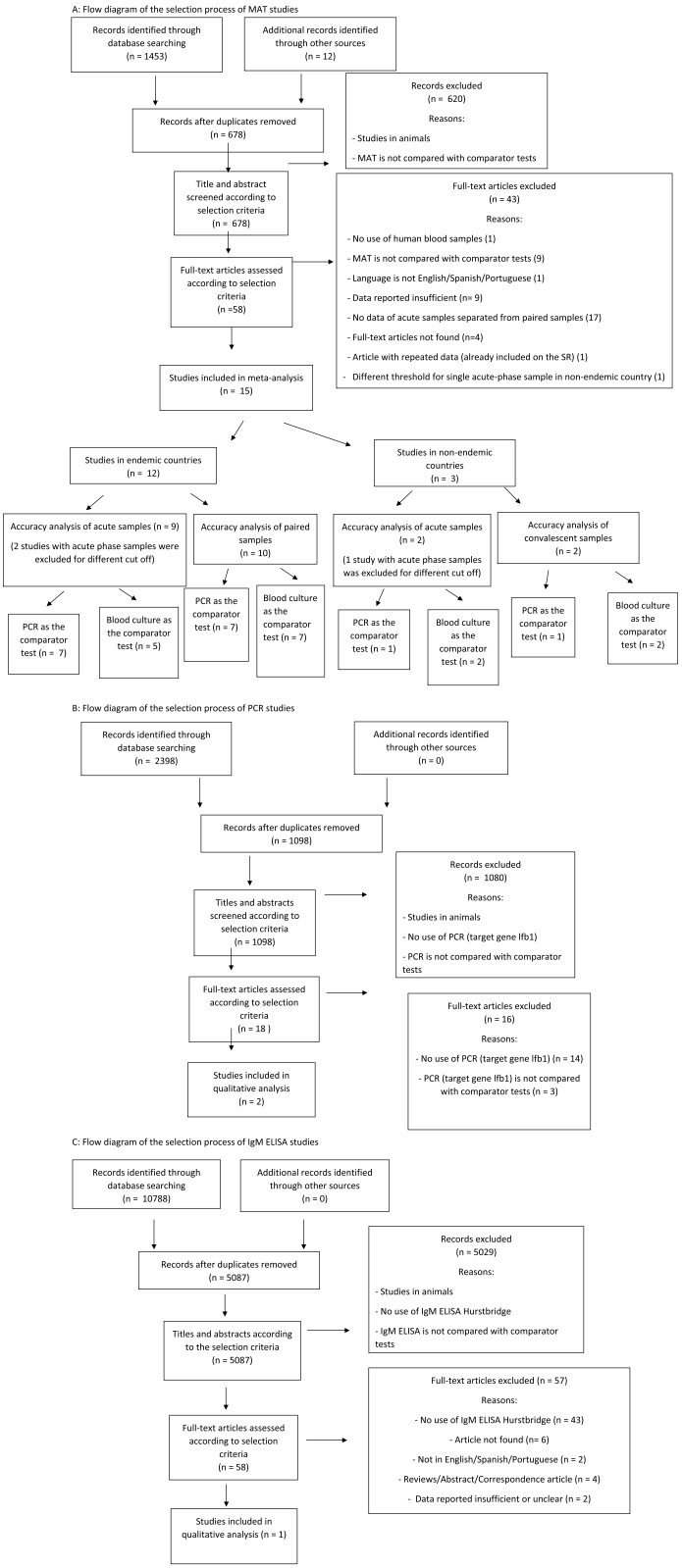


Fig. 1 Study flow diagram for systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022. **A** Flow diagram of the selection process of MAT studies. **B** Flow diagram of the selection process of PCR studies. **C** Flow diagram of the selection process of IgM ELISA studies

ELISA IgM target antigen Leptospira fainei serovar Hurstbridge

Our IgM ELISA review identified 5,092 reports. Of these, 58 (1.1%) were identified as potentially relevant on the basis of title and abstract and underwent full-text review. Of these 58 reports, one (1.7%) article [41] met our selection criteria and was included (Fig. 1).

Study characteristics

Single acute-phase and paired MAT

The characteristics of all included studies are detailed in Table 4. The 15 studies included for MAT (11 (73%) studies were of single-sample MAT, 12 (80%) studies of paired MAT and 8 (53%) studies were of both) were conducted from 2000 through 2020. Of these studies, 14 (93%) of 15 [25–38] included participants with suspected leptospirosis and one (7%) of 15 [39] included participants with fever. Of studies from endemic regions, recruitment occurred in Brazil [28, 29]; Japan [34]; Pacific Island Countries and Territories such as Marquesas Islands, Society Islands, Wallis and Futuna, and New Caledonia [27]; India [32, 33]; Laos [25, 28]; Malaysia [30, 35]; and Thailand [31, 36]. In non-endemic countries, recruitment occurred in New Zealand [39] and Slovenia [37, 38]. All studies were prospective. The MAT panel comprised 20 to 22 serovars in five studies [25, 26, 29, 30, 35], 13 to 15 serovars in three studies [34, 37, 38], and 8 to 11 serovars in three studies [32, 33, 39]. The MAT panel was not described in four studies [27, 28, 31, 36]. The comparator test was blood culture in five studies [29, 32, 33, 36, 37], PCR in four studies [26, 27, 30, 35], and both were used as comparators in six studies [25, 28, 31, 34, 38, 39]. Of studies with PCR as a comparator test, three studies used serum samples [26–28], five used whole blood samples [31, 34, 35, 38, 39], one used both [30], and one study used serum and buffy coat [25]. Recruitment of individuals varied in relation to time of illness onset across studies. The number of days post-onset (DPO) of symptoms at recruitment were 0 to 14 days [34], 1 to 30 days [25, 27], a mean of 6 days [29], and an interquartile range of 2 to 5 [36], 2 to 6 [31], and 3 to 7 days [28]. The DPO of symptoms was not detailed in eight studies [26, 30, 32, 33, 35, 37–39]. The number of days between acute and convalescent samples also varied with reported timeframes including: 7 to 15 days [25, 31, 32], more than 15 days [29, 35, 38], and was not detailed in nine studies [26–28, 30, 33, 34, 36, 37, 39].

PCR target gene lfb1

The two studies included for PCR accuracy analysis were conducted 2004–2005 [27] and 2015–2016 [40]. Both studies included patients with suspected leptospirosis, were prospective, and enrolled in the endemic countries Azores [40], and the Pacific Island Countries and Territories of Marquesas Islands, New Caledonia, Society Islands, and Wallis and Futuna [27]. In one study [27] the comparator test was MAT, in which the MAT panel was not described, and 10 (24%) of 41 patients had paired samples. In other study [40] the comparator test was PCR targeting the *rrs* gene in serum samples. The DPO of symptoms was of 1 to 30 days in one study [27] and was not described in other study [40].

ELISA IgM with antigen Leptospira fainei serovar Hurstbridge

The eligible study included for IgM ELISA accuracy analysis [41] was conducted in France, French Polynesia, Guadeloupe, Guyana, and Martinique, and was a two-gate design study that included patients with suspected leptospirosis and controls from patients with evidence of recent infection for dengue and syphilis, or from healthy blood donors. IgM ELISA was performed in serum samples and the comparator test was MAT. The MAT panel included 22 serovars, and it was not mentioned how many participants had paired samples.

Study quality

The results of bias assessment are shown in Table 5.

Single acute-phase and paired MAT

In the patient domain, all studies were graded as low risk of bias and applicability, because they were all prospective and with a population of suspected leptospirosis or febrile patients. In the index test domain, when studies used single acute-phase samples for a confirmatory diagnosis of leptospirosis [25–31, 35–38], they were graded as high risk of bias. When studies used paired samples for a confirmatory diagnosis of leptospirosis [25–34, 38, 39], they were graded low risk of bias on the basis that the positivity criteria included a four-fold rise or greater, or seroconversion, between samples. Regarding applicability, nine studies were graded low risk because they used a globally representative panel of 20 to 22 serovars [25, 26, 29, 30, 35], or used 10 to 15 locally known circulating serovars [32, 33, 37, 38]. Two studies [34, 39] were graded high risk since the MAT panels composed of 13 serogroups and eight serovars, respectively, and they were not mentioned as being locally representative of the study setting. Finally, four studies [27, 28, 31, 36] were graded high risk because MAT panel composition was not described.

Table 4 Characteristics of studies selected in the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950 – 2022 for MAT, PCR and IgM accuracy evaluation

A. MAT endemic studies

Study first author (ref)	Title	Journal	Year of recruitment	Country	Endemicity	Study setting	Type of study	Participants included	Mean age / Children	Case definition	MAT panel	Comparator test	DPO of fever at recruitment	Sample type for PCR test	Acute sample data/ Convalescent samples data	Interval between acute and convalescent samples
Woods K [25]	A comparison of two molecular methods for diagnosing leptospirosis from three different sample types in patients presenting with fever in Laos	Clin Microbiol Infect	2014	Vietiane, Laos	Yes	Mahosot Hospital	Prospective	Suspected leptospirosis	39 years (0.5–97) / Yes	Titers of ≥ 1:400 or a fourfold rise in titre	22 serovars. Performed at the WHO/FAO/OIE Collaborating Centre for Leptospirosis Reference and Research, Queensland, Australia	Blood culture and PCR	1–30 days	Serum or buffy coat	Yes/Yes	10–14 days
Blanco R [26]	Evaluation of nested polymerase chain reaction for the early detection of Leptospira spp. DNA in serum samples from patients with leptospirosis	Diagn Microbiol Infect Dis	2010	Brazil, Sao Paulo	Yes	Not stated	Prospective	Suspected leptospirosis	Not stated	Threshold: ≥ 1:800 or a fourfold rise	21 serovars most frequently found in São Paulo, Brazil	PCR	Not stated	Serum	Yes/Yes	Not stated
Merien F [27]	A rapid and quantitative method for the detection of Leptospira species in human leptospirosis	FEMS Microbiol Lett	2004–2005	Pacific Island Countries and Territories	Yes	Not stated	Prospective	Suspected leptospirosis	30 years / Yes	Threshold: ≥ 1:400 or a fourfold rise	Not described	PCR	1–30 days	Serum	Yes/Yes	Not stated

Table 4 (continued)

Dittrich S [28]	A Prospective Hospital Study to Evaluate the Diagnostic Accuracy of Rapid Diagnostic Tests for the Early Detection of Leptospirosis in Laos	Am J Trop Med Hyg	2014–2015	Laos	Yes	Mahosot Hospital	Prospective	Suspected leptospirosis or typhus	39 years (0.5–92) / Yes	Threshold: ≥ 1:400 or fourfold rise	MAT was performed and interpreted by the WHO Collaborating Center for Reference and Research on Leptospirosis, Australia	Blood culture and PCR	Interquartile range: 3–7 days	Serum	Yes/Yes	Not stated
Albuquerque A [29]	Validation of a case definition for leptospirosis diagnosis in patients with acute severe febrile disease admitted in reference hospitals at the State of Pernambuco, Brazil	Rev Soc Bras Med Trop	2009	Pernambuco, Brazil	Yes	Hospital Barão de Lucena and Hospital Universitário Oswaldo Cruz	Prospective	Suspected leptospirosis	32.9 (Standard deviation 13.2) / No	Threshold: > 1:800 or a fourfold rise	22 serovars	Blood culture	6.1 ± 2.6 DPO days	Not applicable	Yes/Yes	≥ 14 days
Philip N [30]	Combined PCR and MAT improves the early diagnosis of the biphasic illness leptospirosis	PLoS One	2016–2017	Malasya	Yes	Hospital Serdang, Hospital Tengku Ampuan Rahimah and Hospital Teluk Intan	Prospective	Suspected leptospirosis	Not stated	Threshold: > 1:400 or a fourfold rise	20 serovars local and internationals	PCR	Not stated	Serum and Whole blood	Yes/Yes	Not stated
Dinhuzen J [31]	A prospective study to evaluate the accuracy of rapid diagnostic tests for diagnosis of human leptospirosis	PLoS Negl Trop Dis	2015–2016	Thailand	Yes	15 hospitals in the Srisaket province	Prospective	Suspected leptospirosis	46 (Standard deviation 17) / No	Threshold: > 1:400 or a fourfold rise	Not described	Blood culture and PCR	Interquartile range 2–6 days	Whole blood	Yes/Yes	7 days

Table 4 (continued)

Mullan S [32]	An Important Tool for Early Diagnosis of Leptospirosis Cases	J Clin Diagn Res	2008	India	Yes	New Civil Hospital and peripheral health centre of South Gujarat	Prospective	Suspected leptospirosis	Unclear / No	No	11 serogroups	Blood culture	Not stated	Not applicable	No/Yes	15 days
Vijayachari P [33]	Evaluation of Lepto Dri Dot as a rapid test for the diagnosis of leptospirosis	Epidemiol Infect	2000–2001	India	Yes	3 primary health centres in South Andaman	Prospective	Suspected leptospirosis	Not stated	No	10 serovars commonly encountered in India	Blood culture	Not stated	Not applicable	No/Yes	Not stated
Kakita T [34]	Laboratory diagnostic, epidemiological, and clinical characteristics of human leptospirosis in Okinawa Prefecture, Japan, 2003–2020	PLoS Negl Trop Dis	2003–2020	Japan, Okinawa Prefecture	Yes	Clinics and hospitals in Okinawa Prefecture	Prospective	Suspected leptospirosis	Not stated	No	13 serovar strains of 12 serogroups	Blood culture and PCR	0–14 days	Whole blood	No/Yes	Not stated
Alia S [35]	Diagnostic accuracy of rapid diagnostic tests for the early detection of leptospirosis	J Infect Public Health	2016–2017	Malasya	Yes	Hospital Serdang	Prospective	Suspected leptospirosis	Not stated	Threshold: ≥ 1:400 or a four-fold rise	20 serovars	PCR	Not stated	Whole blood	Yes/No	21–30 days
Sukmark T [36]	Diagnostic accuracy of rapid diagnostic tests for the early detection of leptospirosis	PLoS Negl Trop Dis	2012–2014	Thailand	Yes	11 centers in 8 provinces around Thailand	Prospective	Suspected leptospirosis	Not stated	Threshold: ≥ 1:400 or a four-fold rise	Not described	Blood culture	Interquartile range 2–5 days	Not applicable	Yes/No	Not stated

Table 4 (continued)

B. MAT non-endemic studies

Study first author (ref)	Title	Journal	Year of recruitment	Country	Endemicity	Study setting	Type of study	Participants included	Mean age / Children included	Case definition	MAT panel	Comparator test	DPO of fever at recruitment	Sample type for PCR test	Acute sample data/ Convalescent sample data	Days between acute and convalescent sample
Podgoršek D [37]	Evaluation of real-time PCR targeting the lipL32 gene for diagnosis of Leptospira infection	BMC Microbiology	Not stated	Slovenia	No	Different hospitals in Slovenia	Prospective	Febrile Patients	Not stated	Titers of ≥ 1:100	15 serovars from the geographic area	Blood culture	Not stated	Not stated	Yes/No	Not stated
Podgoršek D [38]	Evaluation of the immunochromatographic (Leptocheck) test for detection of specific antibodies against leptospires	Wien Klin Wochenschr	Not stated	Slovenia	No	Different hospitals in Slovenia	Prospective	Suspected leptospirosis	Not stated	Titers of ≥ 1:100	13 serovars from the geographic area	Blood culture and PCR	No stated	Whole blood	Yes/Yes	14–30 days
Earl L [39]	An evaluation of diagnostic tests in a case series of suspected leptospirosis patients seen in primary care	N Z Med J	Not stated	New Zealand	No	General practices in Waikato and 3 medical centers in Wairoa	Prospective	Suspected for Leptospirosis	39 years (11–73) / Yes	Threshold:> 1:400 andourfold rise	8 serovars	Blood culture and PCR	Not stated	Whole Blood	No/Yes	Not stated

C. PCR studies

Study first author (ref)	Title	Journal	Year of recruitment	Country	Endemicity	Study setting	Type of study	Participants included	Mean age / Children included	DPO of fever at recruitment	Sample type for PCR test	Comparator test	MAT Case definition for a positive acute sample	MAT case definition for a positive convalescent sample	MAT Paired samples	MAT panel

Table 4 (continued)

Merien F [27]	A rapid and quantitative method for the detection of Leptospira species in human leptospirosis	FEMS Microbiology Letters	2004–2005	Pacific Island Countries and Territories	High	Not stated	Prospective	Suspected leptospirosis	30 years / Yes	1–30 days	Serum	MAT	≥ 1:400 titer	Seroconversion or two fold-rise between titers	10/41	Unclear
Esteves L [40]	Diagnosis of Human Leptospirosis in a Clinical Setting: Real-Time PCR High Resolution Melting Analysis for Detection of Leptospira at the Onset of Disease	Scientific Reports	2015–2016	Azores	High	Hospital	Prospective	Suspected leptospirosis	48,2 years (Standard deviation 16,4) / Not stated	Unclear	Serum	PCR rrs	Not applicable	Not applicable	Not applicable	Not applicable

D. IgM ELISA studies

Study first author (ref)	**Title**	**Journal**	**Year of recruitment**	**Country**	**Endemicity**	**Study setting**	**Type of study**	**Participants included**	**Median age / Children included**	**DPO of fever at recruitment**	**Sample type for IgM ELISA**	**Comparator test**	**MAT Case definition for a positive acute sample**	**MAT case definition for a positive convalescent sample**	**MAT Paired samples**	**MAT panel**
Bourhy P [41]	Evaluation of an in-house ELISA using the intermediate species Leptospira fainei for diagnosis of leptospirosis	Journal of clinical microbiology	Not applicable	Mainland France and French overseas territories	High and low	Hospital	Case–control, with mixed population	Suspected leptospirosis, other diseases, healthy donors	44 years / Yes	Not stated	Serum	MAT	≥ 1:400 titer	Seroconversion	Unclear	22 serogroups

Table 5 Bias assessment in the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022

A. Endemic countries

Author, reference	Year	Participant Selection	Bias				Applicability		
			Index Test		Comparator Test	Flow and timing	Participant Selection	Index Test	Comparator Test
			Single MAT studies	Paired MAT studies					
Woods K [25]	2017	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Blanco R [26]	2014	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Merien F [27]	2005	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Unclear Risk	Low Risk
Dittrich S [28]	2018	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Unclear Risk	Low Risk
Albuquerque A [29]	2011	Low Risk	High Risk	Low Risk	Unclear Risk	Low Risk	Low Risk	Low Risk	Unclear Risk
Philip N [30]	2020	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Dinhuzen J [31]	2021	Low Risk	High Risk	Low Risk	Low Risk	Low Risk	Low Risk	Unclear Risk	Low Risk
Mullan S [32]	2016	Low Risk	NA	Low risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Vijayachari P [33]	2002	Low Risk	NA	Low risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Kakita T [34]	2021	Low Risk	NA	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk
Alia S [35]	2019	Low Risk	High Risk	NA	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Sukmark T [36]	2018	Low Risk	High Risk	NA	Low Risk	Low Risk	Low Risk	Unclear Risk	Low Risk

B. Non-endemic countries

Author	Year	Patient Selection	Bias				Applicability		
			Index Test		Comparator Test	Flow and timing	Patient Selection	Index Test	Comparator Test
			Single MAT studies	Paired MAT studies					
Podgoršek D [37]	2020	Low Risk	High risk	NA	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Podgoršek D [38]	2015	Low Risk	Hish risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk
Earl L [39]	2021	Low Risk	NA	Low Risk	Low Risk	Low Risk	Low Risk	High Risk	Low Risk

C. Bias assessment of studies selected for PCR accuracy evaluation

Author	Year	Bias				Applicability		
		Patient Selection	Index Test	Comparator Test	Flow and timing	Patient Selection	Index Test	Comparator Test
Merien F [27]	2005	Low Risk	Low Risk	High Risk	Low Risk	Low Risk	Low risk	Unclear Risk
Esteves L [40]	2011	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low risk	Low Risk

Table 5 (continued)

D. Bias assessment of studies selected for IgM ELISA accuracy evaluation

Author	Year	Bias				Applicability		
		Patient Selection	Index Test	Comparator Test	Flow and timing	Patient Selection	Index Test	Comparator Test
Bourhy P [41]	2013	High Risk	Low Risk	High Risk	Low Risk	High Risk	Low Risk	Low Risk

In the comparator test domain, regarding bias and applicability, 14 studies [25–28, 30–39] were graded low risk because the comparator tests were performed in recruitment samples and according to standard methodology. One study [29] was graded high risk because laboratory procedures were not described or referenced. For the timing and flow domain, all studies were graded low risk of bias because patients were subject to the same comparator tests, and comparator tests and index test were performed on samples taken at the same time for acute phase.

PCR target gene lfb1

In the patient and index test domain both PCR studies [27, 40] were graded low risk for quality concerns because they were prospective, in patients suspected of leptospirosis, and the index test was performed in recruitment samples and according to standard methodology. In the comparator test domain, one study [27] was graded high risk of bias because MAT was the comparator test and less than 75% of the samples were paired samples, and graded as high risk for applicability concerns because the MAT panel composition was not described. The second study [40] was graded low risk for quality concerns since the comparator test was performed according to standard methodology. For timing and flow domain, both studies were graded low risk of bias because patients were subject to the same comparator tests, and comparator tests and index test were performed on samples taken at the same time for acute phase.

ELISA IgM target antigen Leptospira fainei serovar Hurstbridge

The single IgM ELISA study [41] was graded high risk of bias and high risk for applicability concerns in the patient domain, because it was a two-gate design study and controls were healthy blood donors or patients with other diseases. In the index test domain, it was graded low risk for quality concerns since it was performed according to detailed standard methodology and the threshold for positivity defined a priori. In the comparator test domain, it was graded as high risk of bias because MAT was the comparator test and there was no information regarding the use of paired samples for a confirmatory case. For timing and flow domain, it was graded as low risk of bias since patients were subject to the same comparator tests, and comparator tests and index test were performed on samples taken at the same time for acute phase.

Sensitivity and specificity estimates

Single acute-phase and paired MAT

Overall, 11 studies with data on single acute-phase samples representing 2,625 individuals and 12 studies on paired samples representing 1,721 individuals were included in a meta-analysis for MAT. Abstracted data are detailed in Supplementary material, Table S2.

For single acute-phase samples, the pooled sensitivity and specificity of MAT were 14% (95% CrI 3–38%) and 86% (95% CrI 59–96%), respectively, and the predicted sensitivity and specificity were 14% (95% CrI 0–90%) and 86% (95% CrI 9–100%). The estimates for the sensitivity and specificity of MAT in each individual study can be found in Fig. 2 and the summary receiver operating characteristic (SROC) curves representing the pooled and predicted estimates in Fig. 3.

Among paired samples, the pooled sensitivity and specificity of MAT were 68% (95% CrI 32–92%) and 75% (95% CrI 45–93%) respectively, and the predicted sensitivity and specificity were 69% (95% CrI 2–100%) and 75% (95% CrI 2–100%). The estimates for individual studies can be found in Fig. 4 and the SROC curves for pooled and predicted estimates in Fig. 5.

PCR targeting lfb1

Two studies were included in our review of PCR diagnosis, including a total of 253 individuals. The estimated median sensitivity of PCR in Merien, et al. [27] was 92% (95% CrI 72–100%) and median specificity was 66% (95% CrI 49–91%). In Esteves, et al. [40] the median sensitivity of PCR was 98% (95% CrI 90–100%) and the median specificity was 99% (98–100%) (Table 6).

ELISA IgM target antigen Leptospira fainei serovar Hurstbridge

A single study that included 519 individuals was identified in our review of IgM ELISA. The estimated median sensitivity of IgM was 97% (93–100%) and the median specificity was 99% (97–100%) (Table 6).

Discussion

We carried out a systematic review of the sensitivity and specificity of MAT, PCR with the target gene *Lfb1,* and IgM ELISA with the antigen *Leptospira fainei* serovar Hurstbridge for diagnosis of human leptospirosis. Our meta-analysis of 15 studies, including 3,188 participants, found that MAT on single acute-phase samples had a predicted median sensitivity and specificity of 14% and 86%, respectively, for detecting leptospirosis, and using paired samples MAT had a predicted median sensitivity and specificity of 69% and 75%, respectively.

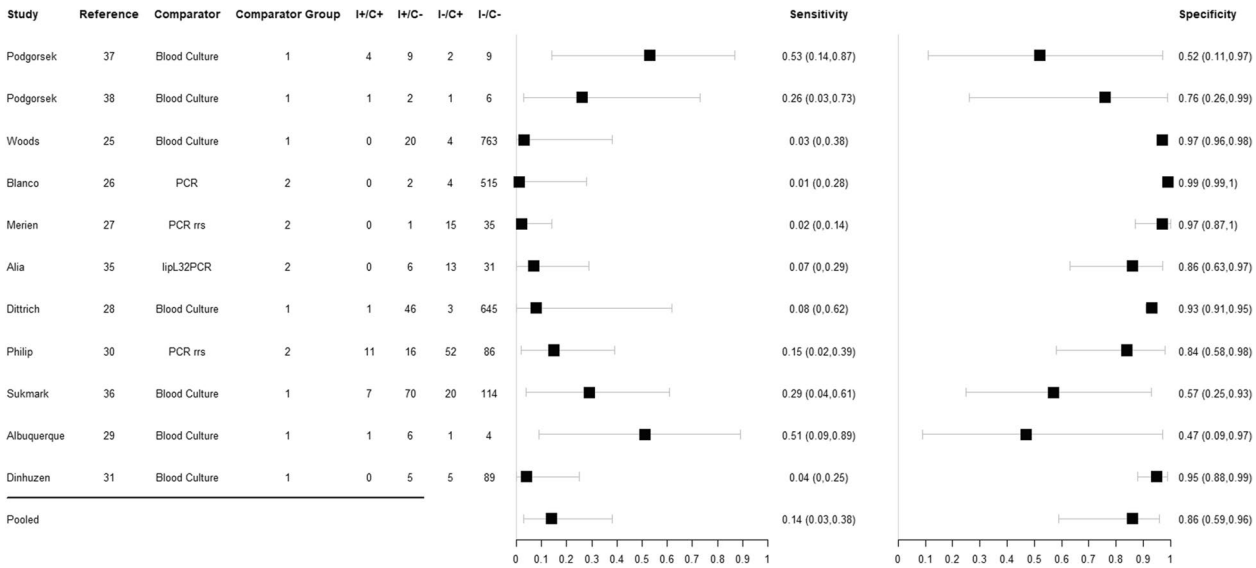


Fig. 2 Forest plot of estimated and pooled sensitivity and specificity of studies evaluating the diagnostic accuracy of MAT in single acute-phase samples, published global and between 1950–2022

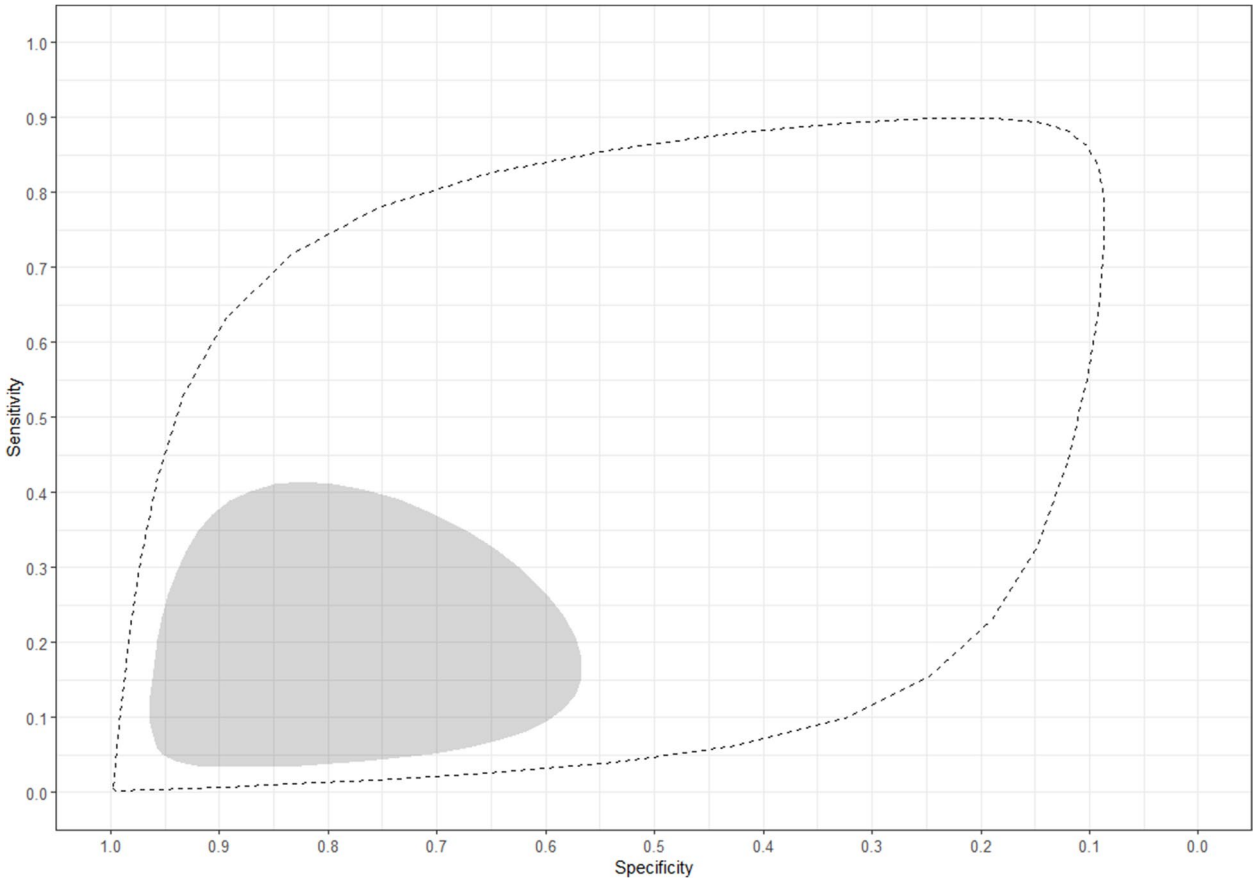


Fig. 3 Roc curve of pooled and predicted sensitivity and specificity of studies evaluating the diagnostic accuracy of MAT in single acute-phase samples, published global and between 1950–2022

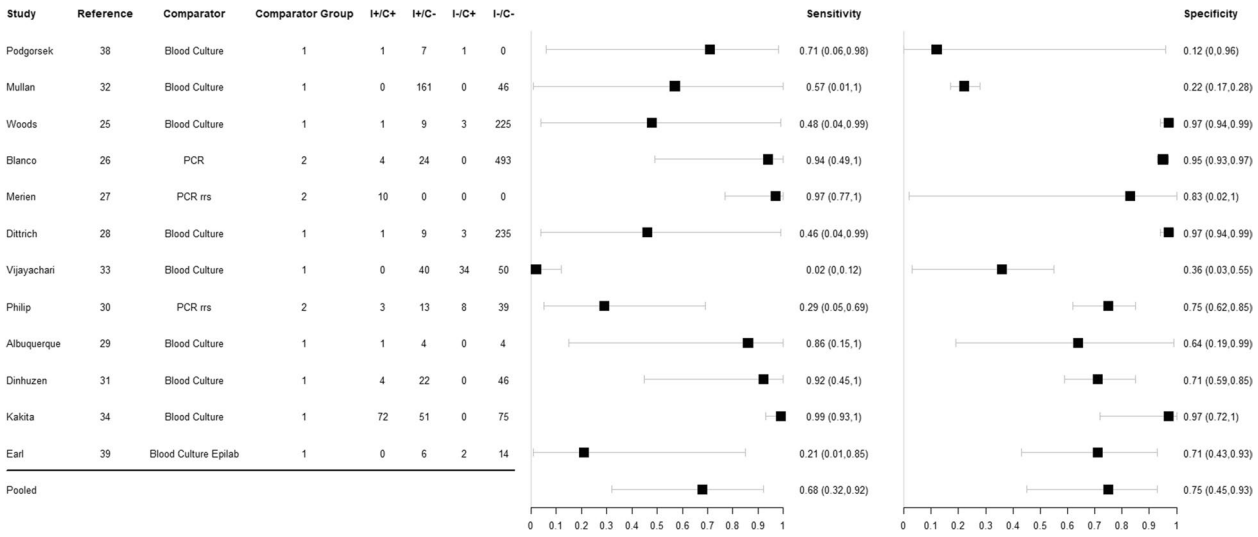


Fig. 4 Forest plot of estimated and pooled sensitivity and specificity of studies evaluating the diagnostic accuracy of MAT in paired samples, published global and between 1950–2022

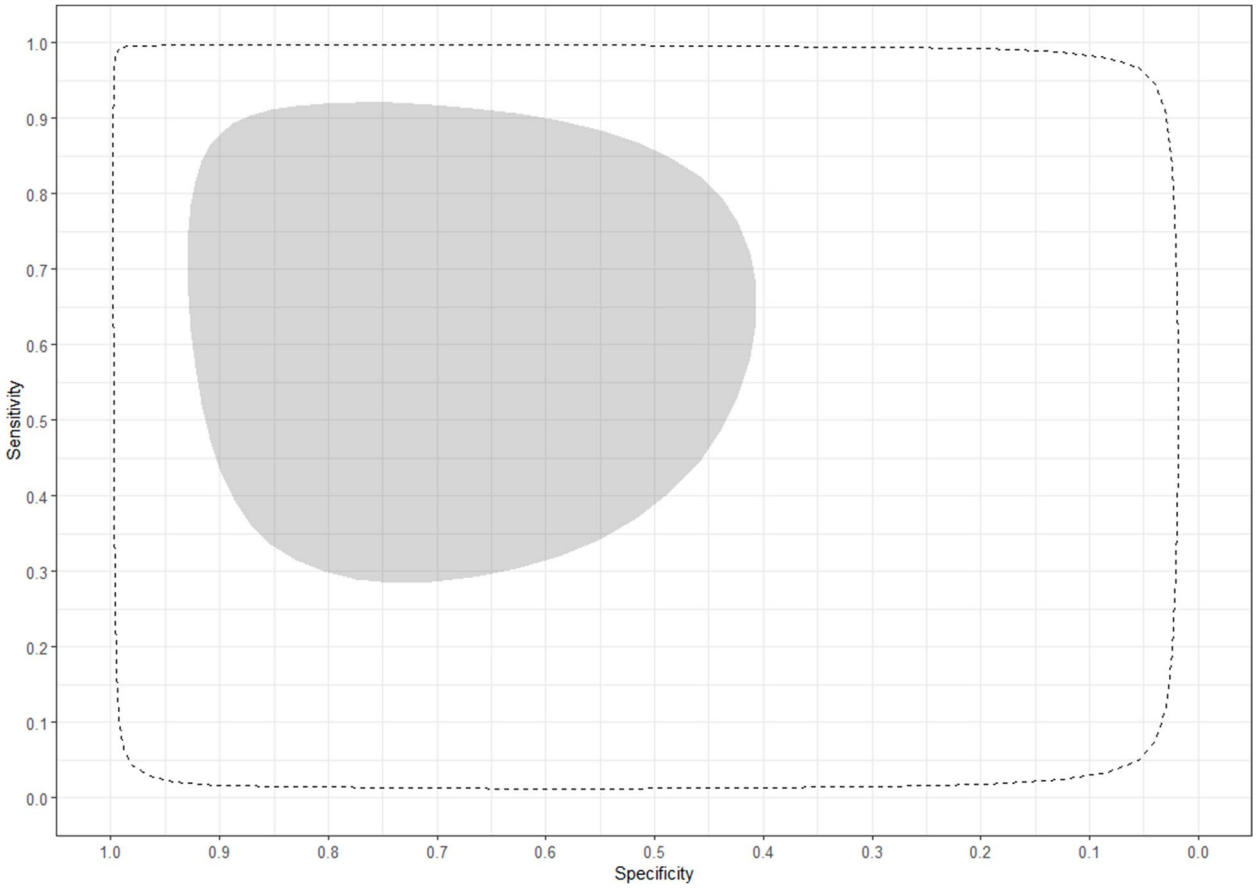


Fig. 5 Roc curve of pooled and predicted sensitivity and specificity of studies evaluating the diagnostic accuracy of MAT in paired samples, published global and between 1950–2022

Table 6 Extracted data, sensitivity and specificity estimates in the systematic review of studies evaluating the diagnostic accuracy of PCR and IgM ELISA, published global and between 1950 – 2022

Study first author, ref	Reference test	Total N samples	Index + / Reference +	Index + / Reference-	Index-/ Reference +	Index-/ Reference-	Sensitivity	Specificity
A. PCR studies								
Merien F [27]	MAT	51	10	15	1	25	92% (72–100)	66% (49–91)
Esteves L [40]	PCR	202	46	0	1	155	98% (90–100)	99% (98–100)
B. IgM ELISA studies								
Bourhy P [41]	MAT	519	298	3	19	199	97% (93–100%)	99% (97–100%)

Our estimates of the sensitivity of MAT in single acute-phase samples were low across all studies, but specificity was generally high. These findings are in line with the dynamics of the humoral immune response and with previous work from studies in a variety of countries including the Barbados [42], Netherlands [15], and Sri Lanka [43]. Moreover, numerous studies have shown the value of adding culture, nucleic acid amplification, or antigen detection to MAT serology during the early phase of the disease [44–50].

In paired samples we estimated to correctly identify just over two-thirds of true leptospirosis cases, and correctly reject the diagnosis for three-quarters of suspected cases. We found a more heterogeneous picture of estimated accuracy but our median estimates of 69% sensitivity and 75% specificity were also in line with previous findings in Barbados [42], Brazil [51], and Thailand [52]. Conversely, another study in Thailand [13], that also used a latent class model, estimated sensitivity to be lower than previous studies at 49.9%, with 95% CI from 37.6 to 60.8%. However, the authors stated that this could have been the result of convalescent-phase samples being collected only ten DPO of symptoms, allowing insufficient time for the antibody response to develop, and that 34% of participants did not have convalescent-phase serum specimens collected. Importantly, the estimate of MAT sensitivity in paired samples was 70.3% was consistent with our analysis.

Heterogeneity among studies is reflected in the wide credible intervals for the predicted sensitivity and specificity in this meta-analysis, particularly among the paired samples. The variability in estimates from single acute-phase samples could be explained by the heterogeneity of DPO of fever in the studies included, as shown by Goris et al. [12]. Single acute-phase samples may have been collected early in the illness, less than seven DPO of fever [11], too early in the humoral immune response for it to be a reliably detect infection. The high variability in the sensitivity of MAT in paired samples could be partially explained by the inclusion of patients with a brief interval, less than 14 days [11], between samples, and thus not reaching seroconversion or a four-fold rise or greater between titers [13]. It also could be attributed to failure to consider patients' use of antimicrobials before testing, particularly relevant when culture was used as a comparator test. It also could be due to MAT panel composition not representing the locally circulating strains [53–55].

Our meta-analysis had several limitations. Firstly, a key assumption of the Bayesian latent class model used is that there exist only two disease classes in the underlying population: diseased and disease-free. If in fact more than two classes exist, this assumption can result in biased estimates of test sensitivity and specificity when conditional independence between tests is assumed [56]. While the results presented in the main text of this paper do not make the assumption of conditional independence between tests, two disease classes are assumed. Further limitations include low geographical diversity, since included studies were conducted in only eight endemic countries, the majority in Southeast Asia, so that our estimates are not representative of all leptospirosis endemic countries. Moreover, our classification of a country's endemicity followed Costa, et al. [1], but these estimates are based on limited data and do not account for sub-national variation in leptospirosis incidence. Our bias assessment (Table 5) highlights the high risk of bias of all studies using single acute-phase samples as a confirmatory test for leptospirosis, and also that some studies do not describe or account for a globally or locally representative MAT panel, an important quality concern. Moreover, data on DPO of symptoms, the interval between paired samples, and the use of antimicrobials prior to testing were widely heterogeneous or unknown. This information was not included in the quality assessment but could be an important source for bias in some of our studies, interfering with the proportion of positive and negative tests results that correctly identify the infection status of individuals. Also, the low number of positive MAT results in the majority of selected studies compromised power. Another limitation was not finding studies that reported titres on acute and convalescent samples that would have allowed the direct evaluation of single sample MAT in the context of paired MAT. A final limitation was the difficulty in assessing QUADAS-2, due

to the lack of detailed data reported on the selected studies and due to the heterogeneity in MAT procedure and panel composition, since laboratories uses diverse antigen panels and every setting has different endemic local *Leptospira* serovars, sometimes unstated.

Our review also has many strengths. To our knowledge, this is the first meta-analysis of MAT accuracy for human leptospirosis diagnosis, and the first using Bayesian latent class modelling to account for the imperfect comparator tests. Our approach took into account different case definitions according to endemicity, and evaluated test results from single acute-phase samples separately from paired samples results. Importantly we used an extensive search strategy, contacted authors for additional data where necessary to complete a 2×2 table, and performed in duplicate and independently the process from study screening to data extraction.

Regarding our review of PCR targeting *lfb1* and ELISA IgM targeting antigen *Leptospira fainei* serovar Hurstbridge, due to the scarcity of data available, no meta-analysis could be performed. Instead, we report the estimated accuracy of each test within the included studies only. These results are not generalizable to other studies but suggest that both IgM ELISA and PCR had a high sensitivity in the included studies (median sensitivity: 92%, 98%, and 97%). Specificity varied in the two studies included for PCR (median specificity: 66% and 99%) and was high for IgM ELISA (99%). A 2017 systematic review of IgM ELISA for leptospirosis diagnosis not specifically targeting the antigen *Leptospira fainei* serovar Hurstbridge found similar results [57].

Conclusions

To our knowledge, this is the first meta-analysis estimating the accuracy of MAT in paired samples for diagnosis of human leptospirosis. Our study found that the sensitivity and specificity of MAT in paired samples were not high. However, MAT on paired sera remains the reference standard until a more accurate diagnostic strategy is developed. A key challenge for our review was the scarcity of high-quality studies driven by a low proportion of participants with paired serum samples, and a lack of detailed reporting of sample timing collection and panel composition. Future studies that use paired samples and that report in detail the sample timing collection and MAT panel composition will improve the certainty of accuracy estimates.

Abbreviations

CrI	Credible interval
DPO	Days post-onset
FIEBRE	Febrile Illness Evaluation in a Broad Range of Endemicities
MAT	Microscopic agglutination test
PCR	Polymerase chain reaction
PROSPERO	International Prospective Register of Systematic Reviews
QUADAS-2	Quality Assessment of Studies of Diagnostic Accuracy Approach- 2
SROC	Summary Receiver Operating Characteristic
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-023-08935-0.

Additional file 1: Appendix S1. Search strategy for the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022.

Additional file 2: Appendix S2. Statistical model the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022. **Table S1.** Sensitivity analysis in acute samples. **Table S2.** Sensitivity analysis in convalescent samples.

Additional file 3: Appendix S3. List of studies excluded dued to not having enough data available for a 2x2 contingent table for single-acute phase samples and/or paired samples.

Additional file 4: Table S3. Standardized extraction sheet form in the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950–2022.

Additional file 5: Table S4. Extracted data in the systematic review of studies evaluating the diagnostic accuracy of MAT, published global and between 1950–2022.

Acknowledgements

We thank Prof David Mabey at LSHTM for general support to this review.

Authors' contributions

MV, JB, HH and QB conceived the study. MV and JB assessed the eligibility of the studies, extracted the data, and assessed the methodological quality of the included studies. JF developed the search strategy and conducted the literature search. SK carried out the statistical analysis. JB, RK and OB advised on the statistical analysis. MV, JB and SK prepared the original draft of the manuscript, with considerable input from HH, QB, JAC, PN and MP. All authors contributed to the interpretations of results and all authors reviewed, edited and approved the final manuscript.

Funding

This review was conducted as part of the FIEBRE study, which is funded by UK aid from the UK government; the views expressed, however, do not necessarily reflect the UK government's official policies. MV is supported by Severo Ochoa Scholarship (Reference PRE2019-088336, Project CEX2018-000806-S-19-1). JB is supported by EDCTP Fellowship grant (EDCTP2 programme Project Code TMA2019SFP2836TF). SK is supported by the Medical Research Council London Intercollegiate Doctoral Training Partnership Studentship (MR/N013638/1). JAC received support from US National Institutes of Health grants R01TW009237 and R01 AI121378. The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Availability of data and materials

Code used for meta-analysis is publicly available at: https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/Leptospirosis. Data included in analyses can be found in Table S2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]ISGlobal, Hospital Clínic - Universitat de Barcelona, Calle Rosselló, 171, Entresol, Barcelona 08036, Spain. [2]Centro de Investigação em Saúde de Manhiça (CISM), Maputo, Mozambique. [3]London School of Hygiene & Tropical Medicine, London, UK. [4]ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain. [5]Pediatrics Department, Hospital Sant Joan de Déu, Universitat de Barcelona, Esplugues, Barcelona, Spain. [6]Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [7]MRC International Statistics and Epidemiology Group, London School of Hygiene and Tropical Medicine, London, UK. [8]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [9]Lao-Oxford-Mahosot Hospital Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Laos. [10]Biology of Spirochetes Unit, French National Reference Centre for Leptospirosis, Institut Pasteur, Université Paris Cité, Paris F-75015, France. [11]Centre for International Health, University of Otago, Dunedin, New Zealand.

Received: 18 July 2023 Accepted: 19 December 2023
Published online: 07 February 2024

References

1. Costa F, Hagan J, Calcagno J, Kane M, Torgerson PR, Matinez-Silveira MS, et al. Global morbidity and mortality of leptospirosis: a systematic review. PLoS Negl Trop Dis. 2015;9:9.
2. Taylor AJ, Paris DH, Newton PN. A systematic review of the mortality from untreated leptospirosis. PLoS Negl Trop Dis. 2015;9:6.
3. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, Lovett MA, Levett PN, Gilman RH, Willig MR, Gotuzzo E, Vinetz JM. Peru-United States Leptospirosis Consortium: leptospirosis: a zoonotic disease of global importance. Lancet Infect Dis. 2003;3:12.
4. Pappas G, Papadimitriou P, Siozopoulou V, Christou L, Akritidis N. The globalization of leptospirosis: worldwide incidence trends. Int J Infect Dis. 2008;12:4.
5. Hartskeerl RA, Collares-Pereira M, Ellis WA. Emergence, control an re-emerging leptospirosis: dynamics of infection in the changing world. Clin Microbiol Infect. 2011;17:4.
6. Prasad N, Murdoch DR, Reyburn H, Crump JA. Etiology of severe febrile illness in low- and middle-income countries: a systematic review. PLoS One. 2015;10:6.
7. Halliday J, Carugati M, Snavely M, Allan K, Beamesderfer J, Ladbury G, Hoyle D, Holland P, Crump JA, Cleaveland S, Rubach M. Zoonotic causes of febrile illness in malaria endemic countries: a systematic review. Lancet Infect Dis. 2020;20:2.
8. Mayxay M, Castonguay-Vanier J, Chansamouth V, Dubot-Pérès A, Paris DH, Phetsouvanh R, Tangkhabuanbutra J, Douangdala P, Inthalath S, Souvannasing P, Slesak G, Tongyoo N, Chanthongthip A, Panyanouvong P, Sibounheuang B, Phommasone K, Dohnt M, Phonekeo D, Hongvanthong B, Xayadeth S, Ketmayoon P, Blacksell SD, Moore CE, Craig SB, Burns MA, von Sonnenburg F, Corwin A, de Lamballerie X, González IJ, Christophel EM, Cawthorne A, Bell D, Newton PN. Causes of non-malarial fever in Laos: a prospective study. Lancet Glob Health. 2013;1:1.
9. Picardeau M, Bertherat E, Jancloes M, Skouloudis AN, Durski K, Hartskeerl RA. Rapid tests for diagnosis of leptospirosis: current tools and emerging technologies. Diagn Microbiol Infect Dis. 2014;78:1.
10. Terpstra WJ. Human leptospirosis: guidance for diagnosis, surveillance and control. World Health Organization; 2003. WHO Reference Number: WHO/CDS/CSR/EPH 2002.23. https://www.who.int/publications/i/item/human-leptospirosis-guidance-for-diagnosis-surveillance-and-control. ISBN: 9241545895.
11. Levett PN. Leptospirosis. In: Clinical microbiology reviews. 2001. p. 296–326.
12. Goris MG, Leeflang MM, Boer KR, Goeijenbier M, van Gorp EC, Wagenaar JF, et al. Establishment of valid laboratory case definition for human leptospirosis. J Bacteriol Parasitol. 2012;3:1000132.
13. Limmathurotsakul D, Turner EL, Wuthiekanun V, Thaipadungpanit J, Suputtamongkol Y, Chierakul W, et al. Fool's gold: why imperfect reference tests are undermining the evaluation of novel diagnostics: a reevaluation of 5 diagnostic tests for leptospirosis. Clin Infect Dis. 2012;55:322e31.
14. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol. 1988;41:9.
15. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics. 2012;68:4.
16. Hopkins H, Bassat Q, Chandler CI, FIEBRE Consortium, et al. Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE): protocol for a multisite prospective observational study of the causes of fever in Africa and Asia. BMJ Open. 2020;10:e035632.
17. Bourhy P, Bremont S, Zinini F, Giry C, Picardeau M. Comparison of real-time PCR assays for detection of pathogenic Leptospira spp. in blood and identification of variations in target sequences. J Clin Microbiol. 2011;49:6.
18. Naze F, Desvars A, Picardeau M, Bourhy P, Michault A. Use of a new high resolution melting method for genotyping pathogenic Leptospira spp. PLoS One. 2015;10:7.
19. Wolff JW. The laboratory diagnosis of leptospirosis. In: Thomas CC, editor. Illinois: Springfield; 1954.
20. Kohl C, McIntosh E, Unger S, Haddaway N, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid. 2018;7:8.
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:8.
22. Opota O, Croxatto A, Prod'hom G, Greub G. Blood culture-based diagnosis of bacteraemia: state of the art. Clin Microbiol Infect. 2015;21:4.
23. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57:1.
24. Stan Development Team. Stan modeling language users guide and reference manual, 2.28. 2021.
25. Woods K, Nic-Fhogartaigh C, Arnold C, Boutthasavong L, Phuklia W, Lim C, Chanthongthip A, Tulsiani SM, Craig SB, Burns MA, Weier SL, Davong V, Sihalath S, Limmathurotsakul D, Dance DAB, Shetty N, Zambon M, Newton PN, Dittrich S. A comparison of two molecular methods for diagnosing leptospirosis from three different sample types in patients presenting with fever in Laos. Clin Microbiol Infect. 2018;24:9.
26. Blanco RM, Romero EC. Evaluation of nested polymerase chain reaction for the early detection of Leptospira spp. DNA in serum samples from patients with leptospirosis. Diagn Microbiol Infect Dis. 2014;78:4.
27. Merien F, Portnoi D, Bourhy P, Charavay F, Berlioz-Arthaud A, Baranton G. A rapid and quantitative method for the detection of Leptospira species in human leptospirosis. FEMS Microbiol Lett. 2005;249:1.
28. Dittrich S, Boutthasavong L, Keokhamhoung D, Phuklia W, Craig SB, Tulsiani SM, Burns MA, Weier SL, Dance DAB, Davong V, Vongsouvath M, Mayxay M, Phetsouvanh R, Newton PN, Woods K. A prospective hospital study to evaluate the diagnostic accuracy of rapid diagnostic tests for the early detection of leptospirosis in Laos. Am J Trop Med Hyg. 2018;98:4.
29. Albuquerque AP, Araújo JG, Souza IQ, Martins LC, Oliveira MI, Silva MJ, Montarroyos UR, Miranda B. Validation of a case definition for leptospirosis diagnosis in patients with acute severe febrile disease admitted in reference hospitals at the State of Pernambuco, Brazil. Rev Soc Bras Med Trop. 2011;44:6.
30. Philip N, Affendy NB, Masri SN, Yuhana MY, Than LTL, Sekawi Z, et al. Combined PCR and MAT improves the early diagnosis of the biphasic illness leptospirosis. PLoS One. 2020;15:9.
31. Dinhuzen J, Limothai U, Tachaboon S, Krairojananan P, Laosatiankit B, Boonprasong S, et al. A prospective study to evaluate the accuracy of rapid diagnostic tests for diagnosis of human leptospirosis: result from THAI-LEPTO AKI study. PLoS Negl Trop Dis. 2021;15:2.
32. Mullan S, Panwala TH. Polymerase chain reaction: an important tool for early diagnosis of leptospirosis cases. J Clin Diagn Res. 2016;10:12.
33. Vijayachari P, Sugunan AP, Sehgal SC. Evaluation of Lepto Dri Dot as a rapid test for the diagnosis of leptospirosis. Epidemiol Infect. 2002;129:3.
34. Kakita T, Okano S, Kyan H, Miyahira M, Taira K, Kitashoji E, et al. Laboratory diagnostic, epidemiological, and clinical characteristics of human leptospirosis in Okinawa Prefecture, Japan, 2003–2020. PLoS Negl Trop Dis. 2021;15:12.

35. Alia SN, Joseph N, Philip N, Azhari NN, Garba B, Masri SN, Sekawi Z, Neela VK. Diagnostic accuracy of rapid diagnostic tests for the early detection of leptospirosis. J Infect Public Health. 2019;12:2.
36. Sukmark T, Lumlertgul N, Peerapornratana S, Khositrangsikun K, Tungsanga K, Sitprija V, et al. Thai-Lepto-on-admission probability (THAI-LEPTO) score as an early tool for initial diagnosis of leptospirosis: result from Thai-Lepto AKI study group. PLoS Negl Trop Dis. 2018;12:3.
37. Podgoršek D, Ružić-Sabljić E, Logar M, Pavlović A, Remec T, Baklan Z, Pal E, Cerar T. Evaluation of real-time PCR targeting the lipL32 gene for diagnosis of Leptospira infection. BMC Microbiol. 2020;20:1.
38. Podgoršek D, Cerar T, Logar M, Lešničar G, Remec T, Baklan Z, Pal E, Ružić-Sabljić E. Evaluation of the immunochromatographic (Leptocheck) test for detection of specific antibodies against leptospires. Wien Klin Wochenschr. 2015;127:23–4.
39. Earl L, Fang F, Janes R, Gedye K, French N, Collins-Emerson J, Benschop J. An evaluation of diagnostic tests in a case series of suspected leptospirosis patients seen in primary care. N Z Med J. 2021;134:1539.
40. Esteves LM, Bulhões SM, Branco CC, Carreira T, Vieira ML, Gomes-Solecki M, Mota-Vieira L. Diagnosis of human leptospirosis in a clinical setting: real-time PCR high resolution melting analysis for detection of Leptospira at the onset of disease. Sci Rep. 2018;8:1.
41. Bourhy P, Vray M, Picardeau M. Evaluation of an in-house ELISA using the intermediate species Leptospira fainei for diagnosis of leptospirosis. J Med Microbiol. 2013;62:6.
42. Cumberland P, Everard COR, Levett PN. Assessment of the efficacy of an IgM-ELISA and microcopic agglutination test (MAT) in the diagnosis of acute leptospirosis. Am J Trop Med Hyg. 1999;61:5.
43. Agampodi SB, Dahanayaka NJ, Nöckler K, Anne M, Vinetz JM. Redefining gold standard testing for diagnosing leptospirosis: further evidence from a well-characterized, flood-related outbreak in Sri Lanka. Am Soc Trop Med Hyg. 2016;95:3.
44. Camargo ED, Emilson D, Branda AP, Silva MV, Abra RUIV. Macroscopic agglutination test for rapid diagnosis of human leptospirosis. J Clin Microbiol. 1998;36:11.
45. Appassakij H, Silpapojakul K, Wansit R, Woodtayakorn J. Evaluation of the immunofluorescent antibody test for the diagnosis of human leptospirosis. Am J Trop Med Hyg. 1995;52:4.
46. Cassadou S, Rosine J, Flamand C, Escher M. Underestimation of leptospirosis incidence in the French West Indies. PLoS Negl Trop Dis. 2016;10:4.
47. Niloofa R, Fernando N, de Silva NL, Karunanayake L, Wickramasinghe H, Dikmadugoda N, Premawansa G, Wickramasinghe R, de Silva HJ, Premawansa S, Rajapakse S, Handunnetti S. Diagnosis of leptospirosis: comparison between microscopic agglutination test, IgM-ELISA and IgM rapid immunochromatography test. PLoS One. 2015;10:6.
48. Ooteman MC, Vago AR, Koury MC. Evaluation of MAT, IgM ELISA and PCR methods for the diagnosis of human leptospirosis. J Microbiol Methods. 2006;65:2.
49. Spinosa C, Fonseca CDA, Lu V. Polymerase chain reaction in comparison with serological tests for early diagnosis of human leptospirosis. Trop Med Int Health. 2006;11:11.
50. Waggoner JJ, Balassiano I, Mohamed-hadley A. Reverse-transcriptase PCR detection of Leptospira: absence of agreement with single-specimen microscopic agglutination testing. PLoS One. 2015;10:7.
51. De Abreu Fonseca C, Teixeira De Freitas VL, Caló Romero E, Spinosa C, Arroyo Sanches MC, Da Silva MV, Shikanai-Yasuda MA. Polymerase chain reaction in comparison with serological tests for early diagnosis of human leptospirosis. Trop Med Int Health. 2006;11:11.
52. Thaipadunpanit J, Chierakul W, Wuthiekanun V, Limmathurotsakul D, Amornchai P, Boonslip S, Smythe LD, Limpaiboon R, Hoffmaster AR, Day NPJ, Peacock SJ. Diagnostic accuracy of real-time PCR assays targeting 16S rRNA and lipl32 genes for human leptospirosis in Thailand: a case-control study. PLoS One. 2011;6:1.
53. Scheer CS, Fuchs C, Gründling M, Vollmer M, Bast J, Bohnert JA, Zimmermann K, Hahnenkamp K, Rehberg S, Kuhn SO. Impact of antibiotic administration on blood culture positivity at the beginning of sepsis: a prospective clinical cohort study. Clin Microbiol Infect. 2019;25:3.
54. Rhodes J, Hyder JA, Peruski LF, Fisher C, Jorakate P, Kaewpan A, Dejsirilert S, Thamthitiwat S, Olsen SJ, Dowell SF, Chantra S, Tanwisaid K, Maloney SA, Baggett HC. Antibiotic use in Thailand: quantifying impact on blood culture yield and estimates of pneumococcal bacteremia incidence. Am Soc Trop Med Hyg. 2010;83:2.
55. Driscoll AJ, Deloria Knoll M, Hammitt LL, Baggett HC, Brooks WA, Feikin DR, Kotloff KL, Levine OS, Madhi SA, O'Brien KL, Scott JAG, Thea DM, Howie SRC, Adrian PV, Ahmed D, DeLuca AN, Ebruke BE, Gitahi C, Higdon MM, Kaewpan A, Karani A, Karron RA, Mazumder R, McLellan J, Moore DP, Mwananyanda L, Park DE, Prosperi C, Rhodes J, Saifullah M, Seidenberg P, Sow SO, Tamboura B, Zeger SL, Murdoch DR, PERCH Study Group. The effect of antibiotic exposure and specimen volume on the detection of bacterial pathogens in children with pneumonia. Clin Infect Dis. 2017;64:3.
56. Schofield MR, Maze MJ, Crump JA, Rubach MP, Galloway R, Sharples KJ. On the robustness of latent class models for diagnostic testing with no gold standard. Stat Med. 2021;40:22.
57. Rosa MI, dos Reis ME, Simon CS, Dondossola ER, Alexandre MC, Meller TC. IgM ELISA for leptospirosis diagnosis: a systematic review and meta-analysis. Cienc Saude Colet. 2017;22:12.

Publisher's Note

Appendix S2: Statistical model the systematic review of studies evaluating the diagnostic accuracy of MAT, PCR, and IgM ELISA, published global and between 1950 – 2022.

The complete model to estimate the sensitivity and specificity of the index test in each study, and hence predict sensitivity and specificity in a future study, consists of two hierarchical levels. The first level captures between-study variations in the sensitivity and specificity through a Hierarchical Summary Receiver Operating Characteristic (HSROC) model (Rutter and Gatsonis, 2001; Dendukuri et al 2012; Dendukuri and Joseph, 2001), which assumes that sensitivity and specificity across studies lie on an ROC curve. The second level accounts for the unknown true disease status of all participants using latent class analysis.

In this analysis of the MAT diagnostic test, a separate meta-analysis was run for acute and convalescent samples. In the main text we present a model relaxing the assumption of conditional independence between test results from the same individual given their disease class. In the following, we first outline this model before presenting the results of a sensitivity analysis. We compared the result relaxing the assumption of conditional independence to those assuming conditional independence as well as presenting the results for a model where only studies from endemic settings are included. This is in comparison to the results presented in the main text where studies from both endemic and non-endemic settings are included.

Starting with the latent class model, assume we have a sample of N individuals who have all undergone two different dichotomous tests defined by T_r $(r = 1,2)$, and let t_{ri} be a random variable denoting the outcome from test r for individual $i, i = 1, .., N$. A positive test result for an individual is denoted by $t_{ri} = 1$, and a negative test result by $t_{ri} = 0$. Also assume that the unknown true disease status of an individual, denoted D, can take one of two values: 'diseased' $(D = 1)$ or 'non-diseased' $(D = 0)$.The true (latent) disease status of the i^{th} individual is denoted d_i . Assuming conditional independence between the two tests, we can write the joint distribution as:

$$\Pr(\mathrm{T}_1 = t_1, \mathrm{T}_2 = t_2) = \qquad 1$$

$$\Pr(\mathrm{T}_1|\mathrm{D}=1)\Pr(\mathrm{T}_2|\mathrm{D}=1)\Pr(\mathrm{D}=1)+\Pr(\mathrm{T}_1|\mathrm{D}=0)\Pr(\mathrm{T}_2|\mathrm{D}=0)\Pr(\mathrm{D}=0)$$

For the r^{th} test, the sensitivity (Se) and specificity (Sp) can be written as $\mathrm{Se}_\mathrm{r}=\Pr(\mathrm{T}_\mathrm{r}=1|\mathrm{D}=1)$ and $\mathrm{Sp}_\mathrm{r}=\Pr(\mathrm{T}_\mathrm{r}=0|\mathrm{D}=0)$ respectively. We can also define prevalence as $\pi=\Pr(D=1)$. We can then specify equation 1 in terms of sensitivity, specificity and prevalence:

$$\Pr(\mathrm{T}_1=t_1,\mathrm{T}_2=t_2)= \mathrm{Se}_1\mathrm{Se}_2(1-\mathrm{Se}_1)(1-\mathrm{Se}_2)\pi+(1-\mathrm{Sp}_1)(1-\mathrm{Sp}_2)\mathrm{Sp}_1\mathrm{Sp}_2(1-\pi) \quad (2)$$

Following this, we can then define the likelihood for this latent class model assuming conditional independence as:

$$\mathrm{L}=\prod_{\mathrm{i}=1}^{\mathrm{N}}(\pi{\mathrm{Se}_1}^{\mathrm{t}_{1\mathrm{i}}}{\mathrm{Se}_2}^{\mathrm{t}_{2\mathrm{i}}}(1-\mathrm{Se}_1)^{1-\mathrm{t}_{1\mathrm{i}}}(1-\mathrm{Se}_2)^{1-\mathrm{t}_{2\mathrm{i}}})+((1-\pi){\mathrm{Sp}_1}^{1-\mathrm{t}_{1\mathrm{i}}}{\mathrm{Sp}_2}^{1-\mathrm{t}_{2\mathrm{i}}}(1-\mathrm{Sp}_1)^{\mathrm{t}_{1\mathrm{i}}}(1-\mathrm{Sp}_2)^{\mathrm{t}_{2\mathrm{i}}}) \quad (3)$$

To relax the assumption of conditional independence and account for conditional dependence between tests, we allow the sensitivity from each test to depend on an individual level random effect s_i. We assume that the sensitivity in the conditionally dependent model takes the form (1):

$$\Pr(\mathrm{T}_{\mathrm{ri}}=1|\,\mathrm{D}_\mathrm{i}=1,)=\mathrm{Se}_{\mathrm{ri}}=\mathrm{g}^{-1}(\mathrm{a}_{\mathrm{rd}=1}+\mathrm{b}_{\mathrm{rd}=1}\mathrm{s}_\mathrm{i}) \quad (4)$$

where $g(\cdot)$ is a link function. In this study we use the logit link, $g^{-1}(y)=1/(1+e^{-y})$, $a_{rd=1}$ and $b_{rd=1}$ are unknown parameters to be estimated where b describes the strength of dependence between two tests and the random effect s_i follows a standard normal distribution. The subject-specific random-effect s_i represents some unobserved characteristic for example infection intensity, that indirectly creates dependence between tests.

The study-level latent class model (equation 3 and 4) is then linked to the between-study level using a HSROC model (equation 5) which models between study variations by assuming that test sensitivity and specificity lie on an ROC curve. In particular, each study j $(j=1,..,J)$ provides the 2x2 table between the test of interest, hereafter called the index test, which is the same in all studies, and a comparator test which may differ between studies. We let T_{1j} denote the index test outcomes in each study and we let T_{2j} denote the

comparator test outcomes in each study. In line with previous descriptions (2,3), we define the sensitivity and specificity of the index test in the jth study by:

$$\begin{aligned}\text{logit}(\text{Se}_{1\text{j}}) &= \Pr(T_{1j} = 1|\text{D} = 1) \ = -(\theta_{\text{j}} \ - \ \alpha_{\text{j}}/2)/\exp(\beta/2) \\ \text{logit}(\text{Sp}_{1\text{j}}) &= \Pr(T_{1j} = 0|\text{D} = 0) \ = (\theta_{\text{j}} \ + \ \alpha_{\text{j}}/2)/\exp(-\beta/2)\end{aligned} \quad 5$$

where θ_j represents the positivity criteria for study j. The positivity criteria, or cut-off value, models the dependence between the true positive fraction and false positive fraction in each study. α_j represents the diagnostic accuracy and measures the mean difference in test accuracy between individuals 'diseased' and individuals 'non-diseased' in study j. β the scale parameter, allows differences in the variation of outcomes between disease positive and disease negative individuals thus allowing asymmetry in the ROC curve. Both θ and α parameters are modelled as random effects with independent normal distributions to incorporate variation between studies:

$$\begin{aligned}\theta_{\text{j}} &\sim \ \text{N}(\Theta, \sigma_\theta) \\ \alpha_{\text{j}} &\sim \ \text{N}(\Lambda, \sigma_\alpha)\end{aligned} \quad 6$$

We present pooled estimates of sensitivity and specificity which are given by:

$$\begin{aligned}\text{logit (Pooled Se)} &= -\big((\Theta - \Lambda/2)/\exp(\beta/2)\big) \\ \text{logit (Pooled Sp)} &= ((\Theta + \Lambda/2)/\exp(-\beta/2))\end{aligned} \quad 8$$

Importantly, we also present a prediction of sensitivity and specificity in a new study. Predicted estimates are important because the pooled estimates in a meta-analysis only represent an average estimate among the studies included in the analysis. Predicted estimates on the other hand account for the variation captured through the modelling framework and can be used as priors for sensitivity and specificity of the index test in a new study. We predict sensitivity and specificity by replacing Θ with $\theta_{newstudy}$ and Λ with $\alpha_{newstudy}$:

$$\begin{aligned}\theta_{newstudy} &\sim N(\Theta, \sigma_\theta) \\ \alpha_{newstudy} &\sim N(\Lambda, \sigma_\alpha)\end{aligned} \quad 10$$

The above description was applied to both acute and convalescent samples independently.

X.2 Prior Specification

We use the following priors: Θ~(0,1), Λ~N(0,2) and, β~$Uniform$(−0.75,0.75). Variance parameters $\sigma_\theta \; and \; \sigma_\alpha$ follow zero-truncated standard normal distributions. We allow prevalence and specificity of the index test in each study to be uniform over 0 to 1. The sensitivity of the index test in each study are also assumed uniform between a lower limit of 1 minus the specificity. This ensures that the probability of a positive test is higher for somebody with disease than without.

In both analyses, the comparator tests are Culture and PCR. For culture, the sensitivity is assumed uniform between 0 and 1 however the specificity is assumed to follow a $beta(50,1)$ distribution. Assuming a $beta(50,1)$ distribution corresponds to an assumption of 95% probability of the specificity being above 94%. We assume the sensitivity and specificity of PCR follow a $beta(5,1)$ distribution. This corresponds to an assumption of 95% probability that the sensitivity and specificity of PCR are above 50%.

X.3 Sensitivity Analysis

Sensitivity analyses were conducted to assess the robustness of our results to assumptions about conditional dependence between tests and to the endemicity level of countries. In the main text of the article a model relaxing the assumption of conditional independence between diagnostic tests is presented. Here we present the results assuming conditional independence, comparing to those presented in the main text. We also present the results when only those studies from endemic studies are included. Table 1 highlights that the pooled and predicted estimates in acute samples did not differ greatly between models. Table 2 shows that in convalescent samples only including endemic countries slightly increased the median estimate of pooled and predicted sensitivity and specificities but that there is little evidence of conditional dependence between tests as results comparing whether conditional independence is assumed or relaxed are similar.

Table S1: Sensitivity analysis in acute samples

Study countries included	Conditional independence (CI) or	Pooled sensitivity	Pooled specificity	Predicted sensitivity	Predicted specificity

	conditional dependence (CD)	median (95% CrI)	median (95% CrI)	median (95% CrI)	median (95% CrI)
All	CI	17(5-41)	88(66-98)	17(1-89)	89(14-100)
All*	CD	14(3-38)	86(59-96)	14(0-90)	86(9-100)
Endemic only	CI	13(3-37)	90(67-98)	13(0-86)	90(14-100)
Endemic only	CD	11(2-33)	87(60-97)	10(0-86)	88(10-100)

*This row represents the results presented in the main text of the manuscript

Table S2: Sensitivity analysis in convalescent samples

Study countries included	Conditional independence (CI) or conditional dependence (CD)	Pooled sensitivity median (95% CrI)	Pooled specificity median (95% CrI)	Predicted sensitivity median (95% CrI)	Predicted specificity median (95% CrI)
All	CI	70(32-92)	74(43-93)	70(2-100)	76(2-100)
All*	CD	68(32-92)	75(45-93)	69(2-100)	75(2-100)
Endemic only	CI	74(38-95)	81(51-95)	76(2-100)	81(4-100)
Endemic only	CD	73(34-95)	81(48-95)	73(2-100)	81(3-100)

*This row represents the results presented in the main text of the manuscript

References:

1. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57(1):158-67.
2. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20(19):2865-84.
3. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics. 2012;68(4):1285-93.

5.4 Paper contributions

The paper presented in section 5.3 describes the results of a diagnostic test accuracy systematic review and meta-analysis for the diagnosis of leptospirosis following the most up to date advice on best practices from the 2023 Cochrane handbook [86]. Further, the code for all analyses is available open access. The paper sought to estimate the accuracy of three tests of interest however very few studies were found to estimate the accuracy of two of the three tests of interest (PCR and IgM ELISA) so, a meta-analysis was only carried out for the microscopic agglutination test (MAT).

Guidelines for the proper use of the MAT test requires the testing of human samples from two time points, commonly day 0 and day 28 [130]. However, from this review, it is evident that what happens in practice is not always what the guidelines stipulate. The systematic review found many studies comparing the result of MAT from day 0 only with another test. As a consequence, we decided to change our original analysis plan and estimate the accuracy of MAT on single acute phase samples as well as estimating the accuracy when the test is performed as per the guidelines. This ad hoc analysis was performed because it is important for researchers to understand the accuracy of a diagnostic test being used, especially if there is a systematic difference in how the test is being used from which it was originally designed for.

Unsurprisingly, the estimated accuracy of MAT on acute phase samples only had a poor sensitivity 14% (95% credible interval(CrI) 3-38%). Despite a poor sensitivity, specificity remains high on the first test (85%, 95% CrI: 59-96%). While it is not advised to use MAT on acute phase samples only, this review presents the first estimates of the accuracy of this test when it is used in this way for the diagnosis of leptospirosis. On the other hand, MAT on paired samples is recognised as the gold-standard in the diagnosis of leptospirosis however, the results found here suggest this test is not perfect with an estimated partially-pooled sensitivity of 68% (95% CrI: 32-92%) and specificity of 75% (95% CrI: 45-93%). This is not a novel finding [132] but, despite this, MAT continues to be treated as a gold-standard in leptospirosis diagnosis [133, 134]. Recognition that MAT is imperfect is important if a diagnostic test with a higher sensitivity is to be found and utilised.

The FIEBRE case-control study found similar limitations with the MAT test. A substantial proportion of participants did not have a day 28 sample and so the MAT test was only carried out on single acute phase samples. Quantifying the accuracy of the first test only meant that for

individuals in FIEBRE missing data on day 28 we could still utilise their day 0 sample MAT result as we can generate a prior distribution from the results of the meta-analysis presented here.

Part II

Fever Aetiology Estimation

Chapter 6

Disease aetiology research

6.1 Preamble

Part 1 of this thesis focused on diagnostic test accuracy estimation in the absence of perfect reference tests. The predicted sensitivity and specificity of diagnostic tests of interest were estimated via Bayesian latent class meta-analysis. Part 2 uses the estimates of diagnostic test accuracy as priors in a Bayesian analysis of a case-control study to investigate the aetiology of fever, applied to data from the FIEBRE study.

In this first chapter of Part 2, I discuss what aetiology research is, why understanding fever aetiology is important and the study designs that can be used in aetiology research. Two examples of aetiology studies, that differ by syndrome of interest and primary statistical analysis approach and were influential in the design of the FIEBRE study, are outlined.

6.2 Aetiology research

Aetiology is derived from the Greek word *aitiologia* which means giving a reason for. In this vein, aetiology in epidemiological research is the identification of determinants (reasons) involved in the cause, risk or development of disease, conditions and ill health [135]. This could include the identification of endogenous factors, environmental or external factors or behaviours and lifestyle factors. Ultimately, the aim of these studies, is to understand pathways that lead to an outcome so that we may intervene on these pathways [136]. Thus, in aetiology research, we are interested in causation.

A classic epidemiological example of an aetiology study is that of lung cancer. Smoking was hypothesised and found to be a key exposure related to having lung cancer [137]. As a result of multiple studies indicating such an association [138], interventions in the form of adver-

tisement campaigns were launched in an attempt to reduce smoking and the result was a global reduction in lung cancer cases and associated mortality [139]. Another example is an aetiology study investigating diarrhoeal disease where it was hypothesised that measles was as a key cause. Following further investigation, measles vaccination was rolled out and incident cases and deaths from diarrhoea declined [140]. Both of these examples highlight the cornerstone of aetiology studies, that is, to identify an exposure (smoking and measles) that is believed to be an important contributor to a disease or syndrome (lung cancer and diarrhoeal disease) and then introduce an intervention or policy (advertisement campaigns and vaccination) that can reduce or eliminate cases of the disease or syndrome that were due to the exposure.

In this thesis, I am interested in aetiology studies of a case-control study design for syndromes where the primary aim is to estimate how much of a syndrome can be attributed to specific exposures. In the fever context, fever is the syndrome/outcome of interest and the exposures considered are infections caused by different pathogens. Aetiology studies for syndromes with a large morbidity and mortality burden are particularly important as identifying a key contributor may lead to large reductions in morbidity and mortality. A syndrome is a set of recognisable symptoms for which the direct cause is not always understood [141]. Examples of syndromes with significant morbidity and mortality burdens are fever, diarrhoea and pneumonia [4]. In each of these examples, it is assumed that elucidation of pathogens that contribute the largest share of the burden can reduce morbidity and mortality by leading to better access to treatments and/or vaccines or to the development of treatments, vaccines or better diagnostics for these pathogens. The example above of measles vaccination to reduce diarrhoea is one such example. Multi-site case-control aetiology studies have been carried out for both diarrhoea [142] and pneumonia [38] but were lacking for fever until the inception of the FIEBRE study [20].

While the focus in this thesis is on case-control studies, other observational study designs could also be used in aetiology research for example, an early example of an aetiology study using a cohort design is of liver cancer [143]. In a cohort study, individuals are recruited based on exposure status and followed up over time to establish which individuals develop the outcome of interest. A key limitation of this study design in comparison to case-control studies is the time period required to allow the outcome to manifest which can be prohibitively costly [144] and the resources required to follow-up recruited participants. In the exam-

ple of common infections, discussed further in section 6.3, the addition of controls is also necessary to distinguish between pathogens which are common in the whole population and those pathogens that are important in only the cases with the outcome of interest.

6.2.1 Measures of disease association in case-control studies

In this thesis, I am interested in case-control aetiology studies where the goal is to estimate the association between an outcome and possible exposures. Specifically, I am interested in quantifying the association between fever (the outcome) and fever-causing pathogens (exposures). In outlining some of the possible measures of disease association I will use the example introduced briefly above of Doll and Hill investigating smoking as a cause of lung-cancer.

In the study of lung cancer in London by Doll and Hill [137], they recruited incident cases of lung cancer between April 1948 and October 1949 that attended any of the twenty included hospitals from the greater London area. While cases were being interviewed about their smoking history (exposure), controls who also attended the same hospital but did not have lung cancer and were similar to cases on sex and age were also recruited. Exposure status was collected in detail by number of cigarettes smoked per day but here we focus on the simple binary exposure of individuals who have ever been a smoker and individuals who have never smoked. The study population can be described by the 2x2 table shown in Table 6.1a with the actual numbers of cases and controls recruited in the study by exposure status shown in Table 6.1b.

Table 6.1: Example of a 2x2 table for a case-control study with a binary exposure

(a) Dummy table for reporting the results of a case-control study with a binary exposure

	Exposure	
	Exposed	Not-exposed
Cases	A	B
Controls	C	D

(b) 2x2 table for the Doll and Hill case-control study [137] with exposure treated as a binary variable

	Exposure	
	Smoker	Never smoked
Cases	688 (97%)	21
Controls	650 (91%)	59

Table 6.1b shows that, ignoring sampling variation, a higher proportion of cases (97%) than controls (91%) had been a smoker. It follows, that smoking is more common among those with lung cancer.

From Table 6.1b, we can estimate the odds that a case has ever been a smoker (688/21)), also referred to as the odds of exposure among the cases (i.e.Pr($E|D$)) and the odds that a control has ever been a smoker (i.e. Pr($\bar{E}|D$)), or the odds of exposure among the controls (650/59). From these two quantities we can then estimate the odds ratio or, more specifically, the exposure odds ratio as:

$$OR = \frac{A/B}{C/D} \tag{6.2.1}$$

where, A,B,C,D are taken from the 2x2 table in Table 6.1a. However, the quantity of most interest is not the probability of being exposed given you have disease (i.e. Pr($E|D$)) but the probability of being diseased given that you are exposed (Pr($D|E$)). The ratio of the odds of disease in the exposed to the odds among those not exposed is called the risk ratio (RR) or the relative risk and the relationship between these two quantities (the odds ratio and the risk ratio) is the cornerstone of the analysis of a case-control study. Assuming the controls are sampled from the source population of the cases, Cornfield showed that the ratio of the odds of a binary outcome given exposed to that in the unexposed is the same as the ratio of the odds where the roles of outcome and exposure are reversed, if the disease being studied is rare [145]. However, while case-control studies are particularly efficient for investigating rare diseases as cases are selected on outcome status, they are also useful for common diseases and it has subsequently been shown that the odds ratio also approximates the risk ratio for common diseases when cases and controls are recruited concurrently [146, 147].

From Doll and Hill's case-control study, the odds ratio is 2.97 ((688/21)/(650/59)) which, suggests that odds of lung cancer are almost three times greater in smokers than non-smokers. From Doll and Hill's findings they concluded that within the population from which their cases arose, smoking was likely a key cause of lung cancer. Central to this conclusion, is the assumption that the exposure distribution in controls is similar to that in the population from which cases arose. This highlights the importance of the selection of controls in case-control studies. In fact, if the control group is not representative of the population from which the cases have arisen, then any results found are not applicable beyond the particular group studied [145].

Another measure of disease association of interest in aetiology studies

is what has been termed the population attributable fraction (PAF) [138]. Defined as the fraction of all cases of a particular outcome in a population that is attributable to a specific exposure or, the fraction of cases of a particular outcome that would theoretically be prevented if an exposure were eliminated [148]. The population attributable fraction has also been referred to as the attributable risk [138], the population attributable risk percent (if multiplied by 100) [149], the aetiologic fraction [150] or more recently the cause-specific case fraction [151]. The PAF was defined by Miettinen as [146]:

$$PAF = \eta(1 - 1/RR) \tag{6.2.2}$$

where RR is the risk ratio, approximated by the odds ratio in a case control design, describing the odds of the outcome in those exposed to the odds among those not exposed and η is the proportion of cases among which the risk factor is observed. This differs from the original definition of Levin [138] where the overall prevalence in the population is required. From equation 6.2.2 it is clear that the PAF depends fundamentally on the odds associated with the exposure and the prevalence of the exposure in the cases. Up to this point, I have only discussed the odds between two groups (exposed and not exposed) however, it is more likely that the study population has been divided into I strata based on age, sex and other factors that may confound the association between the exposure and outcome. Not taking into account these factors in 6.2.2 would give a biased estimate of the PAF. However, plugging in an adjusted odds ratio (adjusted for example by confounding variables such as age and sex by conditional logistic regression or standardisation) in equation 6.2.2 allows estimation of an adjusted PAF which can remove confounder bias [152].

Using our illustrative example from the Doll and Hill seminal work, the estimated odds ratio is 2.97 and the prevalence of the exposure in the cases is 0.97 (97%). Thus, the PAF would be 64% $(0.97(1 - 1/2.97))$ which can be interpreted as 64% of lung cancer cases are attributed to smoking within the population studied (greater London). While using the terminology 'attributed' is inline with the original definition of the PAF, unless all confounding variables are known to be accounted for, it may still be more appropriate to refer to associations or relationships between exposures and an outcome rather than causation [146].

Case-control studies were traditionally the study design of choice for rare diseases however, they have now been used in the study of common infections including diarrhoea and respiratory illness. With a common disease, the case-control design is no longer clearly more efficient than the cohort study in regards to sample size [153] but it does still avoid

the ethical issues inherent in prospective cohort and intervention studies, since the disease status of participants is already determined. In addition, a case-control study may be simpler and quicker to conduct if it can, for example, be conducted entirely within a health facility. For these reasons, along with the observation that the odds ratio is invariant under study design (cohort or case-control) are reasons why case-control studies have and are still used to study common diseases like diarrhoea and acute respiratory illness.

6.2.2 Fever

Fever is a common symptom of many infections [1] and is a associated with substantial morbidity and mortality [4]. A fever can be caused by many different pathogens and common causes of fever vary considerably across space [6, 7], time [8, 9] and patient characteristics such as age [10, 11] and co-morbidities [12, 13]. Each potential cause of fever may also require unique treatment and control strategies. As a result, to maximise the utility of scarce public health resources and ensure appropriate treatments are available, it would be beneficial to know which pathogens cause the greatest proportion of the fever burden in different locations. Estimating the aetiology of fever is therefore paramount to reducing its disease burden but could also have positive consequences for antimicrobial overuse and thus antimicrobial resistance as the use of antimicrobials is common in unspecified fever [154].

Currently, estimates of fever aetiology are rare [5]. Studies that have looked at fever aetiology are often limited by a lack of access to diagnostic tests for the wide variety of potential causes of fever, small sample sizes and recruiting cases only [10]. The challenge with these limitations is that different studies find discrepant results. While this is not surprising given the limited generalisability of each individual study, it does lead to challenges of unclear guidance on which treatments should be made accessible and which diagnostic tests are required. Specifically, limited access to any diagnostics as well as limited access to the most sensitive and specific diagnostics for particular fever-causing pathogens, means that many fever aetiology studies carried out in low-resource settings are reliant on confirming diagnoses with imperfect diagnostic tests. For example, the gold-standard confirmatory diagnostic test for dengue is viral neutralization test (VNT) [155], however, this test is not often accessible in low-resource settings [156]. Instead, dengue diagnosis in these settings is often determined based on serological tests with lower diagnostic sensitivity than VNT [157]. This is just one example of a common

phenomenon in diagnostic accessibility and the result of this, for dengue and in general, is that current estimates of the causes of fever reliant on tests with imperfect sensitivity are an underestimate of the true number of cases for a specific cause. On the other hand, assuming sensitivity is perfect but specificity is imperfect, current estimates are an overestimate of the true number of cases for a specific cause.

Studies with small sample sizes are challenging because absence of particular infections may reflect a cohort that, by nature of its size, does not include individuals with those infections or indeed it may reflect that in this population that infection is not important. Alternatively, absence of evidence is not the same as evidence of absence. Finally, the challenge of recruiting cases only in fever aetiology studies limits the interpretation of results. For example, a respiratory pathogen may appear common in a cohort of fever cases. Without a control population it could be inferred that respiratory pathogens are a key pathogen related to fever. However, the same respiratory pathogen may appear equally prevalent in controls. Thus, controls allow for the distinguishing of a pathogen that is related to fever with a pathogen that is a common colonizer of individuals in this population and is not related to fever.

A further gap in current fever aetiology studies is the large proportion of fever cases that remain without an aetiological diagnosis highlighting major gaps in our understanding of the causes of fever. A recent systematic review [158] investigated the extent of fever with an unidentified aetiology in adults across East Africa, the geographic sub-region of Africa with arguably the best data on fever aetiology to date [159]. The authors found 20 studies that met their inclusion criteria. Included studies ranged in size from 90 to 1425 patients, all studies were carried out in a single country and the percentage of fever cases with unidentified aetiology varied from 11% to over 90%. Overall, the reviews of Maze *et al.* detailed above, [159] and also Nooh *et al.* [158] highlight that diagnostic testing for only a small number of pathogens and small sample sizes have been a common limitation in aetiology studies of fever to date. In 2023, a fever aetiology study carried out in Uganda also had the same limitations with a total of 1281 participants and a limited number of pathogens investigated [157].

The aim of the FIEBRE study, is to improve upon previous studies by recruiting a large number of cases (both inpatients and outpatients) and controls from multiple countries, where previous estimates of fever aetiology studies are scarce, as well as testing for a large number (>50) of pathogens. The FIEBRE study will be described in more detail in

Chapter 7.

6.3 Multi-site case-control aetiology studies for syndromes

The following section outlines two examples of aetiology studies that used a case-control design. Each example includes a brief introduction to the methodology used to analyse the data collected, to estimate the aetiologies for each disease/syndrome. Each of these examples, along with the FIEBRE study, was designed to overcome limitations of earlier studies and improve the understanding of the aetiology for the given syndrome with the goal of identifying key pathogens that can be targeted for interventions, vaccines or diagnostic test development.

6.3.1 Global Enteric Multicenter Study

The Global Enteric Multicenter Study (GEMS) investigated the aetiology of moderate-to-severe diarrhoea in children across seven countries in sub-Saharan Africa and South Asia [38]. Diarrhoeal disease is one of the top two causes of death in young children in the developing world [160] and there is potential to reduce mortality from diarrhoeal disease through social interventions such as improved housing and provision of sanitation and safe water but also via targeting the major aetiological causes. A successful example of targeting a key diarrhoeal causing pathogen to reduce the burden of morbidity and mortality from diarrhoeal disease, is the measles vaccine [140]. These characteristics make diarrhoea a good example of when an aetiology study can be particularly useful [148].

In GEMS approximately 9500 cases (between 700 and 2000 per site) were enrolled alongside 13000 matched controls (between 1300 and 2500 per site) [142]. Cases were recruited from children seeking care at an included sentinel health facility (hospitals, urgent care facilities and community clinics) and who met selection criteria. Controls were identified from catchment areas of selected sentinel health centres [161]. A case-control study design, as opposed to a cohort design, was required as cases of moderate-to-severe diarrhoea represent only a small fraction of diarrhoeal cases. GEMS researchers also desired a multi-site study with common research protocols and standardised epidemiologic and microbiologic methods to facilitate inferences from the findings across sites[162]. Diagnostic testing included; culture from stool specimens, multiplex polymerase chain reaction (PCR) and ELISA. The selected tests were chosen

based on a balanced consideration of cost, robustness and performance [163], and are likely all imperfect.

The primary analysis for estimating the proportion of cases of moderate-to-severe diarrhoea attributable to specific pathogens uses the ideas introduced above where the risk ratio is approximately estimated by the odds ratio. In the simple case, where it is assumed that there is just a single exposure, pathogen (A), that causes diarrhoea, then, if a case is denoted $Y = 1$ and control $Y = 0$ the odds ratio for diarrhoea, comparing presence versus absence for pathogen A, is estimated by fitting the conditional logistic regression model as in Blackwelder *et al.* [164]

$$\log\left(\frac{\Pr(Y_i = 1|X_{Ai})}{1 - \Pr(Y_i = 1|X_{Ai})}\right) = \alpha + \beta_A X_{Ai} \tag{6.3.1}$$

where, α is the intercept, X_{Ai} represents the presence or absence of pathogen A for individual i and β is the corresponding coefficient. In this case $\exp(\beta_A)$ is the ratio of the odds of diarrhoea when pathogen A (the exposure) is present to the odds of diarrhoea when pathogen A is absent. Then the unadjusted population attributable fraction for pathogen A is given by:

$$PAF = \Pr(A|Diarrhoea)(1 - 1/\exp(\beta_A)) \tag{6.3.2}$$

as in equation 6.2.2, where $\Pr(A|Diarrhoea)$ represents the proportion of diarrhoea cases in which pathogen A is present (η in 6.2.2). A secondary analysis also considered the attributable fraction for a specific pathogen adjusted for the presence of other pathogens. This was due to a high proportion of cases and controls having positive diagnostic tests for more than one pathogen. To consider the adjusted PAF, the conditional logistic regression model is extended to include additional covariates for each additional pathogen as well as the interaction of the effects of pathogen A and each additional pathogen.

Limitations of this approach are that all diagnostic tests used were assumed 100% sensitive which underestimates the role of a pathogen if the true sensitivity is less than 100%. However, the presence of controls and their use in the analysis accounts for imperfect diagnostic test specificity. Estimation of the odds ratio requires that tests must be applied to both cases and controls. As a result, tests not used on controls cannot be utilised. Further, only the results of a single diagnostic test for each pathogen can be incorporated. Additionally, only pathogens with odds ratios statistical significantly greater than one were considered associated with case status, those not meeting this criteria are excluded as a

potential cause. Finally, with this approach, the sum of the attributable fractions is not constrained to 100%. If the PAFs for each pathogen do sum to greater than 100%, it is not clear if this reflects the fact that more than one pathogen is simultaneously responsible for diarrhoea or a result of bias in estimating the odds ratio [165].

6.3.2 Pneumonia Etiology Research for Child Health Study

The Pneumonia Etiology Research for Child Health (PERCH) study investigated the causes of pneumonia in children in nine study sites across seven countries [38]. Across all study sites over 4000 cases and over 5000 community controls were enrolled. A case-control study design was again found to be favourable to ensure a sufficient number of cases with severe and very severe pneumonia were enrolled and controls to allow estimation of aetiologic fractions [38].

Within PERCH, as in all case-control studies, much consideration was taken when selecting the most appropriate controls so that they are the most representative of the general population from which the cases arose. The catchment area, representing the geographic location of where both cases and controls were found, was defined by examining the residence of cases from hospital logs in the previous year [166]. This approach circumvents the issue that some cases will come from further afield than a hospital's pre-defined catchment area due to factors including quality and cost of services or severity of illness. Systematic sampling was then used to recruit cases who showed up to selected health facilities and met inclusion criteria. Community controls were randomly selected to be representative of the defined catchment area. In PERCH, this meant including controls with respiratory symptoms so long as they did not meet the criteria for a case [167].

The data collected in PERCH led to novel analytical challenges [165]. Firstly, multiple tests were used to detect the same pathogen, for example, to detect pneumococcal both PCR and blood culture were used. Similarly multiple specimen types were used for the same pathogen, for example, both nasopharyngeal and induced sputum specimens were tested by PCR to detect a variety of pathogens. This created an analytical challenge that did not have to be addressed in GEMS. Secondly, the imperfect test accuracy (notably test sensitivity) was recognised as a key issue in PERCH that needed to be incorporated into analyses. To address both of these challenges PERCH researchers developed what they called the Perch Integrated Analysis [37].

The primary analysis for estimating the proportion of cases of severe and very severe pneumonia using the Perch Integrated Analysis relies on a Bayesian partial latent class model [168]. The model is only partially latent as by design, the controls are known to have no pathogen infecting the lung and thus their status is not unknown but known. Detection of a pathogen in a control participant then indicates a false-positive result and provides direct estimates of diagnostic test specificity. This analysis approach requires four inputs:

1. The list of pathogens that are potential causes of pneumonia
2. Prior distributions for each aetiologic fraction
3. Prior distributions for the sensitivity and specificity of each diagnostic test
4. Observed diagnostic test results

Prior distributions offer the opportunity to include information from previous studies and clinical and biological understanding into the model that would otherwise not be included [169]. Since direct estimates of test specificity are available from control participants test results, uninformative prior distributions are used. This Bayesian analysis approach, including the full model specification applied to the FIEBRE study, will be discussed in detail in Chapter 7.

Limitations of this approach, include that if more than one pathogen was found to be in the lung it is not possible to determine whether one or both pathogens are the cause. Another limitation, that the current thesis tried to address, is the validity of the assumptions regarding diagnostic test sensitivity. Priors for diagnostic test sensitivity were obtained from infectious disease and laboratory experts [170] who may have subjective beliefs about a test's accuracy. However, sensitivity analyses showed their estimates were robust to the choice of prior distributions for the sensitivity of each diagnostic test.

Chapter 7

Febrile Illness Evaluation in a Broad Range of Endemicities: The FIEBRE study

7.1 Preamble

This chapter introduces the Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) study to which the statistical methods explored in this thesis are applied to estimate fever aetiology. The FIEBRE study design is outlined focusing on the details most relevant to the statistical analysis approach. I then provide a descriptive summary of enrolled participants and of the observed test results that serve as one of the key inputs to the analysis of fever aetiology.

Data collection for the FIEBRE study was from 2018 to 2021, so was impacted by the COVID-19 pandemic. One challenge as a result of this, was the length of time it took to organise flights to send collected samples to reference laboratories for diagnostic testing because at this time, most flights were cancelled. As a result of these delays, data were received later than planned and members of staff originally employed to undertake the FIEBRE data management were no longer in place to carry out these duties. Consequently, I undertook an unexpected data management role for the FIEBRE study and my contribution to this aspect of the study is outlined in section 7.3.

7.2 FIEBRE overview

The Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) study is a multisite case-control study of the causes of fever in sub-Saharan Africa and Asia. The three primary objectives of FIEBRE were:

1. To determine the treatable and/or preventable causes of fever in children aged $\geqslant$ 2 months and in adults presenting as outpatients, and among those admitted to hospitals, in areas represented by the study sites.

2. To determine how fever aetiology varies according to patient age, geographical area, local malaria and HIV prevalence, and other risk factors.

3. To determine the prevalence and spectrum of antimicrobial resistance among bacterial pathogens identified in clinical specimens from febrile patients.

In this thesis I sought to answer the first two primary objectives using Bayesian latent class estimation.

The motivation behind FIEBRE is that fever is a leading cause of healthcare seeking behaviour [1] and contributes to significant morbidity and mortality [4] in low resource settings. Further, reductions in the malaria burden in recent decades [171] mean that it's increasingly important to identify other causes of fever. It is hoped that elucidation of key fever causing pathogens, that are treatable and/or preventable, can help reduce this burden. The study sites were selected due to a paucity of published data on fever and its underlying causes. They were also selected to ensure variability in the HIV and malaria epidemiology which is representative of the wider study regions. Both HIV and malaria may be reasons for different fever aetiologies and so it was important to be able to capture this variability. Similarly, a variety of rural and urban sites were desired (See Table 7.1). Despite the differences between sites, the FIEBRE protocol aimed to ensure harmonised research protocols across sites. This standardised approach ensures that across all sites the same diagnostic tests were used, the same sampling techniques and the same inclusion criteria so that results are comparable across sites.

A full protocol has been published elsewhere [20] but briefly, paediatric (2 months and older and less than 15 years) and adult (15 years and older) inpatients and outpatients presenting with fever and meeting the recruitment criteria were enrolled from sites across four countries (Laos, Malawi, Mozambique and Zimbabwe - See Figure 7.1). Cases were drawn from the population who seek healthcare at the included healthcare facilities and all cases met the following inclusion criteria [20]:

1. Tympanic or axillary temperature of $\geqslant$37.5 degrees Celsius at presentation.

Table 7.1: Characteristics of the study sites in the FIEBRE case-control study

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
Name of health facilities where patients are recruited	Phonhong Vientiane Provincial Hospital	Chikwawa District Hospital	Manhiça District Hospital	Harare Central Hospital, Chitungwiza General Hospital and three primary care clinics in Harare City
Region of country	Northwest	South	South	North central
Demographic classification	Peri-urban and rural	Rural	Rural	Urban
HIV epidemiology (2018 national seroprevalence among adults aged 15-49 years)	0.3%	9.2%	12.6%	12.7%
Malaria epidemiology	Low transmission of *P. falciparum* and *P.vivax*	Perennial transmission of *P.falciparum* with marked seasonality	Perennial transmission of *P.falciparum* with marked seasonality	No local malaria transmission

Note. Table recreated from [20]

2. Not having been hospitalised in the previous month.

3. Age ⩾2 months.

4. For outpatients, residence within the defined catchment around an included health facility.

5. For outpatients aged ⩾15 years, absence of symptoms of lower respiratory infection and of diarrhoeal disease.

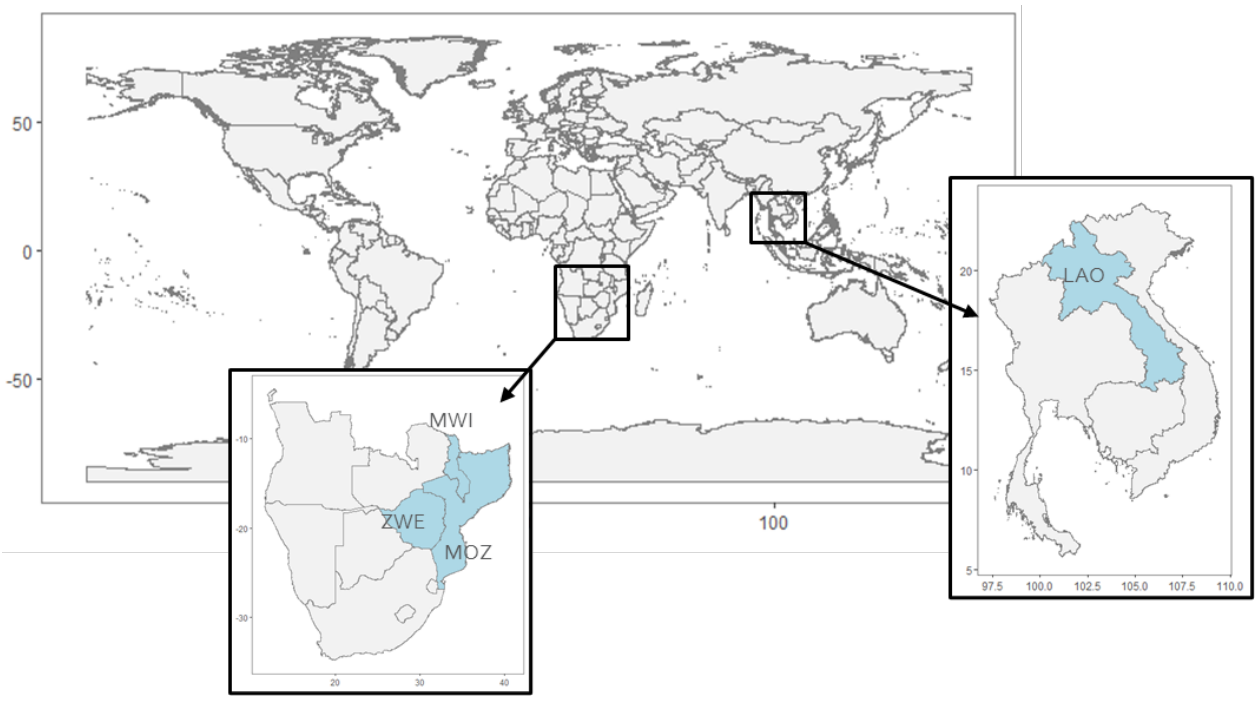
6. For outpatients aged ⩾2 months and <15 years, absence of symptoms of diarrhoeal diseases.

7. Willingness and ability to provide demographic and clinical information, and clinical samples at the time of enrolment and 28 days later.

8. Provision of written informed consent for adult participants; or for children, provision of written consent from a parent/guardian.

Controls were recruited from the catchment areas of study site health facilities. One control was recruited for every two outpatient cases and controls were matched to the outpatient cases by month of enrolment, age, sex and location of residence [20].

Figure 7.1: FIEBRE study countries (LAO=Laos, MWI=Malawi, MOZ=Mozambique and ZWE=Zimbabwe)



FIEBRE began recruiting participants across sites in 2018 and recruitment finished in March 2021. Similar to GEMS and PERCH, FIEBRE

chose a case-control study design. In FIEBRE this design provided several advantages. Recruiting cases based on fever status ensured a large number of fever cases were recruited and by recruiting on the outcome status, FIEBRE didn't have to follow participants over time and wait for the outcome to happen as in prospective cohort studies. Community controls allow the estimation of the background prevalence of pathogens and estimation of disease incidence which would be the first of their kind for some infections. Community controls were matched to outpatient cases only, as inpatients may be referred from outside the healthcare facility catchment area and so finding appropriate controls requires more resources. A consequence of inpatients arising from outside the health facility catchment area is that they may be less representative of the fever epidemiology within the catchment area due to different exposures.

From all participants, nasophyaryngeal and/or oropharyngeal swabs as well as a venous blood sample were collected plus urine samples from selected participants. These samples were assessed by a list of predetermined diagnostic tests for over 50 different pathogens specified in Table 7.2. FIEBRE tested for those pathogens which are known to be treatable and/or preventable causes of fever so that if a pathogen is found to dominate the causes of fever, FIEBRE results would be able to help to guide empirical therapy, control measures, resource allocation and prioritisation of clinical diagnostics [20], to reduce the burden of fever from that pathogen.

Table 7.2: List of infections or pathogens sought in FIEBRE in alphabetical order and the diagnostic test/s used

Infection or pathogen	**Diagnostic Test**
Acinetobacter baumannii	Blood culture
Adenovirus	Luminex respiratory pathogen panel
Brucella	BrucellaCapt, Microscopic Agglutination
Burkholderia pseudomallei	Blood and urine culture
Chikungunya	PCR, IgM ELISA, IgG ELISA, microneutralization
Chlamydophila pneumoniae	Luminex respiratory pathogen panel
Coronavirus 229e	Luminex respiratory pathogen panel
Coronavirus hku1	Luminex respiratory pathogen panel
Coronavirus nl63	Luminex respiratory pathogen panel

Continued on next page

Table7.2 – continued from previous page

Infection or pathogen	**Diagnostic Test**
Coronavirus oc43	Luminex respiratory pathogen panel
Cryptococcus species	Rapid Diagnostic Test (RDT), Blood culture
Dengue	PCR, IgM ELISA, IgG ELISA, microneutralization
Enterobacterales other than *E.Coli*	Blood and urine culture
Enterococcus faecalis	Blood and urine culture
Escherichia coli (E.coli	Blood and urine culture
Haemophilus (H.) Influenzae	Blood culture
Histoplasma	Histoplasma antigen test
Human bocavirus	Luminex respiratory pathogen panel
Human metapneumovirus	Luminex respiratory pathogen panel
Influenza A	Luminex respiratory pathogen panel
Influenza AH1	Luminex respiratory pathogen panel
Influenza AH3	Luminex respiratory pathogen panel
Influenza B	Luminex respiratory pathogen panel
JEV	PCR, IgM ELISA, IgG ELISA, microneutralization
Klebsiella pneumoniae	Blood and urine culture
Legionella pneumophila	Luminex respiratory pathogen panel
Leishmaniasis	ELISA
Leptospirosis	ELISA, MAT, PCR
Malaria	RDT, microscopy
Mycobacteria other	Blood culture
Mycobacterium tuberculosis complex	ULAM, Blood culture
Mycoplasma pneumoniae	Luminex respiratory pathogen panel
Non-typhoidal *Salmonella*	Blood culture
O'nyong'nyong	PCR, microneutralization
Parainfluenza 1	Luminex respiratory pathogen panel
Parainfluenza 2	Luminex respiratory pathogen panel
Parainfluenza 3	Luminex respiratory pathogen panel
Parainfluenza 4	Luminex respiratory pathogen panel
Qfever	IgM and IgG IFA, PCR

Continued on next page

Table7.2 – continued from previous page

Infection or pathogen	Diagnostic Test
RSV A	Luminex respiratory pathogen panel
RSV B	Luminex respiratory pathogen panel
Rhinovirus / Enterovirus	Luminex respiratory pathogen panel, PCR
Scrub Typhus	IgM and IgG IFA, PCR
Spotted Fever group	IgM and IgG IFA, PCR
Staphylococcus Aureus	Blood and urine culture
Streptococcus other	Blood culture
Streptococcus pneumoniae	Blood culture
Streptococcus pyogenes	Blood culture
Talaromyces marneffei	Blood culture
Typhoidal *Salmonella*	Blood culture
Typhus Group	IgM and IgG IFA, PCR
Urinary tract infection	Urine dipstick and urine culture
Zika	PCR, IgM ELISA, IgG ELISA, microneutralization

The target sample size at each site was 600 cases within each stratum (children-inpatients, children-outpatients, adult-inpatients, adult-outpatients). Controls were aimed to be matched one control to two participating outpatients by month of enrolment, age, gender and location of residence so that at least 600 community controls (matched to 1200 outpatients) were recruited at each site.

7.3 FIEBRE data management

This section briefly outlines the four months of data management work carried out during this PhD for the FIEBRE study. This was a time consuming piece of work, and distracted me from the main purpose of my PhD. However, upon reflection it has given me practical knowledge of issues in study conduct and data management and overall, I think the experience was a valuable one.

The FIEBRE study includes data collected both at the sites and at reference laboratories. Delays arising from COVID-19 meant the final data from reference laboratories were delayed and data that had been received was not cleaned and stored appropriately before the end of the contract for the FIEBRE data management team. As a result, I along

with Polycarp Mogeni (another FIEBRE statistician) and Heidi Hopkins (FIEBRE scientific programme coordinator) took on the remaining data management from February 2023.

The FIEBRE dataset can be divided into three parts: the first contains clinical data collected at point of care/enrollment for cases; the second contains clinical data collected at point of enrollment for controls; and the third is the diagnostic test results received from reference laboratories in 11 different data sets. For each part, the data forms and data collection had already occurred but the data were poorly organised and little data cleaning had been done.

For the case and control datasets, which each contained over 2000 variables, no data dictionary existed. My first task, undertaken with clinical input from my colleague Heidi Hopkins, was to review each variable and create a user-friendly data dictionary so that not only could FIEBRE primary objectives be met but that future researchers would also be able to use the rich data collected. The second job that was required was cleaning. There were many issues that needed to be addressed. These included, data being recorded inconsistently (for example data for test results entered as Neg, Negative, negative, or 0). Sometimes the same information being split over multiple variables (for example, when a question was worded slightly differently at one site to the other three and this one site's data was in a separate variable to the other three sites' data), so I created derived variables containing the pertinent information. There were also many quality and missing data checks to carry out.

For each of the 11 reference laboratory datasets a similar but slightly different process was followed as these data had not previously been integrated into the case or control datasets. The following steps were carried out for each reference laboratory dataset:

1. Create a single dataset for each of the 11 reference laboratory results (often data was sent across multiple Excel sheets with different variable names in each sheet)

2. Quality check

    (a) Compare the number of results against shipping lists to see if we are missing results for participants that should have a result.

    (b) Are there any implausible values? (for example, sometimes titre threshold results originally entered as text are unintentionally converted to numbers during handling in Excel)

   (c) Are there any duplicates (two results for the same specimen) due to labelling errors?

   (d) Check back with reference laboratories to resolve any queries

3. Create a data dictionary

4. Tidy inconsistent data (for example, the recording of positive and missing results)

5. Derive final variables for analysis

   (a) Where there are paired samples for serological tests, a positive test result is a combination of the results on the acute and convalescent sample so either a single variable that accounted for seroconversion or a four-fold rise between samples was needed

   (b) Where results are continuous deciding a threshold or cut-off value for a positive test result which may differ by site.

   (c) Where an individual had multiple results for the same pathogen (i.e. multiple specimens were tested) and results were divergent, deciding a final result in collaboration with clinicians

6. Create descriptive summary tables to share with FIEBRE co-investigators (See Table 7.3 for an example)

7.4 FIEBRE data description

7.4.1 Characteristics of enrolled cases and controls

This section describes the data collected in FIEBRE that is specifically related to the estimation of the attributable fractions. Other clinical information on presentation and exposures was collected and not discussed in this thesis. Overall, 10252 participants were enrolled representing 7851 cases and 2401 controls. Table 7.4 describes the demographic characteristics of included cases and controls as well as whether cases were enrolled as inpatients or outpatients. Table 7.4 shows that within each site, close to 2000 cases were enrolled (less than the target sample size of 2400). The number of community controls enrolled was also less than the goal of 600 in all sites except Malawi where 908 controls were enrolled. In total, among the cases more outpatients than inpatients were enrolled (56.5%) but adults (participants aged 15 years or over) and children (participants

Table 7.3: Descriptive summary table for the CrAg results in Zimbabwe. Results are displayed as number (%)

	Total cases enrolled	**Total tested**	**Test positive**
Overall count	1924	496 (25.8)	14 (2.8)
Admission status			
Inpatient	801	436 (54.4)	13 (3.0)
Outpatient	1123	60 (5.3)	1 (1.7)
Age group			
<5 years	474	83 (17.5)	0 (0.0)
≥5 to <15 years	402	72 (17.9)	0 (0.0)
≥15 years	1048	341 (32.5)	14 (4.1)
Sex			
Female	952	236 (24.8)	4 (1.7)
Male	972	260 (26.7)	10 (3.8)
HIV status			
Negative	1662	284 (17.1)	0 (0.0)
Positive	231	206 (89.2)	14 (6.8)
Don't know	31	6 (19.4)	0 (0.0)
Day 28 clinical status			
Died	55	43 (78.2)	4 (9.3)
Alive recovered	1376	299 (21.7)	7 (2.3)
Alive improved	149	64 (43.0)	2 (3.1)
Alive but same as day 0	3	2 (66.7)	0 (0.0)
Alive worse than day 0	9	5 (55.6)	0 (0.0)
Lost to follow up	193	50 (25.9)	0 (0.0)
Alive no information	0	0	0
Missing	139	33 (23.7)	0 (0.0)

*Note.*No controls were tested for CrAg

less than 15 years) as well as males and females were roughly evenly enrolled between inpatients and outpatients. In controls, more adults were recruited (57.1%) as well as more females (59.5%).

Table 7.5 describes the demographics of cases by inpatient and outpatient status. Similar numbers of adult and child inpatients were enrolled but this varied by site. In Malawi, 61.3% of inpatients enrolled were children whereas in Laos, only 40.5% of those inpatients enrolled were children. Similar proportions of male and female inpatients were enrolled across all sites. In outpatients, a greater proportion of adults were enrolled (53.1%) and a greater proportion of females (56.8%).

7.4.2 Diagnostic test data

The diagnostic test results are a central component of the analysis to estimate the attributable fraction for each pathogen. The design of the FIEBRE study means that all participants provided specimens so that they could be tested for all of the pathogens listed in Table 7.2. However, there were a subset of pathogens for whom only cases (fever patients) were tested. These included all microbiology and mycobacterium which were tested by culture which is known to have a very high, almost perfect specificity (i.e. recruited controls would almost certainly never have the pathogen). The result of this assumption, is that we can assume a high specificity without needing the control data to estimate specificity. Further categories of tests where we do not expect all participants to have a result are listed below.

1. Certain diagnostic tests require the comparison between a sample taken at enrollment (on day 0) and a sample taken 28 days later. In these instances, loss to follow up means not all participants had a sample taken at day 28. When a test ideally would have two samples but in a group of individuals we only have a day 0 sample, we treat these data as a separate type of test. We would not expect the same sensitivity and specificity from just the day 0 sample as if we could compare between day 0 and day 28.

2. Small sample volume. For some participants, the sample volume was not sufficient to be able to send a sample to all the reference laboratories. In cases of small sample volume, a pre-specified order of which laboratories to send samples to was followed.

3. Malaria microscopy in Malawi. Errors were made during malaria microscopy in Malawi such that no microscopy data were used from the Malawi site but malaria RDT data were still available.

Table 7.4: Summary of numbers of cases and controls by site, shown overall and by in/out-patient status, age group, and sex. Results are number (%)

	Laos		**Malawi**		**Mozambique**		**Zimbabwe**		**Total**	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Total enrolled	1972	485	1773	908	2182	572	1924	436	7851	2401
In/Outpatient										
Inpatient	1023 (53.2)	-	569 (32.1)	-	1026 (47.0)	-	801 (41.6)	-	3419 (43.5)	-
Outpatient	949 (49.3)	-	1204 (67.9)	-	1156 (53.0)	-	1123 (58.4)	-	4432 (56.5)	-
Age										
Age<15 years	760 (38.5)	185 (38.1)	952 (53.7)	399 (43.9)	1161 (53.2)	302 (52.8)	876 (45.5)	141 (32.3)	3749 (47.8)	1027 (42.8)
Age⩾15 years	1212 (61.5)	300 (61.9)	821 (46.3)	509 (56.1)	1021 (46.8)	266 (46.5)	1048 (54.5)	295 (67.7)	4102 (52.2)	1370 (57.1)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	4 (0.2)
Sex										
Female	979 (49.6)	240 (49.5)	991 (55.9)	559 (61.6)	1199 (54.9)	367 (64.2)	952 (49.5)	262 (60.1)	4121 (52.5)	1428 (59.5)
Male	993 (50.4)	245 (50.5)	782 (44.1)	349 (38.4)	983 (45.1)	203 (35.5)	972 (50.5)	174 (39.9)	3730 (47.5)	971 (40.4)
Missing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.1)

Table 7.5: Summary of numbers of in/out patient cases by site, shown overall and by age group, and sex. Results are number (%)

	Laos		**Malawi**		**Mozambique**		**Zimbabwe**		**Total**	
	In	Out	In	Out	In	Out	In	Out	In	Out
Total enrolled	1023	949	569	1204	1026	1156	801	1123	3419	4432
Age										
Age<15 years	414 (40.5)	346 (36.5)	349 (61.3)	603 (50.1)	554 (54.0)	607 (52.5)	355 (44.3)	521 (46.4)	1672 (48.9)	2077 (46.9)
Age⩾15 years	609 (59.5)	603 (63.5)	220 (38.7)	601 (49.9)	472 (46.0)	549 (47.5)	446 (55.7)	602 (53.6)	1747 (51.1)	2355 (53.1)
Sex										
Female	497 (48.5)	482 (50.8)	267 (46.9)	724 (60.1)	493 (48.1)	706 (61.1)	348 (43.4)	604 (53.8)	1605 (46.9)	2516 (56.8)
Male	526 (51.4)	467 (49.2)	302 (53.1)	480 (39.9)	533 (51.9)	450 (38.9)	453 (56.6)	519 (46.2)	1814 (53.1)	1916 (43.2)

4. Screening test followed by a confirmatory test. For certain infections, for example, Leptospirosis and Brucella, the diagnostic testing algorithm involved a screening test and only those positive by the screening test went on to have the second or confirmatory test. This approach is common and is a result of time and money constraints.

5. Leishmaniasis. Initial plans were that all participants would be tested for visceral leishmanisis by the direct agglutination test. However, none were tested using this originally-planned test. Instead, only a 5-10% random sample of participants were tested by a different test, ELISA, due to cost and time constraints with the broad goal of targeting 100 participants per site with a balance of cases and controls.

Binary diagnostic test results were sought for all pathogens. Data received from reference laboratories were varied including, binary data, ordered categorical data, continuous test results and results from paired samples taken on the day or presentation to a health facility and 28 days later. In collaboration with each reference laboratory, a binary result was derived from all received data. Due to the reasons listed above, as well as errors in data labelling, there is missing data. That is, participants who do not have an observed test result for all diagnostic tests. Table 7.6 reports the percentage of cases for whom we have at least one diagnostic test result for each pathogen sought and Table 7.7 the same for controls.

Overall, 99.5% (n=10198) of participants have an observed test result for at least one diagnostic test of any type. Of those participants with missing data for all diagnostic tests (n=54), 83.3% (n=45) were controls, 57.4% (n=31) were children, 87% (n=47) were from Malawi, 5.6% (n=3) from Mozambique and 7.4% (n=4) from Zimbabwe. Figure 7.2 highlights that participants did not all undergo all tests. The number of tests participants had depended on whether or not a participant was a case or control. The modal number of diagnostic tests that a control had was 18 and for a case it was 33. This difference is highlighted by the different distributions for cases and controls seen in Figure 7.2.

Figure 7.2: A histogram summarising the A) counts and B) cumulative number of diagnostic tests each participant had (of those with at least one diagnostic test result). Cases are shown in blue and controls in orange.

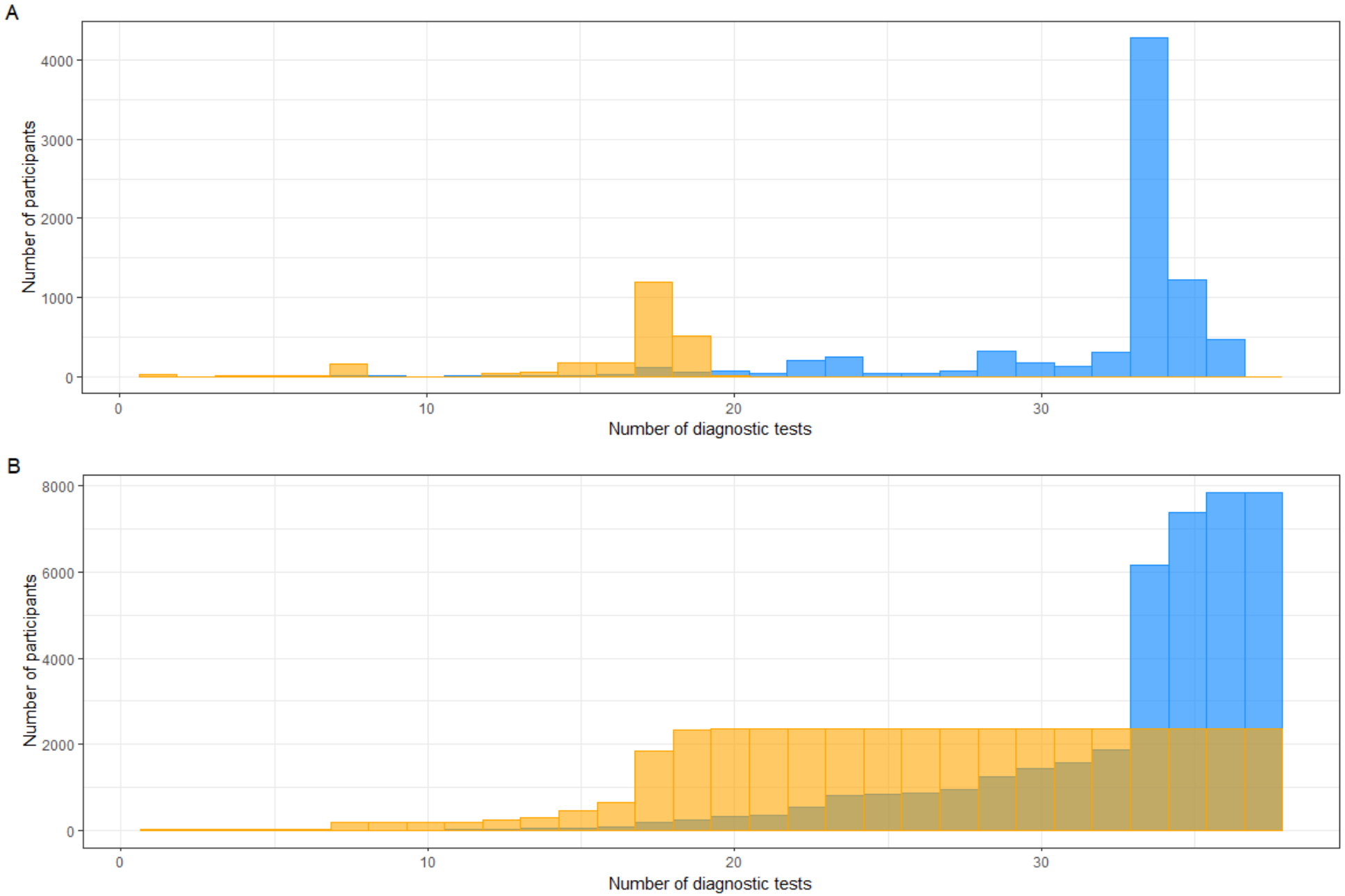


Table 7.6: Percentage of patients (In- and Out-) who received at least one diagnostic test for each infection group, shown by site and age group. Strata where less than 10% of enrolled participants were tested are indicated by bold font.

	Laos				Malawi				Mozambique				Zimbabwe			
	<15		15+		<15		15+		<15		15+		<15		15+	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Malaria	87.4	92.2	99.3	98.8	98.6	99.8	99.5	99.8	99.3	99.5	99.4	99.8	91.3	96.9	98.4	100.0
Respiratory pathogens	88.9	96.8	88.8	97.5	54.2	96.4	61.8	77.9	94.8	98.7	94.7	97.6	98.0	96.5	98.4	94.7
Histoplasma	**8.5**	**0.0**	83.4	**0.3**	61.6	**0.0**	80.9	**0.0**	41.7	**0.0**	76.5	**0.2**	27.0	**0.0**	55.8	**0.0**
Brucella	39.9	49.4	93.3	81.9	79.4	91.2	94.1	97.0	80.0	79.9	95.6	96.7	53.8	74.1	81.4	92.7
Leptospirosis	84.5	90.2	99.3	99.0	91.1	98.0	97.7	98.2	99.3	98.8	98.5	99.8	82.5	93.3	96.2	99.8
CrAg	47.6	**0.6**	65.8	46.8	45.0	**1.2**	59.1	**9.5**	48.4	**3.3**	96.6	44.3	41.1	**1.7**	65.2	**8.5**
Leishmaniasis	**3.6**	12.1	**7.6**	**7.0**	**3.4**	10.4	**8.2**	**8.2**	**2.7**	**3.6**	**2.5**	**7.5**	**7.0**	**5.4**	**3.8**	**7.3**
Arboviruses[1]	51.7	63.6	95.1	90.0	85.4	95.2	93.2	97.5	93.7	93.9	97.2	99.3	67.9	83.9	91.5	99.0
JEV[2]	51.7	63.6	95.1	90.0	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
O'nyong'nyong	**0.0**	**0.0**	**0.0**	**0.0**	85.4	95.2	93.2	97.5	93.7	93.9	97.2	99.3	67.9	83.9	91.5	99.0
Blood culture pathogens[4]	100	100	100	100	98.9	99.0	99.5	99.8	93.3	97.7	79.0	99.3	95.2	98.7	99.3	100
TB	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	39.3	46.5	**0.2**	**0.8**	77.7	21.1	**0.0**	**0.0**	55.1	19.1
Urinary Tract Infection	13.0	**2.0**	**9.7**	**1.8**	**8.0**	**4.1**	10.5	**5.3**	**8.7**	10.4	19.3	24.0	**5.1**	**6.3**	**4.3**	**5.0**
Rickettsioses[3]	52.4	56.1	92.3	85.7	58.5	78.9	63.2	85.5	57.6	51.2	58.1	74.0	40.6	48.9	60.5	72.9

[1] Arboviruses includes dengue, chikungunya and zika virus

[2] Japanese Encephalitis Virus

[3] Rickettsioses include scrub typhus group, spotted fever group, typhus group and Q fever

[4] Blood culture pathogens includes all those pathogens detected in blood cultures (*Burkholderia pseudomallei*, Tyhpoidal *Salmonella*, Non-typhoidal *Salmonella*, *Staphylococcus aureus*, *Escherichia coli*, Other enterobacterales, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*)

Malaria, respiratory infections, arboviruses (not including Japanese Encephalitis Virus (JEV) or O'nyong'nyong virus (ONNV)), brucella, leptospirosis and pathogens detected by blood culture were pathogens where at least 75% of enrolled participants were tested. While at least 75% of all participants tested, there were strata where fever participants were tested. For example, for respiratory infections, only 54% of inpatient children and 62% of inpatient adults in Malawi have diagnostic test results compared with over 85% in all other strata. Similarly, brucella test results were available for over 75% of all enrolled participants but under 50% of child participants in Laos have an observed test result for brucella. JEV was not tested for in African sites and ONNV was not tested for in Laos, but where JEV and ONNV were investigated, over 75% of participants were tested.

Blood cultures were performed for all cases in Laos and greater than 90% of cases across all other sites with the notable exception of inpatient adults in Mozambique where 79% of cases had a blood culture performed. Blood cultures were not performed in controls. The testing strategy for histoplasma, CrAg, urinary tract infections and TB by blood culture did not include all participants and so despite low percentages of participants tested across all strata, this is not unexpected. The testing strategy and diagnostic test for leishmaniasis changed over the course of the project due to time and money constraints. Ultimately the goal was to test a small random sample from each site and from those results decide whether or not to carry out more testing. Surprisingly, less than 75% of all enrolled participants had an observed result for ricketssioses despite a testing strategy that included all participants. This is perhaps not so surprising when we consider the diagnostic tests used for rickettsioses which relied on paired samples. In patient strata, outpatient children in Laos had the greatest proportion of patients with an observed test result at 92.3% and inpatient children in Zimbabwe had the lowest proportion of patients with an observed test result at 40.6% (Table 7.6).

The next sections present the percentage of participants with a positive diagnostic test result, for each infection sought, out of those tested. Where more than one diagnostic test was used to detect an infection, an and/or rule was applied to determine test positivity for the purpose of these crude summaries only.

Table 7.7: Percentage of controls who received at least one diagnostic test for each infection group, shown by site and age group. Strata where less than 10% of enrolled controls were tested are indicated by bold font.

	Laos		Malawi		Mozambique		Zimbabwe	
	<15	15+	<15	15+	<15	15+	<15	15+
Malaria	100.0	100.0	93.5	95.3	100.0	100.0	95.7	94.9
Respiratory pathogens	99.5	99.3	85.5	67.4	99.7	99.2	98.6	99.0
Histoplasma	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Brucella	54.1	85.0	66.7	91.2	60.9	95.1	26.2	32.2
Leptospirosis	73.5	93.3	91.5	92.7	98.7	99.2	73.8	74.6
CrAg	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Leishmaniasis	20.0	21.7	**1.5**	**3.1**	**4.0**	**5.6**	**4.3**	**2.4**
Arboviruses[1]	76.2	96.0	86.7	92.3	67.5	93.6	81.6	90.8
JEV[2]	76.2	96.0	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
O'nyong'nyong	**0.0**	**0.0**	86.7	92.3	67.5	93.6	81.6	90.8
Blood culture pathogens [4]	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
TB	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Urinary tract infection	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Rickettsioses[3]	28.1	39.7	**5.3**	**2.8**	**2.3**	**3.0**	**0.7**	**2.0**

[1] Arboviruses includes dengue, chikungunya and zika virus

[2] Japanese Encephalitis Virus

[3] Rickettsioses include scrub typhus group, spotted fever group, typhus group and Q fever

[4] Blood culture pathogens includes all those pathogens detected in blood cultures (*Burkholderia pseudomallei*, Tyhpoidal *Salmonella*, Non-typhoidal *Salmonella*, *Staphylococcus aureus*, *Escherichia coli*, Other enterobacterales, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*)

Respiratory pathogens

The total number of positive test results for each respiratory pathogen detected via the Luminex RPP is reported in Table 7.8 for cases and Table 7.9 for controls. Coronaviruses (not including COVID-19) are grouped into a single cause (coronavirus 229e, coronavirus hku1, coronavirus nl63, coronavirus oc43), as well as Influenza A (Influenza AH1 and Influenza AH3), Parainfluenza's (parainfluenza 1, parainfluenza 2, parainfluenza 3 and parainfluenza 4) and RSV (RSV A and RSV B). The respiratory pathogen, *Legionella pneumophila*, which most commonly presents as pneumonia, had no positive test results and is not included in any subsequent analyses.

The respiratory pathogen with the greatest percentage of positive test results in a participant stratum was rhinovirus/enterovirus in Malawi outpatient children at 45.8%. Rhinovirus/enterovirus consistently had the largest percentage of positive test results compared to other respiratory pathogens in both cases and controls. While not consistent across all sites and strata generally, there was a trend for a greater propor-

tion of positive test results in children than adults. In controls, respiratory pathogens where more than 10% of a strata tested positive are rhinovirus/enterovirus, adenovirus and human bocavirus.

Table 7.8: Percentage of those patients (In- and Out-) tested for respiratory infections with Luminex RPP that tested positive, shown by site and age group

	Laos				Malawi				Mozambique				Zimbabwe			
	< 15		15+		< 15		15+		< 15		15+		< 15		15+	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Influenza A	6.0	14.0	2.6	5.4	2.1	8.6	3.7	6.8	2.9	9.3	4.3	14.9	9.5	8.0	2.7	7.9
Influenza B	1.4	13.4	1.7	4.1	2.1	3.4	0.0	3.0	1.5	12.0	0.2	2.4	3.4	10.1	1.6	2.3
RSV	1.4	2.7	0.2	0.7	6.3	4.1	0.0	0.6	6.5	5.3	1.1	1.9	7.5	3.0	0.5	1.2
Adenovirus	4.6	2.4	0.0	0.3	9.5	13.9	0.7	1.5	7.0	12.7	1.3	2.8	7.8	6.4	0.5	0.4
Rhinovirus/Enterovirus	15.8	15.2	5.4	8.0	30.7	45.8	8.1	15.0	28.8	23.0	11.9	18.3	19.8	19.1	14.6	8.6
Human Bocavirus	1.9	1.8	0.2	0.2	6.9	16.0	2.2	2.4	8.0	4.8	1.3	3.0	4.9	5.0	3.2	1.8
Parainfluenza	6.3	4.5	0.7	1.4	6.3	16.7	2.2	3.2	5.3	5.0	1.1	3.2	4.0	5.0	3.0	1.6
Coronavirus	2.2	2.1	1.1	1.7	3.2	5.3	1.5	1.9	3.2	4.0	1.8	2.4	2.3	1.6	3.4	1.8
Human metapneumovirus	1.6	0.0	0.2	0.3	2.1	6.2	1.5	1.1	2.5	2.5	0.4	1.7	1.4	1.4	0.5	0.0
Legionella pneumophila	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlamydophila pneumoniae	0.5	0.0	0.0	0.0	1.6	1.5	0.0	0.0	1.5	0.8	0.0	0.0	0.6	0.4	0.0	0.2
Mycoplasma pneumoniae	0.0	0.0	0.0	0.2	0.0	0.9	0.0	0.2	0.2	1.0	0.2	0.9	0.6	0.4	0.2	0.0

Table 7.9: Percentage of controls tested for respiratory infections by the Luminex RPP that tested positive, shown by site and age group

	Laos		Malawi		Mozambique		Zimbabwe	
	< 15	15+	< 15	15+	< 15	15+	< 15	15+
Influenza A	0.5	0.0	0.6	0.0	1.3	0.4	0.7	1.4
Influenza B	0.5	0.0	0.3	0.0	0.0	0.4	0.7	1.0
RSV	2.2	1.0	4.1	1.5	1.7	0.0	0.7	0.0
Adenovirus	0.0	0.0	10.9	2.0	4.0	0.0	2.2	2.4
Rhinovirus/Enterovirus	19.0	5.4	38.1	16.9	29.6	6.8	13.7	8.9
Human Bocavirus	1.1	0.3	17.6	3.5	3.7	0.0	1.4	1.4
Parainfluenza	0.5	0.7	8.5	1.2	1.7	0.8	2.2	1.0
Coronavirus	4.3	1.0	4.1	2.3	3.0	0.4	0.7	2.1
Human metapneumovirus	0.5	0.0	2.6	0.6	1.0	0.0	1.4	0.0
Legionella pneumophila	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlamydophila pneumoniae	1.6	0.0	3.2	0.6	0.7	0.0	0.0	0.0
Mycoplasma pneumoniae	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0

Malaria

In Malawi and Mozambique, the percentage of inpatient children positive for malaria was high with 50.3% and 47.3% test positivity, respectively (Table 7.10). The percentage of positive test results in Malawi was also high in outpatient children and adults irrespective of whether they were inpatients or outpatients but was considerably lower in other strata in Mozambique outpatients (less than 15%). There were no positive test results for malaria in Laos (cases or controls). In controls, across all sites except Malawi, malaria positivity was less than 1% (Table 7.11).

Histoplasma, Brucella and CrAg

Test positivity in cases for histoplasma, brucella and CrAg was low, between 0 and 1% in all stratum except adults in Zimbabwe where the test positivity for CrAg was 4.5% in inpatients and 2.0% in outpatients (Table 7.10). Histoplasma and CrAg were not tested in controls. Although only very low levels of test positivity in cases for brucella, in adults in Zimbabwe and Mozambique, a similar percentage of controls tested positive as did outpatients (Table 7.11).

Leptospirosis and Leishmanaisis

Leptospirosis showed similar proportions of test positivity in patient strata as controls (See Tables 7.10 and 7.11). Two notable exceptions where cases showed a greater percentage of test positivity were in adult inpatients in Laos (16.5% of cases tested positive) and adult outpatients in Zimbabwe (16.8% of cases tested positive). For leishmaniasis, the strata with the greatest proportion of participants test positive was not

in cases but in controls. In Malawi, 31.3% of adults controls tested positive for leishmaniasis and in all strata, except child controls in Zimbabwe and Malawi, test positivity for leishmanaisis was above 2%.

Urinary tract infections

A small percentage of cases were tested for urinary tract infections (UTI) but of those tested, greater than 10% were diagnostic test positive in all African sites (Table 7.10). 66.7% of Mozambique adult outpatients tested were positive for a UTI. Across all sites, there was a trend for a greater proportion of positive urinary tract infection test results in adults compared with children. There were no positives in Laos outpatients and a only small percentage of those tested were positive in Laos inpatients. Despite a high proportion of positive test results, the testing strategy for UTIs was to collect urine samples for all cases under two years and from older cases who had symptoms of a UTI [20]. However, in practice in many cases under two years of age it was not possible to get a urine sample. Consequently, the proportion of positive test results is not necessarily a reflection of the proportion of positive test results in the wider FIEBRE case population given the limited number of participants tested.

Rickettsioses and TB blood cultures

Across patient strata, the percentage of cases that were tested for rickettsioses that had a positive test varied from 0.5% in child outpatients in Laos to 9.5% in child outpatients in Malawi (Table 7.10). There was no trend for a greater proportion of positive test results by age group or patient status. No controls tested positive for a rickettsioses (Table 7.11). Positive TB blood cultures were only found in inpatients adults. 5.1% of inpatient adults in Malawi, 7.7% of inpatient adults in Mozambique and just over a quarter (25.4%) of inpatient adults in Zimbabwe who were tested had a positive TB blood culture (Table 7.10). Similar to the results from UTIs, it is hard to infer much of TB positive blood cultures in the general fever case population given the limited number of participants tested. No other strata had any participants with a positive TB blood culture.

Arboviruses

A greater proportion of participants (cases and controls) tested positive for an arbovirus in Laos and Malawi than Mozambique and Zimbabwe. In

Table 7.10: Percentage of patients (In- and Out-) that tested positive for malaria, histoplasma, brucella, CrAg, leptospirosis, leishmanisis, TB blood cultures, urinary tract infections and rickettsioses, shown by site and age group

	Laos				Malawi				Mozambique				Zimbabwe			
	< 15		15+		< 15		15+		< 15		15+		< 15		15+	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Malaria	0.0	0.0	0.0	0.0	50.3	39.5	25.1	28.2	47.3	5.5	12.2	9.9	1.9	0.4	3.9	0.5
Histoplasma	0.0	-	1.2	0.0	0.5	-	0.6	-	0.0	-	1.1	0.0	0.0	-	0.8	-
Brucella	1.2	0.0	0.5	0.6	0.4	0.5	0.5	1.0	0.9	0.8	0.2	0.9	0.5	0.5	0.0	1.3
CrAg	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.4	0.0	1.3	0.4	0.0	0.0	4.5	2.0
Leptospirosis	6.9	5.4	16.5	9.5	2.2	3.4	4.7	11.2	2.4	2.7	4.7	8.6	5.1	6.0	6.8	16.8
Leishmanisis	13.3	2.4	2.2	4.8	8.3	20.6	5.6	30.6	26.7	4.5	16.7	14.6	12.0	17.9	11.8	9.1
TB	-	-	-	-	-	-	5.1	0.0	0.0	0.0	7.7	0.0	-	-	25.4	0.0
Urinary tract infection	1.9	0.0	6.8	0.0	25.0	16.0	26.1	21.9	12.5	15.9	33.0	66.7	16.7	12.1	21.1	40.0
Rickettsioses	2.3	0.5	8.2	4.8	5.9	9.5	2.2	4.5	5.0	3.2	4.4	3.9	1.4	2.4	3.7	6.6

Table 7.11: Percentage of controls that tested positive for malaria, histoplasma, brucella, CrAg, leptospirosis, leishmanisis, TB blood cultures, urinary tract infections and rickettsioses, shown by site and age group

	Laos		Malawi		Mozambique		Zimbabwe	
	< 15	15+	< 15	15+	< 15	15+	< 15	15+
Malaria	0.0	0.0	20.1	7.4	0.3	0.0	0.0	0.0
Histoplasma	-	-	-	-	-	-	-	-
Brucella	0.0	2.0	0.8	0.0	0.5	1.2	0.0	1.1
CrAg	-	-	-	-	-	-	-	-
Leptospirosis	5.9	5.7	3.0	8.9	4.0	8.0	3.8	7.3
Leishmanisis	5.4	6.2	0.0	31.3	8.3	6.7	0.0	14.3
TB	-	-	-	-	-	-	-	-
Urinary tract infection	-	-	-	-	-	-	-	-
Rickettsioses	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Laos, 17.3% to 31.3% of patient strata tested positive for dengue (Table 7.12). In Laos, dengue was the arbovirus with the greatest proportion of positive test results in cases but JEV was also common. Across all sites, chikungunya positivity was mostly found in child inpatients and zika in adults. In Malawi, chikungunya was the arbovirus with the greatest proportion of positive test results. There were no positives for O'nyong'nyong virus and so no further analyses are carried out for O'nyong'nyong virus.

In Laos, Mozambique and Zimbabwe, positivity to arboviruses in controls was less than 8% (Table 7.13) in each strata however, in Malawi greater than 10% of controls tested positive for chikungunya (children and adults) as well as for dengue (adults only).

Table 7.12: Percentage of patients (In- and Out-) that tested positive for arboviruses (dengue, chikungunya, zika virus, Japanese Encephalitis Virus (JEV) and O'nyong'nyong virus), shown by site and age group

	Laos				Malawi				Mozambique				Zimbabwe			
	< 15		15+		< 15		15+		< 15		15+		< 15		15+	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Dengue	23.8	17.3	26.1	31.3	16.1	9.4	17.1	16.4	13.5	3.0	5.9	6.1	5.4	3.0	7.4	4.7
Chikungunya	17.3	4.5	8.3	2.2	28.9	14.5	24.9	23.9	22.5	6.7	9.8	8.8	7.5	6.4	9.6	9.1
Zika	3.7	2.7	11.1	9.4	14.8	9.9	13.2	7.8	15.6	2.5	4.8	2.9	2.5	0.7	3.2	0.2
JEV	18.2	13.6	20.7	18.2	-	-	-	-	-	-	-	-	-	-	-	-
O'nyong'nyong	-	-	-	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 7.13: Percentage of controls that tested positive for arboviruses (dengue, chikungunya, zika virus, Japanese Encephalitis Virus (JEV) and O'nyong'nyong virus), shown by site and age group

	Laos		Malawi		Mozambique		Zimbabwe	
	<15	15+	<15	15+	<15	15+	<15	15+
Dengue	5.7	5.9	7.2	12.3	0.0	0.4	0.0	4.9
Chikungunya	7.1	2.4	11.8	18.9	5.4	3.6	0.9	3.4
Zika	0.0	0.7	5.8	5.7	1.5	0.4	0.0	0.4
JEV	2.1	2.8	-	-	-	-	-	-
O'nyong'nyong	-	-	0.0	0.0	0.0	0.0	0.0	0.0

Blood cultures

Overall, blood cultures positive for any pathogen were low across all sites (Table 7.14). The exception, is for the pathogen group Typhoidal *Salmonella* where a substantial proportion of outpatients in Zimbabwe were positive. 11.7% of child outpatients and 25.2% of adult outpatients in Zimbabwe tested positive for Typhoidal *Salmonella* but only 3.0% of child inpatients and 1.6% of adult inpatients tested positive also for Typhoidal *Salmonella*. Also of note, both adult strata in Mozambique had 3% of those cases tested positive for *Escherichia coli* (*E.coli*). As no control specimens were examined by blood culture, there are no positive test results to report.

Table 7.14: Percentage of patients (In- and Out-) that tested positive for different pathogens via blood culture, shown by site and age group

	Laos				Malawi				Mozambique				Zimbabwe			
	< 15		15+		< 15		15+		< 15		15+		< 15		15+	
	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out	In	Out
Typhoidal *Salmonella*	0.7	0.3	0.8	0.5	0.6	0.2	2.3	0.8	0.2	0.0	0.5	0.0	3.0	11.7	1.6	25.2
Non-typhoidal *Salmonella*	0.0	0.0	0.0	0.0	1.2	0.5	0.0	0.3	0.2	0.0	1.1	0.4	0.0	0.0	0.0	0.2
Talaromyces marneffei	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Escherichia coli	0.0	0.0	1.0	0.3	0.0	0.2	1.4	0.3	0.6	0.7	3.5	2.9	0.0	0.2	0.9	1.8
Enterobacterales other than *E.coli*	0.0	0.0	0.8	0.2	0.0	0.2	0.5	0.0	0.6	0.0	1.3	0.2	0.3	0.6	0.5	0.2
Klebsiella pneumoniae	0.0	0.0	0.3	0.0	0.0	0.0	0.5	0.0	0.4	0.0	1.1	0.0	0.3	0.2	0.2	0.0
Burkholderia pseudomallei	0.0	0.3	1.5	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Staphylococcus aureus	0.5	0.3	0.2	0.5	0.6	0.2	0.0	0.0	2.7	0.3	2.4	0.2	0.6	0.2	0.2	0.2
Streptococcus other	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.0	0.0	0.0	0.2
Streptococcus pneumoniae	0.2	0.3	0.0	0.0	0.3	0.0	0.0	0.0	2.1	0.0	0.3	0.7	0.0	0.2	0.2	0.5
Acinetobacter baumannii	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Streptococcus pyogenes	0.2	0.0	0.0	0.0	0.0	0.2	0.5	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Mycobacterium tuberculosis complex	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	1.6	0.0	0.0	0.0	3.2	0.0
Cryptococcus sp.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.2	0.2
H.influenza	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Enterococcus faecalis	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mycobacteria other	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0

Co-infections

Overall, 57.2% (n=5832) of enrolled participants (cases or control) with at least one diagnostic test result had a positive test result for at least one pathogen (a positive test result here is the same definition as throughout this chapter where this means positive on any of the diagnostic methods utilised to investigate a pathogen). Of interest is how many individuals were positive for more than one pathogen and whether there are pairs of pathogens that frequently appear together. An important aspect of this is how many diagnostic tests did each individual have and this is shown in Figure 7.2.

In cases, who had at least one test (n=7842) 62.8% of participants had a positive result for a diagnostic test for at least one pathogen sought. 27.0% had a positive diagnostic test result for two pathogens, 11.6% were positive for three different pathogens and 4.5% four or more positive results (where the denominator for each is the number of cases with at least two (n=7842), three (n=7841) or four (n=7840) diagnostic test results). The patient with the greatest number of positive test results was positive for ten different pathogens: malaria, RSV, adenovirus, rhinovirus/enterovirus, human metapnuemovirus, human bocavirus, parainfluenza, leishmaniasis, chikungunya and zika.

In controls who had at least one diagnostic test (n=2356), 38.4% of participants had at least one positive diagnostic test result. 12.5% of controls had positive diagnostic test results for two pathogens and 4.9% three or more. The denominator for each is the number of controls with at least two diagnostic tests (n=2334) or at least three diagnostic tests (n=2333).

Figure 7.3 highlights which pathogens, if any, appear more frequently together in cases. From Figure 7.3 it is apparent that only a small proportion of cases had positive test results for the same two pathogens. The most common 'co-infections' are malaria and other arboviruses. Another common co-infection appears to be malaria and rhinovirus/enterovirus. While I have referred to this section as co-infections, it is more appropriate to refer to these as co-detections. Multiple pathogens have been detected but they are not necessarily true positives. Investigations by site showed little difference to those for all sites combined.

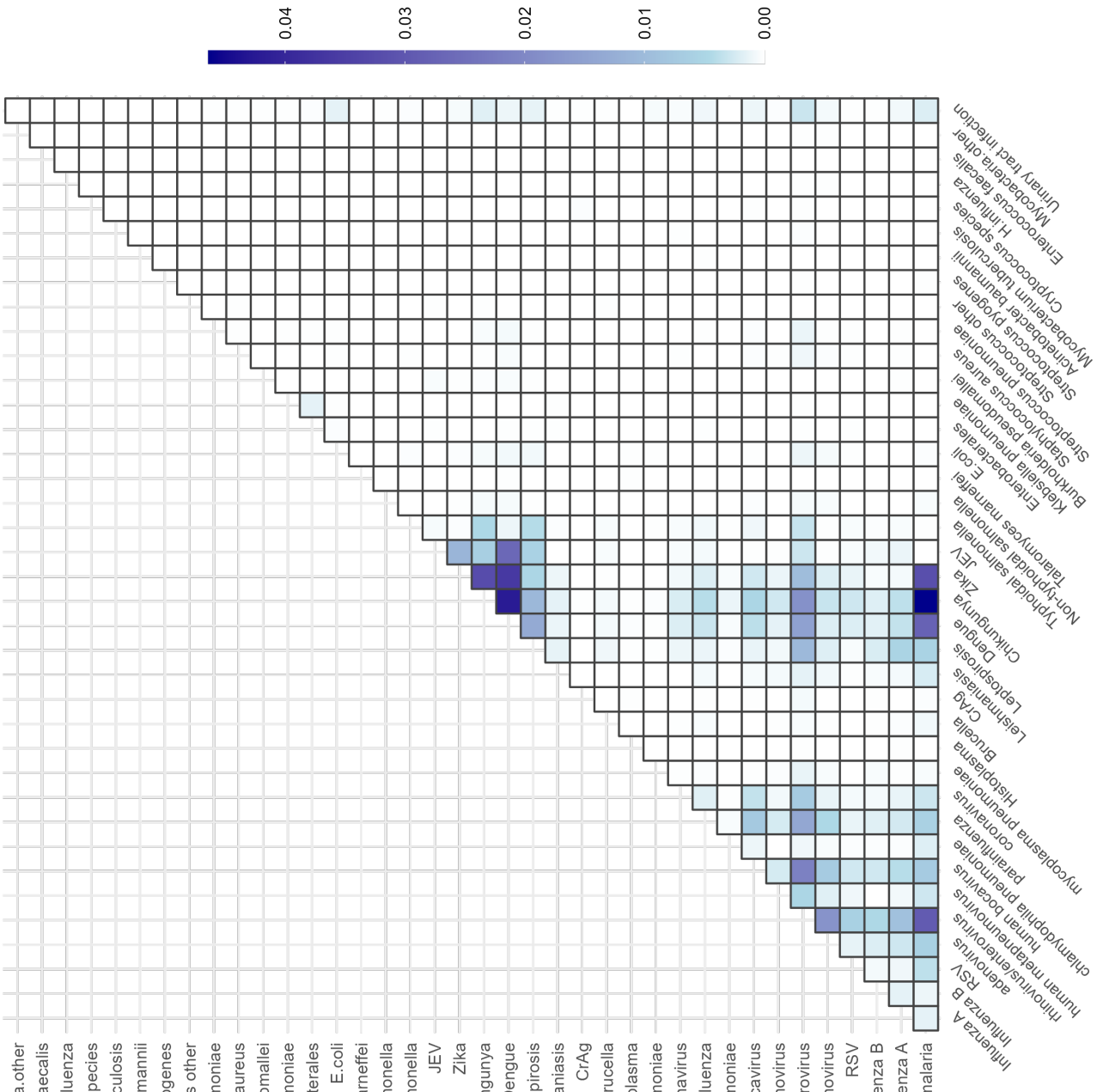


Figure 7.3: Proportion of co-detections in cases for each pair of infections across all sites. Each cell represents the proportion of patients with positive results for both infections out of those patients with test results for both infections

Summary

In this thesis, the results from diagnostic tests form a key input into the final planned analysis. This data consists of at least one diagnostic test result (for any pathogen) from 10198 recruited participants (7842 cases and 2356 controls), 99.5% of those enrolled. There was no pathogen where all participants received at least one diagnostic test. To limit potential biases from missing data, it was decided that only pathogens where at least 75% of all participants were tested would be included in further analyses. Those pathogens where less than 75% of participants received a diagnostic test include: Histoplasma, CrAg, leishmaniasis, urinary tract infections, rickettsioses and TB blood cultures (Table 7.6).

Overall, test positivity to any pathogen was lower than anticipated across all sites [172]. Exceptions, where at least 75% of participants were tested and over a third of participants in a strata tested positive for a pathogen include:

1. Rhinovirus/enterovirus in Malawi child outpatients (45.8% - Table 7.8)
2. Rhinovirus/enterovirus in Malawi child controls (38.1% - Table 7.9)
3. Malaria in Malawi child inpatients and outpatients (50.3% and 39.5% respectively - Table 7.10)
4. Malaria in Mozambique inpatients (47.3% - Table 7.10)

As a result of low test positivity, pathogens where at least 75% of participants were tested but less than five participants tested positive were also excluded from further analyses. These were: the respiratory pathogen *Legionella pneumophila* and pathogens detected by blood culture *Talaromyces marneffei*, *Streptococcus pyogenes*, *Acinetobacter baumannii*, *Haemophilus influenzae*, *Cryptococcus* species and *Enterococcus faecalis*.

In general, co-infections did not appear to be common and the results from exploratory analyses of co-infections led to no further exploration of co-infections in this thesis.

Chapter 8

Bayesian estimation of the causes of fever

8.1 Preamble

In this chapter, I present the statistical methods used to estimate the aetiology of fever in FIEBRE study sites and the results from this analysis. The approach taken relies on the latent class model framework and is an extension of the analysis used in the PERCH study [37] applied to FIEBRE to estimate the fraction of fever cases due to different fever-causing pathogens. Key differences between the analysis used here and that in PERCH are:

i) No diagnostic tests are assumed to have 100% sensitivity and specificity (called gold standard in the PERCH analysis) or 100% specificity (called silver standard in the PERCH analysis)

ii) Priors for diagnostic test sensitivity and specificity are derived from diagnostic test accuracy meta-analyses where possible as opposed to the consensus of an expert working group

Here I present two methods. I start with the simpler case where each pathogen is considered in isolation. Under this approach, for each pathogen, I consider a model in which it is assumed that all fever cases fall into one of two latent states - fever is either caused by that pathogen or is not. I refer to these as *single-pathogen models*. The second approach extends this by considering all pathogens in a single model. The number of latent states in this *multi-pathogen model* is equal to the number of pathogens investigated plus one category to capture 'other causes'.

Separate analyses are run for All patients, Inpatients and Outpatients combined and within each of these groups, analyses are further stratified by site (Laos, Malawi, Mozambique and Zimbabwe) and age group (0

to 14 years referred to as children, and 15 years and over referred to as adults). All results are presented with the median and 95% credible interval (CrI).

8.2 Model Specification

8.2.1 Notation

As in the diagnostic test accuracy application, I assume the true cause of a case's fever is unknown. In this fever aetiology application, the relative probabilities of fevers having been caused by particular pathogens, is estimated using data from multiple different diagnostic tests for different treatable and/or preventable causes of fever. The latent indicator Z_i now represents the true pathogen causing fever for case i, taking possible values $1, ..., J+1$, where J denotes the number of pathogens considered as potential causes of fever. A case's fever may be caused by something other than the pathogens sought and so I consider an additional category representing when a case's fever is caused by something "Not Specified". Hence, the number of possible latent subgroups is $J+1$. This applies to both single- and multi-pathogen models but in the single-pathogen model, it is always the case that $J = 1$.

Let K_j denote the number of diagnostic tests used to detect a pathogen j ($j = 1, .., J$). In the FIEBRE data, K ranges from one to four. Then, let $\mathbf{M}_j$ denote the matrix of test results for pathogen j with dimensions NxK_j, where N refers to participants $i = 1, ..., N$. Y^i indicates whether a participant is a case ($Y^i = 1$) or control ($Y^i = 0$). The probability of pathogen j being the cause of a case's fever is given by $\pi_j, j = 1, ..., J+1$ where, by the axioms of probability, $\sum_{j=1}^{J+1} \pi_j = 1$. As in Wu *et al.* [168] I let ψ_j, denote the false positivity rate (1 minus the specificity) for the test for pathogen j however, I add a further subscript to accommodate multiple tests for the same pathogen. Thus, ψ_{j_k} where k takes the value $k = 1, .., K_j$, denotes the false positivity rate for a test for pathogen j and θ_{j_k}, $k = 1, ..., K_j$, to denote the sensitivity. π_j, ψ_{j_k} and θ_{j_k} can be written as the probabilities:

$$\begin{aligned} \pi_j &= \Pr(Z_i = j | Y_i = 1), j = 1, ..., J+1 \\ \psi_{j_k} &= \Pr(\mathbf{M}_{ij} = 1 | Z_i \neq j), j = 1, ..., J \\ \theta_{j_k} &= \Pr(\mathbf{M}_{ij} = 1 | Z_i = j), j = 1, ..., J \end{aligned} \tag{8.2.1}$$

The FIEBRE study design means that the sensitivity and specificity have an interpretation as the probability of detecting a positive/negative test

result for a truly diseased/non-diseased individual who has a fever and sought care. The estimand in this Bayesian analysis is π_j, which is defined as the probability that a cases fever is caused by pathogen j given they are a case. This definition takes into account diagnostic test accuracy (this is different to η in Chapter 6 which is the observed proportion of fever cases that tested positive for pathogen j). Although estimated by a different methodology, π_j has the same interpretation as the PAF introduced in Chapter 6 (equation 6.2.2). That is, it is the proportion of fever cases attributed to pathogen j or equivalently, the proportion of fever cases that would theoretically be prevented if pathogen j were eliminated.

8.2.2 Likelihood

The FIEBRE study aimed to investigate febrile illness in paediatric and adult outpatients and inpatients [20]. Controls have no pathogen that caused a fever that led them to seek health care. Based on the study design we can therefore assume that any positive test result for a control is a false-positive for *infection causing fever and health-care seeking*. Controls can have a fever or even asymptomatic or sub-clinical infection. However, by being recruited as controls they did not meet the criteria for inclusion as a case. Making use of this assumption that any positive result in a control represents a false-positive, the contribution to the multi-pathogen model likelihood from a control is:

$$P_i^0 = \Pr(\mathbf{M}_i = \mathbf{m}_i|\psi) = \prod_{j=1}^{J}\sum_{k=1}^{K_j}(\psi_{j_k})^{m_{ij_k}}(1-\psi_{j_k})^{1-m_{ij_k}} \tag{8.2.2}$$

In the single-pathogen model, for pathogen j, this reduces to:

$$P_i^0 = \Pr(\mathbf{M}_{ij} = \mathbf{m}_{ij}|\psi_j) = \sum_{k=1}^{K_j}(\psi_{j_k})^{m_{ij_k}}(1-\psi_{j_k})^{1-m_{ij_k}} \tag{8.2.3}$$

which is of the same form as equation 2.4.4 from the diagnostic test accuracy single study scenario. I assume a single cause of infection, such that, for a case truly infected by pathogen j, positive test results for pathogens other than j are false positive results with probability equal to ψ_{j_k}. Borrowing the false positivity rate estimated from controls (for pathogens where we have control data) allows us to distinguish between pathogens that are the true cause of a health-care seeking fever and those that are just background colonizers. The likelihood contribution from a

case in the multi-pathogen model is then:

$$P_i^1 = \Pr(\mathbf{M}_i = \mathbf{m}_i|\pi, \theta, \psi) = \prod_{j=1}^{J+1} \pi_j \left(\sum_{k=1}^{K_j} ((\theta_{j_k})^{m_{ij_k}} (1-\theta_{j_k})^{1-m_{ij_k}} \prod_{l \neq j} (\psi_{l_k})^{m_{il_k}} (1-\psi_{l_k})^{1-m_{il_k}}) \right)^{I_{j \neq J+1}} \tag{8.2.4}$$

as described in Wu *et al.* [168] where, the component $l \neq j$ comes from the fact that for a case infected with pathogen j, under the assumption of a single cause, no other pathogen can be the true cause and positive test results must be false positives. And, $I_{j \neq J+1}$ indicates that when $j = J+1$, the latent state for the category denoted "Not Specified", there are no observed test results and π_{J+1} is estimated by summing over $\pi_j, j = 1, .., J$ (i.e. $1 - \sum_{j=1}^{J} \pi_j$). It is assumed that given the true pathogen status Z_i, test results for tests of different pathogens are assumed independent e.g. that detection of Dengue in a patient with Dengue does not impact the probability of detecting any other pathogen. This is different to the conditional independence assumption discussed so far applied to tests for the same pathogen. Extensions to relax the local independence assumption have been discussed [173]. This additional element of independence between pathogens is a component that does not need to be considered in the single-pathogen models, when each pathogen is considered in turn.

Similar to equation 8.2.3, the single-pathogen model for pathogen j, the contribution to the likelihood for a case reduces to:

$$P_i^1 = \Pr(\mathbf{M}_{ij} = \mathbf{m}_{ij}|\pi, \theta_j, \psi_j) = \prod_{j=1}^{J+1} \pi_j \left(\sum_{k=1}^{K_j} ((\theta_{j_k})^{m_{ij_k}} (1-\theta_{j_k})^{1-m_{ij_k}}) \right)^{I_{j \neq J+1}} \tag{8.2.5}$$

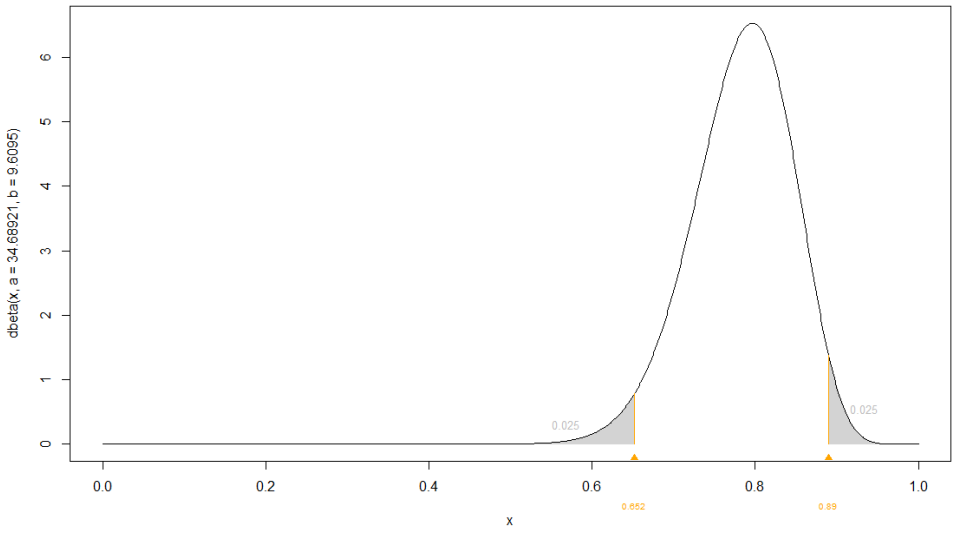
8.2.3 Priors

All parameters in equations 8.2.2 through 8.2.5 (π_j, θ_{j_k} and ψ_{j_k}) require a prior distribution to be specified. The prior for the probability of a pathogen (including the category of "Not Specified") being the cause of a cases fever (π_j) is assumed to follow a Dirichlet distribution that favours no pathogen over another ($\pi_j \sim Dirichlet(1, ..., 1)$). Prior information for diagnostic test sensitivities (θ_{j_k}) is derived based on diagnostic test accuracy meta-analyses and are in the form of a 95% credible interval. Priors for the sensitivity of each diagnostic test are then beta distributions where the 2.5% and 97.5% quantiles of the beta distribution match the 95% credible interval from the prior information [174]. For example, the malaria RDT is assumed to have a sensitivity between 65.2%

and 89.0% with 95% probability. Using these specified quantiles, the computed beta parameters are $\alpha = 34.69$ and $\beta = 9.61$ (See figure 8.1). While initial plans were to perform a meta-analysis for all tests, where no meta-analysis was carried out, I relied on expert opinion to derive a 95% credible interval. Expert elicitation of probability distributions is a complex task. There are papers exploring the different systematic approaches that can be taken to elicit expert opinion [175] however, these methods are outside the scope of this thesis and instead I relied on subjective, expert opinion from a collective group of FIEBRE co-investigators. These potentially biased opinions are paired with sensitivity analyses to assess robustness.

For all specificities, I assume a uniform prior over the interval 0 to 1 (i.e. 0% to 100%) when test results from controls are available. It is assumed that any positive test result from a control participant is a false positive and so the specificity can be estimated directly from the observed data hence a uniform prior distribution can be assumed. However, when data from controls are not available, the same process as for sensitivity is followed, if no meta-analyses was carried out a 95% credible interval is derived from expert elicitation.

Figure 8.1: Beta prior distribution for the sensitivity of the malaria RDT using 2.5 and 97.5% quantiles of 65.2 and 89.0 [174]



8.3 Estimation and model implementation

The primary goal is to estimate the posterior distribution of the cause-specific case fractions $(\pi_j, j = 1, ., , , J + 1)$. Unlike in the diagnostic test

accuracy implementation of LCM, here, the addition of controls allows direct estimation of test specificities $(1 - \psi_{j_k})$ for detecting pathogens that cause health-care seeking fevers. These parameters are therefore identifiable leading to this model being described as a *partially identifiable model* [119] and distinguishing these latent class models from those discussed in Part 1 of this thesis. Even with these identifiable parameters, there are still many non-identified parameters within the model leading to weak identifiability. Unidentified parameters are defined by Garrett and Zeger [176] as parameters where the only information we have about the parameter of interest is supplied by the prior distribution (i.e. from equation 2.4.5 $p(\theta|Y) \approx p(\theta)$). Model identifiability is a challenge in this application for the same reasons as in the diagnostic test accuracy scenario (See Section 2.4.1). Though the Bayesian framework can help with this challenge, identifiability is not guaranteed [60]. This characteristic highlights the importance of the prior distributions used and sensitivity analyses that investigate how robust estimates are to the choice of prior distribution.

Traditional analysis of a matched case-control study commonly involves matching at the analysis stage, however this is not a requirement [177]. In particular, in the analyses presented in this thesis we do not estimate an odds ratio. Variables that were matched on are however controlled for in the analysis by means of stratification.

8.3.1 Subgroup analyses

Separate analyses are run for each participant group (Outpatients and Inpatients) as well as for All patients combined. Within each analysis (Outpatients, Inpatients and All patients), I further stratify by four sites and two age groups. To do this, a variable is generated which indicates to which strata an individual belongs. For example, Laos/child and Mozambique/adult represent two out of eight possible strata considering four sites and two age groups. Within each analysis (Inpatients, Outpatients and All patients) diagnostic test sensitivity is assumed to be the same across strata but it could be allowed to vary if prior information indicated that sensitivity is expected to differ. Specificity on the other hand, is allowed to vary by all strata (site and age group) when control data are available to allow direct estimation of specificity. This means, that $\psi_{j_{k_c}}$, denotes the false positivity rate of a test in stratum c where $c = 1, ..., C$. Then, π_j is also expanded to estimate the probability of pathogen j being the cause of a case in stratum c, $\pi_{j_c} = \Pr(Z_i = j|Y_i = 1, C_i = c)$. Further, within each subgroup analysis (Outpatients, Inpatients and All

patients), the same control participants are used.

8.3.2 Software implementation

As in section 4.5, the unknown parameters are estimated via Markov Chain Monte Carlo methods. All models are implemented using CmdStan in R [122]. Model code for the malaria single-pathogen model (from which all other models can be derived) is included in Appendix E and on github. Models were run using 4 chains [123] and convergence was assessed visually with the *Bayesplot* R package [63]. I only report results where the rank normalised split-R-Hat statistic is < 1.01, the total effective sample size is at least 400 and there are no warnings for energy fraction of missing information or divergent transitions [123]. Where the only warnings reported are divergent transitions, attempts to diagnose the cause of divergences are made and where appropriate, the target acceptance rate is increased to help sampling in areas of the posterior that exhibit high curvature and require small step sizes to explore [178].

8.3.3 Missing data

Not all participants had a test result for all tests carried out (See Figure 7.2). However, a participant does not need complete data (a test result for all tests for all pathogens) to be included in the current analysis. A participant's data are included in the multi-pathogen model if they have an observed test result for at least one included pathogen. A participant's data is only included in the single-pathogen model if they have an observed test result for one of the tests for the pathogen under consideration. Included pathogens are those where at least 75% of all cases have an observed test result. Those pathogens where fewer than 75% of patients have a diagnostic test result for at least one of the diagnostic methods used (JEV, ONNV, histoplasma, CrAg, leishmaniasis, urinary tract infections, rickettsioses and TB blood cultures - Table 7.6), are not included in analyses and I report the number of positive test results only (Table 7.10). In utilising the data in this way, I assume that any missingness is solely a result of the observed case or control status, i.e. given case or control status, test results are missing at random.

The testing strategy for JEV and ONNV included only certain sites and for histoplasma and CrAg included only individuals with particular characteristics. This testing strategy was based on the assumption that these pathogen are not a cause of fever in the sites where testing did not occur or in individuals without that characteristic. From this, I could

have made the assumption that all individuals not tested would have had a negative test result and included these data in the analysis. However, I chose not to make this assumption as there is some evidence to suggest these pathogens may be found in those sites/individuals not tested. For example, CrAg was only tested for in HIV positive individuals but there is evidence of CrAg in HIV negative individuals [179] and JEV was not investigated for in the Africa sites but there has been cases of JEV in Africa [180].

8.4 Results from single-pathogen models

This section presents the results from the single-pathogen models. Considering each potential cause of fever in turn, within each single-pathogen model, two latent subgroups ($Z_i = 1, 2$) are assumed. This represents that the fever is either attributed to the pathogen under consideration or not however, the way this is implemented, means that the results are a probability of membership to each latent subgroup. From the case-control study, the key input into this analysis are the binary test results for each test for each pathogen. The results are presented by pathogen below.

For each pathogen a table is provided describing the diagnostic test/s used, testing strategy and priors as well as a presentation of the primary results by participant group (All patients, Outpatients and Inpatients), site (Laos, Malawi, Mozambique and Zimbabwe) and age group (less than 15 years referred to as children and, 15 years or older referred to as adults). Only those pathogens with at least five positive test results for five individuals and at least 75% of participants tested are included in analyses.

The final part of each single-pathogen model section is a sensitivity analysis. Depending on the pathogen and number of diagnostic tests used, a sensitivity analysis comparing the results from the primary analysis model with either results from a model with a different conditional dependence structure or results from a model that used more diffuse prior distributions is presented. Finally, section 8.4.7 presents a synthesis of the results across the single-pathogen models.

8.4.1 Malaria

Malaria was investigated using two diagnostic tests; a rapid diagnostic test (RDT) and microscopy (Table 8.1). Although microscopy was used at all sites, ultimately, only the microscopy data from three sites (Laos,

Mozambique and Zimbabwe) was deemed of high enough quality (when compared to reference laboratory microscopy) to be included in analyses. This means that there are no control participant results from microscopy to estimate the specificity in Malawi and instead an informative prior assuming the specificity is between 95% and 99.9% with 95% probability is used. The malaria RDT was carried out on 98% of participants while the microscopy testing strategy meant many fewer (31%) were tested by that method. 3155 individuals (31%) had a test result for both tests.

Table 8.1: A summary of model inputs (observed diagnostic test results and accuracy priors) for malaria

	Test type	
	RDT	**Microscopy**
Test details	Hrp2 and/or pan positive on day 0 sample	Two readers, third reader if disagreement from day 0 samples
Testing strategy	All	All RDT positive and a 10% sample of RDT negative
Number tested (%)[1]	10004 (97.6%)	3167 (30.9%)
Prior sensitivity[2]	65.2%-89.0%[4] [181]	0-100%
Prior specificity[3]	0-100%	0-100%; Malawi 95-100%

[1] % of enrolled participants
[2] The same sensitivity priors are used for all sites and age groups unless specified.
[3] Specificity is allowed to vary by site and age group where data on controls is available.
[4] Sensitivity is assumed to be between 65.2% and 89.0% with 95% probability.

The primary analysis model for malaria included results from both microscopy and the RDT and accounted for conditional dependence between the two tests in individuals with malaria. Table 8.2 presents the results from this primary analysis model. In each analysis (All patients, Inpatients and Outpatients) in both adults and children, malaria is estimated to have little to no contribution to the causes of fever in Laos and 5% or less in Zimbabwe. In all children (inpatients and outpatients combined), just over half (53%) of fever cases in Malawi are attributed to malaria with a 95% credible interval ranging from 35% to 70%. Considering only children from the Outpatient analysis, the estimate is lower (42%, 95% CrI: 30-69%) and considering only Inpatients, the estimate is higher (74%, 95% CrI: 55-92%). A similar trend is seen in Mozambique where fewer fever cases are attributed to malaria in Outpatients (6%, 95% CrI: 4-8%) compared with Inpatients (70%, 95% CrI: 58-81%).

A marked difference between adults and children is visible in Malawi and Mozambique particularly in Inpatients. One reason for this may be the local malaria endemicity. In malaria endemic settings, older chil-

dren and adults develop immunity to severe and even mild malaria after repeated infections and so are less likely to be seeking care for a fever caused by malaria [182]. In the current analysis, these sub-clinical or asymptomatic cases of malaria infection are not accounted for.

A sensitivity analysis compares the estimates from the primary analysis model, that accounts for conditional dependence between malaria diagnostics in disease positive individuals, to estimates from a model that assumes conditional independence between malaria diagnostics, given the true disease status of participants. Figure 8.2 shows how each of those estimates presented in Table 8.2 differs by assuming a more simple model of conditional independence between the results of these tests. In general, we would expect the results from a model accounting for conditional dependence to have wider credible intervals as a result of taking into account the uncertainty in individual-level test sensitivity.

Results in adults show little difference between estimates from a model that assumes conditional independence to a model relaxing the conditional independence assumption. Credible intervals are generally narrower in the conditionally independent model as expected from not taking into account the additional uncertainty in individual-level test sensitivity. Point estimates from the conditionally independent model also appear to be shifted slightly lower. This may reflect an overestimation in diagnostic test sensitivity from the conditionally independent model which impacts the percentage of fever cases attributed to malaria. In children, little difference is also seen between these two models in Zimbabwe and Laos. In Malawi and Mozambique, there is a small difference in All patients and a large difference in Inpatients. This was due to larger differences in the estimated sensitivity of the diagnostic tests used in the analyses.

Table 8.2: Median (95% credible interval) estimate of the percentage of fevers attributed to malaria

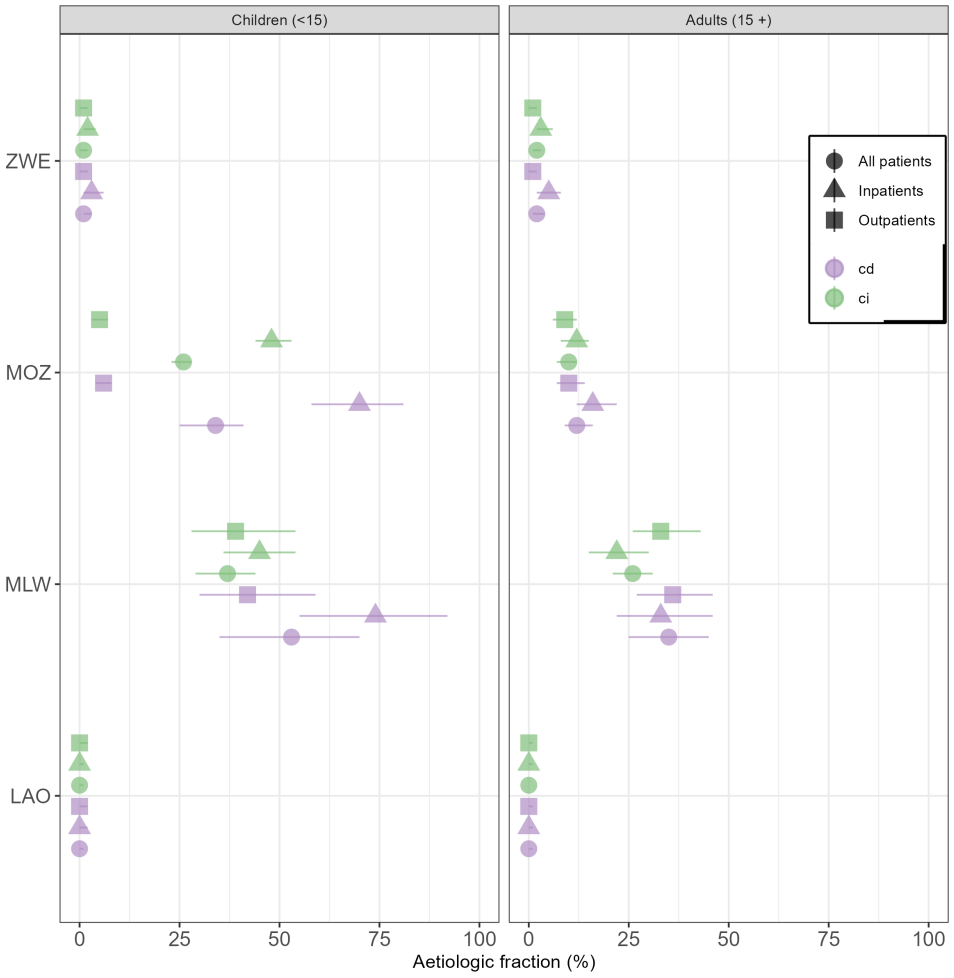
(a) Estimates in children aged less than 15 years

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
All patients	0 (0-1)	53 (35-70)	34 (25-41)	1 (1-3)
Outpatients	0 (0-2)	42 (30-59)	6 (4-8)	1 (0-2)
Inpatients	0 (0-2)	74 (55-92)	70 (58-81)	3 (1-6)

(b) Estimates in adults aged 15 years or more

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
All patients	0 (0-1)	35 (25-45)	12 (9-16)	2 (1-4)
Outpatients	0 (0-1)	36 (27-46)	10 (7-14)	1 (0-2)
Inpatients	0 (0-1)	33 (22-46)	16 (12-22)	5 (2-8)

Figure 8.2: A comparison of the estimated percentage of fever cases attributed to malaria by site, age, participant status group and, whether or not conditional dependence was accounted for between malaria diagnostic tests



Note. cd = conditional dependence (estimates arise from a model that assumes conditional dependence between test results given disease status), ci = conditional independence (estimates arise from a model that assumes conditional independence between test results given disease status)

8.4.2 Respiratory Infections

This section presents the results from single-pathogen models for respiratory pathogens investigated by the Luminex respiratory pathogen panel (RPP) (See Table 8.4). For the single-pathogen models (and all subsequent analyses) certain respiratory pathogens were grouped, these are: coronaviruses (coronavirus 229e, coronavirus hku1, coronavirus nl63, coronavirus oc43), Influenza A (Influenza AH1 and Influenza AH3), parainfluenza's (parainfluenza 1, parainfluenza 2, parainfluenza 3 and parainfluenza 4) and RSV (RSV A and RSV B). New groups were made using an and/or rule to determine positive test results.

Meta-analyses were carried out to estimate the accuracy of the Lu-

minex RPP to detect Influenza A, Influenza B and RSV, however, no meta-analysis was carried out for the other pathogens. Instead, prior information on the test sensitivity was provided by the Luminex RPP manufacturer [183]. The priors for all respiratory pathogens were narrow ranges (less than 10% percentage points between 2.5% and 97.5% quantiles). Sensitivity analyses for the respiratory pathogens included using more diffuse priors assuming 70% to 99% with 95% probability. On the Luminex RPP, detection of rhinovirus and enterovirus is not differentiated. Further PCR tests were used to differentiate these two viruses only and secondary analyses that estimate the percentage of fever cases attributed to rhinovirus and enterovirus separately is presented in the sub-section titled **Rhinovirus/Enterovirus**.

Table 8.3: A summary of model inputs for respiratory pathogens investigated by the Luminex Respiratory Pathogen Panel

	Test type	
	Luminex RPP[1]	**PCR**
Test details	Day 0 respiratory samples	Day 0 samples
Testing strategy	All	All Luminex positive for rhinovirus/enterovirus
Number tested (%)[2]	9335 (91.1%)	1691 (16.5%)
Prior sensitivity	90-99%[3] [183]	70-99% [184]
Prior specificity	0-100%	0-100%

[1] Respiratory Pathogen Panel
[2] % of enrolled participants
[3] Sensitivity is assumed to be between 90.0% and 99.0% with 95% probability for each pathogen separately

The estimated percentage of fever cases attributed to each respiratory pathogen in children is presented in Table 8.4. Fever cases are consistently, across all sites attributed to Influenza A and Influenza B more than any other respiratory pathogen. Interestingly, this burden is greater in Outpatients than Inpatients. For example, in Laos 14% (95% CrI: 10-19%) of fevers are attributed to Influenza A in Outpatients and only 5% (95% CrI: 3-9%) in Inpatients. Likewise for Influenza B, in Mozambique 12% (95% CrI: 10-16%) of fever cases are attributed to Influenza B in Outpatients and only 1% (95% CrI: 0-3%) in Inpatients. In contrast, fever cases are attributed to a greater proportion of RSV in Inpatients than Outpatients across African sites. While fever cases are attributed to similar percentages of some pathogens (for example, adenovirus and coronavirus) in children across all sites, others show marked differences. Rhinovirus/enterovirus is estimated to have been responsible for between 0 and 3% of fevers in Laos and Mozambique but 13% (95% CrI: 3-24%) in Malawi. Similarly, parainfluenza is estimated to be responsible for less

than 6% of fevers in all sites and participant groups in Laos, Mozambique and Zimbabwe but 9% of fever cases in Outpatient children in Malawi.

The estimated percentage of fever cases attributed to each respiratory pathogen in adults is presented in Table 8.5. Similarly to children, in adults the greatest proportion of fever cases is consistently attributed to Influenza A across all sites. However, on the whole a smaller percentage of fever cases are attributed to respiratory pathogens in adults than children. In contrast to fever cases in children, where rhinovirus/enterovirus is estimated to have been responsible for very few fevers in Mozambique, in adults rhinovirius/enterovirus is estimated to be responsible for 9%, 13% and 6% of fever cases in All patients, Outpatients and Inpatients respectively.

Sensitivity analyses for the respiratory pathogens compare the results from a model that assumed the Luminex RPP had a sensitivity of 90% to 99% with 95% probability to the results from a model that assumed a more diffuse prior of 70% to 99% with 95% probability. Figures 8.3, F.1 and F.2 show how the estimates presented in tables 8.4 and 8.5 differ when a more diffuse prior distribution for test sensitivity is used. In general, we would expect the results from a model using a more diffuse prior to have wider credible intervals (i.e. more uncertainty) around estimates. In Outpatients, our sensitivity analyses shows wider credible intervals for almost all estimates when a more diffuse prior is used (Figure 8.3). The point estimate is also shifted higher with a more diffuse prior. Similar findings are present between age groups and in Inpatients as well as All patients (See Appendix F).

Table 8.4: Median (95% credible interval) estimate of the percentage of fevers attributed to respiratory pathogens in all children (All), outpatient children (Out) and inpatient children (In)

		Laos	**Malawi**	**Mozambique**	**Zimbabwe**
Influenza A	All	9 (7-12)	7 (5-9)	5 (3-7)	8 (5-10)
	Out	14 (10-19)	8 (6-11)	8 (5-11)	7 (4-10)
	In	5 (3-9)	2 (0-5)	2 (0-4)	9 (5-13)
Influenza B	All	6 (4-9)	3 (1-4)	7 (6-9)	7 (3-9)
	Out	13 (9-18)	3 (2-5)	12 (10-16)	9 (6-13)
	In	1 (0-3)	2 (0-5)	1 (0-3)	3 (0-5)
RSV[1]	All	1 (0-2)	1 (0-3)	4 (2-6)	4 (1-6)
	Out	1 (0-4)	1 (0-3)	4 (1-6)	2 (0-3)
	In	1 (0-2)	3 (0-8)	5 (2-9)	7 (3-10)
Adenovirus	All	3 (2-5)	3 (0-7)	6 (3-9)	5 (1-8)
	Out	2 (0-4)	4 (0-9)	9 (5-13)	4 (1-7)
	In	5 (2-7)	2 (0-7)	3 (0-7)	6 (1-10)
Rhinovirus/Enterovirus	All	1 (0-5)	6 (0-14)	1 (0-5)	5 (0-12)
	Out	2 (0-7)	13 (3-24)	1 (0-5)	7 (1-13)
	In	2 (0-7)	2 (0-10)	3 (0-10)	7 (1-15)
Human Bocavirus	All	1 (0-2)	1 (0-4)	3 (0-5)	3 (1-6)
	Out	1 (0-3)	2 (0-6)	2 (0-4)	3 (1-6)
	In	1 (0-3)	1 (0-4)	5 (1-8)	3 (1-7)
ParaInfluenza[2]	All	5 (2-7)	6 (2-10)	4 (1-5)	2 (0-5)
	Out	4 (1-7)	9 (5-14)	3 (1-6)	3 (0-6)
	In	6 (3-9)	1 (0-5)	4 (1-6)	2 (0-5)
Coronavirus[3]	All	0 (0-2)	1 (0-4)	1 (0-3)	1 (0-3)
	Out	1 (0-2)	2 (0-4)	1 (0-4)	1 (0-3)
	In	1 (0-2)	1 (0-4)	1 (0-3)	2 (0-4)
Human metapneumovius	All	1 (0-1)	3 (0-5)	1 (0-3)	1 (0-2)
	Out	0 (0-1)	4 (1-7)	2 (0-3)	1 (0-2)
	In	1 (0-3)	1 (0-4)	2 (0-4)	1 (0-3)
Chlamydophila pneumonia	All	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)
	Out	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)
	In	0 (0-1)	1 (0-3)	1 (0-3)	1 (0-2)
Mycoplasma pneumonia	All	0 (0-1)	0 (0-1)	1 (0-1)	0 (0-1)
	Out	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)
	In	0 (0-1)	0 (0-2)	0 (0-1)	1 (0-2)

[1] RSV includes RSV A and RSV B

[2] Parainfluenza includes Parainfluenza 1, Parainfluenza 2, Parainfluenza 3 and Parainfluenza 4

[3] Coronavirus includes Coronavirus 229e, Coronavirus hku1, Coronavirus nl63 and Coronavirus oc43

Table 8.5: Median (95% credible interval) estimate of the percentage of fevers attributed to respiratory pathogens in all adults (All), outpatient adults (Out) and inpatient adults (In)

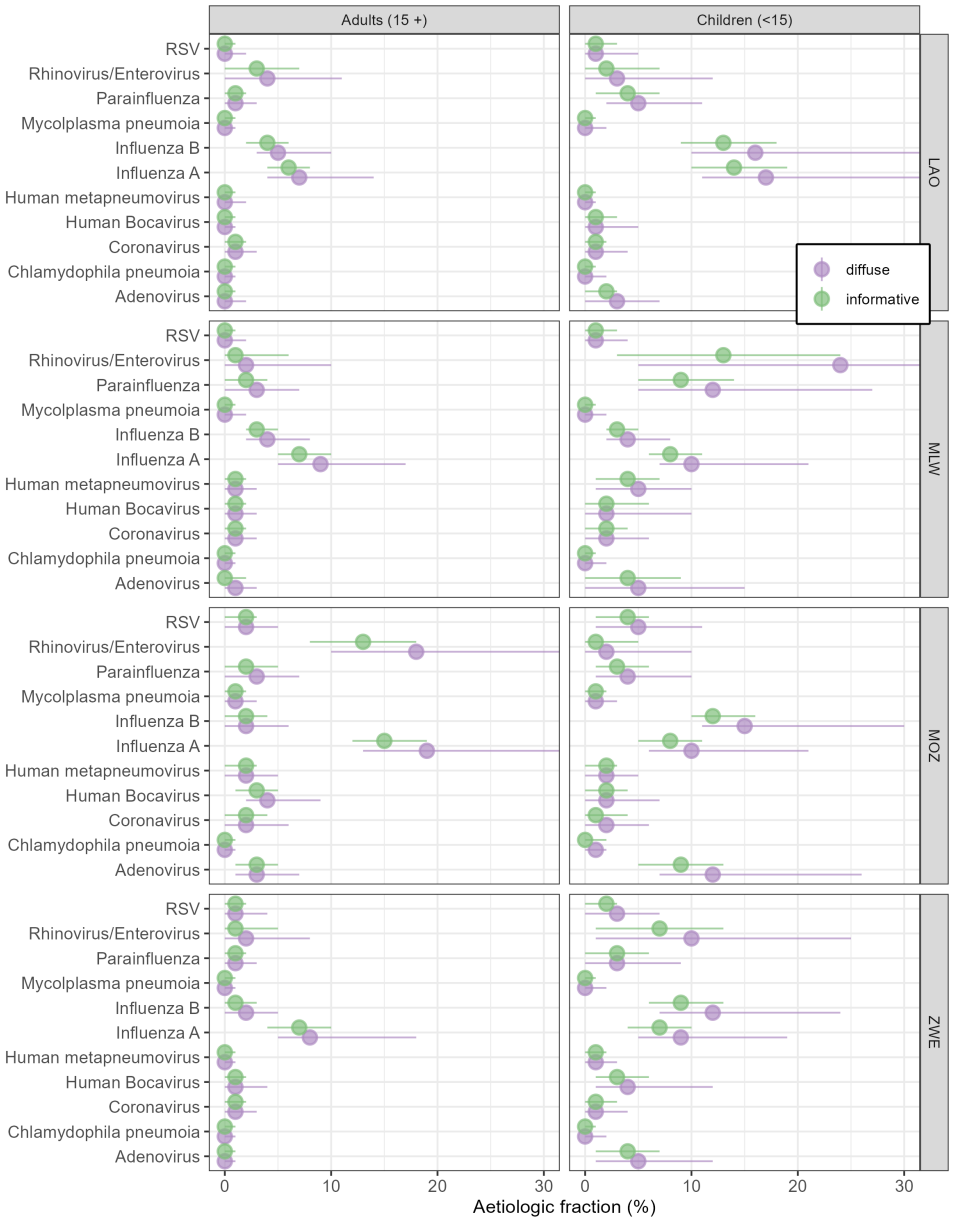
		Laos	**Malawi**	**Mozambique**	**Zimbabwe**
Influenza A	All	4 (3-5)	6 (4-9)	10 (8-12)	4 (2-6)
	Out	6 (4-8)	7 (5-10)	15 (12-19)	7 (4-10)
	In	3 (1-4)	4 (1-8)	4 (2-6)	2 (0-4)
Influenza B	All	3 (1-4)	2 (1-4)	1 (0-2)	1 (0-2)
	Out	4 (2-6)	3 (2-5)	2 (0-4)	1 (0-3)
	In	2 (0-3)	1 (0-3)	0 (0-1)	1 (0-3)
RSV[1]	All	0 (0-1)	0 (0-1)	1 (0-2)	1 (0-2)
	Out	0 (0-1)	0 (0-1)	2 (0-3)	1 (0-2)
	In	0 (0-1)	1 (0-3)	1 (0-2)	0 (0-2)
Adenovirus	All	0 (0-1)	0 (0-2)	2 (1-3)	0 (0-1)
	Out	0 (0-1)	0 (0-2)	3 (1-5)	0 (0-1)
	In	0 (0-1)	1 (0-3)	1 (0-3)	0 (0-1)
Rhinovirus/Enterovirus	All	2 (0-4)	1 (0-4)	9 (4-12)	2 (0-6)
	Out	3 (0-7)	1 (0-6)	13 (8-18)	1 (0-5)
	In	1 (0-4)	1 (0-5)	6 (1-10)	6 (2-12)
Human Bocavirus	All	0 (0-1)	0 (0-2)	2 (1-3)	1 (0-3)
	Out	0 (0-1)	1 (0-2)	3 (1-5)	1 (0-2)
	In	0 (0-1)	1 (0-4)	1 (0-3)	2 (0-4)
ParaInfluenza[2]	All	1 (0-1)	2 (0-4)	1 (0-3)	1 (0-3)
	Out	1 (0-2)	2 (0-4)	2 (0-5)	1 (0-2)
	In	0 (0-1)	2 (0-5)	1 (0-5)	2 (0-4)
Coronavirus[3]	All	1 (0-2)	0 (0-2)	1 (0-3)	1 (0-2)
	Out	1 (0-2)	1 (0-2)	2 (0-4)	1 (0-2)
	In	1 (0-2)	1 (0-4)	1 (0-3)	2 (0-4)
Human metapneumovius	All	0 (0-1)	1 (0-2)	1 (0-2)	0 (0-1)
	Out	0 (0-1)	1 (0-2)	2 (0-3)	0 (0-1)
	In	0 (0-1)	1 (0-5)	0 (0-2)	0 (0-2)
Chlamydophila pneumonia	All	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
	Out	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
	In	0 (0-1)	1 (0-3)	0 (0-1)	0 (0-1)
Mycoplasma pneumonia	All	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
	Out	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)
	In	0 (0-1)	1 (0-3)	0 (0-1)	0 (0-1)

[1] RSV includes RSV A and RSV B

[2] Parainfluenza includes Parainfluenza 1, Parainfluenza 2, Parainfluenza 3 and Parainfluenza 4

[3] Coronavirus includes Coronavirus 229e, Coronavirus hku1, Coronavirus nl63 and Coronavirus oc43)

Figure 8.3: A comparison of the estimated percentage of fever cases attributed to respiratory pathogens in Outpatients using a more informative or diffuse prior for diagnostic test sensitivity by site (LAO=Laos, MLW=Malawi, MOZ=Mozambique, ZWE=Zimbabwe) and age group.



Rhinovirus/Enterovirus

On the Luminex RPP, detection of rhinovirus and enterovirus is not differentiated. A further PCR test was used to differentiate these two viruses only in people with a positive Luminex test Rhinovirus/enterovirus. The previous results (Table 8.4 and 8.5) estimate the percentage of fevers attributed to rhinovirus/enterovirus ($\pi_{rhino/entero}$) however, assuming that $\pi_{rhino/entero} = \pi_{rhinovirus} + \pi_{enterovirus}$ and utilising the additional PCR test results, we can estimate the percentage of fevers attributed to rhinovirus and enterovirus separately.

Table 8.6: Median (95% credible interval) estimate of the percentage of fevers attributed to Rhinovirus/Enterovirus (Rhino/Entero) as well as Rhinovirus and Enterovirus separately in all patients

(a) Estimates in children aged less than 15 years

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
Rhino/Entero	1 (0-5)	6 (0-14)	1 (0-5)	5 (0-12)
Rhinovirus	0 (0-1)	3 (0-8)	1 (0-3)	3 (0-6)
Enterovirus	0 (0-3)	1 (0-4)	0 (0-1)	2 (0-4)

(b) Estimates in adults aged 15 years or more

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
Rhino/Entero	2 (0-4)	1 (0-4)	10 (5-14)	3 (0-7)
Rhinovirus	0 (0-1)	0 (0-2)	7 (3-9)	2 (0-4)
Enterovirus	0 (0-2)	0 (0-0)	2 (1-3)	0 (0-1)

In a parallel model to the rhinovirus/enterovirus single-pathogen model, I ran a modified version of this model, where $J = 2$, representing rhinovirus or enterovirus, $Z_i = 1, .., J + 1$ and there is a single PCR test for each j. Using this information we can re-scale the estimates from this parallel model according to our estimate of $\pi_{rhino/entero}$ from the single-pathogen model. Table 8.6 shows the results for $\pi_{rhino/entero}$, $\pi_{rhinovirus}$, and $\pi_{enterovirus}$ separately for All patients.

Taking advantage of the additional PCR tests in this way highlights the flexible framework of latent class models. Incorporating this additional information shows that in strata where fever cases are attributed to rhinovirus/enterovirus, it is rhinovirus and not enterovirus that is the cause (Table 8.6).

8.4.3 Brucella

Brucella was investigated using two different diagnostic tests, Brucellacapt was a single-step screening test carried out on all individuals (78.9% in practice), followed by Microscopic agglutination (MAT) on those participants with a positive test for Brucellacapt (Table 8.7). The MAT test works by comparing the test result from MAT on a day 0 sample with a second sample collected at day 28. However, not all individuals returned for a day 28 sample (68% of those with a MAT on day 0 had a MAT on day 28 sample) and so we have some participants with just a single acute sample tested by MAT and some individuals with the result of the paired day 0 and day 28 sample. The single sample MAT is not how the test is advised to be used and is treated as a different test to the results of MAT paired as these results would have a different expected accuracy (sensitivity and specificity). While in isolation the results of a

single acute sample tested by MAT might not be helpful, in combination with the results from PCR and MAT paired these results can still be utilised in a latent class model. Day 28 samples were only requested for cases and so there are no control data to estimate the specificity of MAT on paired samples. Instead, this test is assumed highly specific with a prior specificity of 90-99%.

Table 8.7: A summary of model inputs (observed diagnostic test results and accuracy priors) for brucella

	Test type		
	BrucellaCapt	**MAT[1] acute**	**MAT paired**
Test details	Day 0 samples	Day 0 sample titre 1:160	Day 0 and day 28 samples
Testing strategy	All	All Brucellacapt positive	All Brucellacapt positive
Number tested (%)[2]	8077 (78.9%)	69 (0.7%)	47 (0.5%)
Prior sensitivity	68%-99%[3]	55-99%	90-99%
Prior specificity	0-100%	0-100%	90-99%

[1] MAT, microscopic agglutination test
[2] % of enrolled
[3] Sensitivity is assumed to be between 68% and 99% with 95% probability.

The primary analysis model for brucella included results from three different diagnostic tests and considered these tests independent. In all analyses, in both adults and children, brucella is estimated to have little to no contribution to the causes of fever in FIEBRE study sites (Table 8.8). Exploring the impact of the prior distributions, a sensitivity analysis that used a more diffuse prior distribution (70-99%, with 95% probability) for the sensitivity of Brucellacapt found similar results with greater uncertainty (Figure 8.4).

Table 8.8: Median (95% credible interval) estimate of the percentage of fevers attributed to brucella

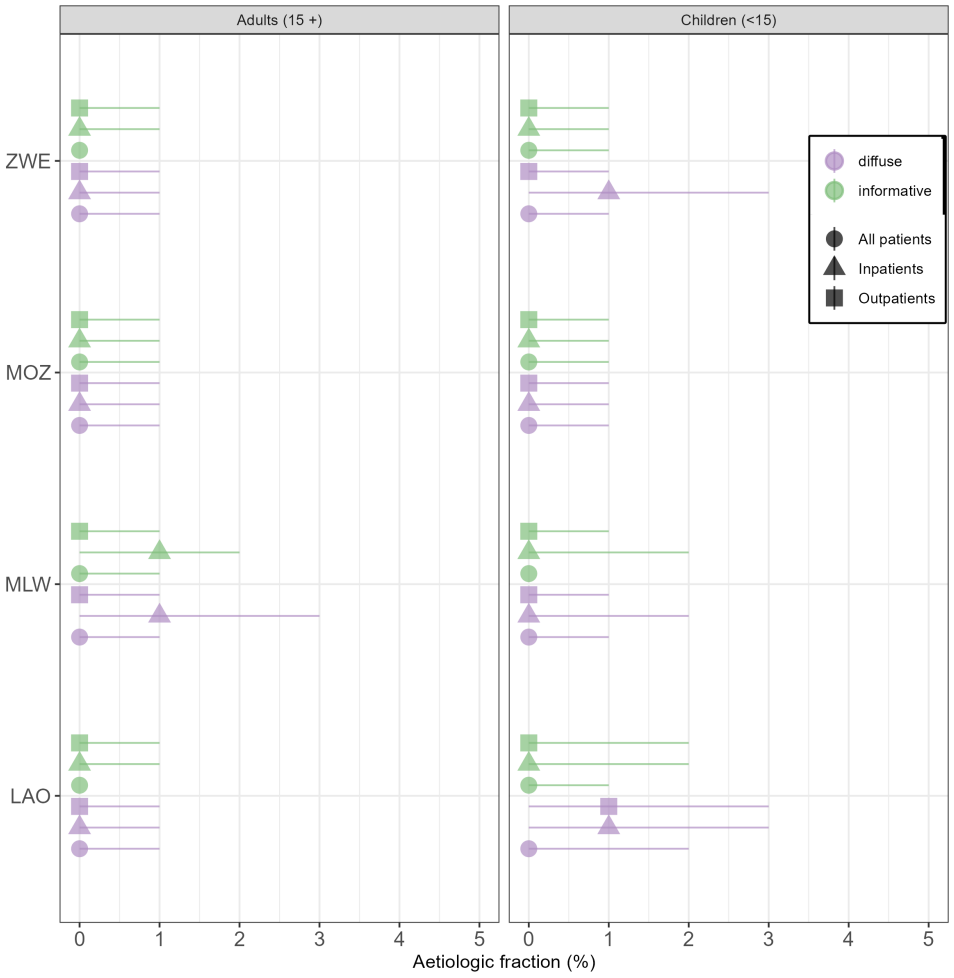
(a) Estimates in children aged less than 15 years

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
All patients	0 (0-1)	0 (0-0)	0 (0-1)	0 (0-1)
Outpatients	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)
Inpatients	0 (0-2)	0 (0-2)	0 (0-1)	0 (0-1)

(b) Estimates in adults aged 15 years or more

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
All patients	0 (0-0)	0 (0-1)	0 (0-1)	0 (0-0)
Outpatients	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Inpatients	0 (0-1)	1 (0-2)	0 (0-1)	0 (0-1)

Figure 8.4: A comparison of the estimated percentage of fever cases attributed to brucella by site, age, participant status group and, whether or not an informative or diffuse prior distribution was used for the Brucellacapt test



8.4.4 Arboviruses

Arboviruses included in this analysis are: dengue, zika, chikungunya and JEV. JEV was only investigated in Laos but in all other regards, the diagnostic tests used and the testing strategy implemented is the same across sites (Table 8.9). Although IgM ELISA is used on a large proportion of enrolled participants, a single sample IgM ELISA carried out early in symptom onset is not sensitive [131]. Hence, the prior for IgM ELISA test sensitivity is low. IgM ELISA detects antibodies that may not yet be detectable in samples depending on how many days post onset of symptoms a participant showed up to a health facility.

Overall, from the arbovirus single-pathogen models, the only arbovirus that fever cases were attributed to in FIEBRE study sites is dengue. In Laos only, fever cases are attributed to dengue in between 9 and 10% of cases in children (Table 8.10) and between 15 and 23% in adults (Table 8.11). Sensitivity analyses considered that IgM ELISA

test sensitivity was between 10 and 80%, with 95% probability and that IgG ELISA sensitivity was between 70 and 99%, with 95% probability. Figures 8.5, 8.6 and 8.7 show how the estimated percentage of fever cases attributed to arboviruses from a model with these alternative prior distributions for test sensitivities differs to the primary analysis model.

Table 8.9: A summary of model inputs (observed diagnostic test results and accuracy priors) for arboviruses

	Test type		
	PCR	**IgM ELISA**	**IgG ELISA**
Test details	Day 0 samples	Day 0 samples	Day 0 and day 28 samples
Testing strategy	All	All	Patients only
Number tested (%)[1]	9079 (88.6%)	9079 (88.6%)	4950 (63.0%)
Prior sensitivity	68%-99%[2]	0-50%	90-99%
Prior specificity	0-100%	0-100%	90-99%

[1] % of enrolled
[2] Sensitivity is assumed to be between 68% and 99% with 95% probability.

Table 8.10: Median (95% credible interval) estimate of the percentage of fevers attributed to arboviruses in all children (All), outpatient children (Out) and inpatient children (In)

		Laos	**Malawi**	**Mozambique**	**Zimbabwe**
dengue	All	10 (7-13)	0 (0-0)	0 (0-0)	0 (0-1)
	Out	10 (6-15)	0 (0-1)	0 (0-1)	0 (0-1)
	In	9 (5-14)	0 (0-2)	0 (0-1)	0 (0-2)
zika	All	0 (0-1)	0 (0-0)	0 (0-0)	0 (0-1)
	Out	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)
	In	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)
chikungunya	All	0 (0-1)	0 (0-1)	0 (0-0)	0 (0-1)
	Out	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)
	In	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)
JEV	All	0 (0-1)	-	-	-
	Out	0 (0-2)	-	-	-
	In	0 (0-2)	-	-	-

Note. JEV testing only carried out in Laos

Table 8.11: Median (95% credible interval) estimate of the percentage of fevers attributed to arboviruses in all adults (All), outpatient adults (Out) and inpatient adults (In)

		Laos	**Malawi**	**Mozambique**	**Zimbabwe**
dengue	All	19 (15-23)	0 (0-1)	0 (0-1)	0 (0-1)
	Out	23 (18-28)	0 (0-1)	0 (0-1)	0 (0-1)
	In	15 (11-20)	0 (0-2)	0 (0-2)	0 (0-1)
zika	All	0 (0-0)	0 (0-1)	0 (0-0)	0 (0-0)
	Out	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
	In	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)
chikungunya	All	0 (0-0)	0 (0-1)	0 (0-1)	0 (0-0)
	Out	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
	In	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)
JEV	All	0 (0-0)	-	-	-
	Out	0 (0-1)	-	-	-
	In	0 (0-1)	-	-	-

Note. JEV testing only carried out in Laos

Figure 8.5: Estimated percentage of fever cases attributed to each arbovirus by site and age in All patients depending on whether or not an informative or diffuse prior distribution was used for IgM and IgG ELISA.

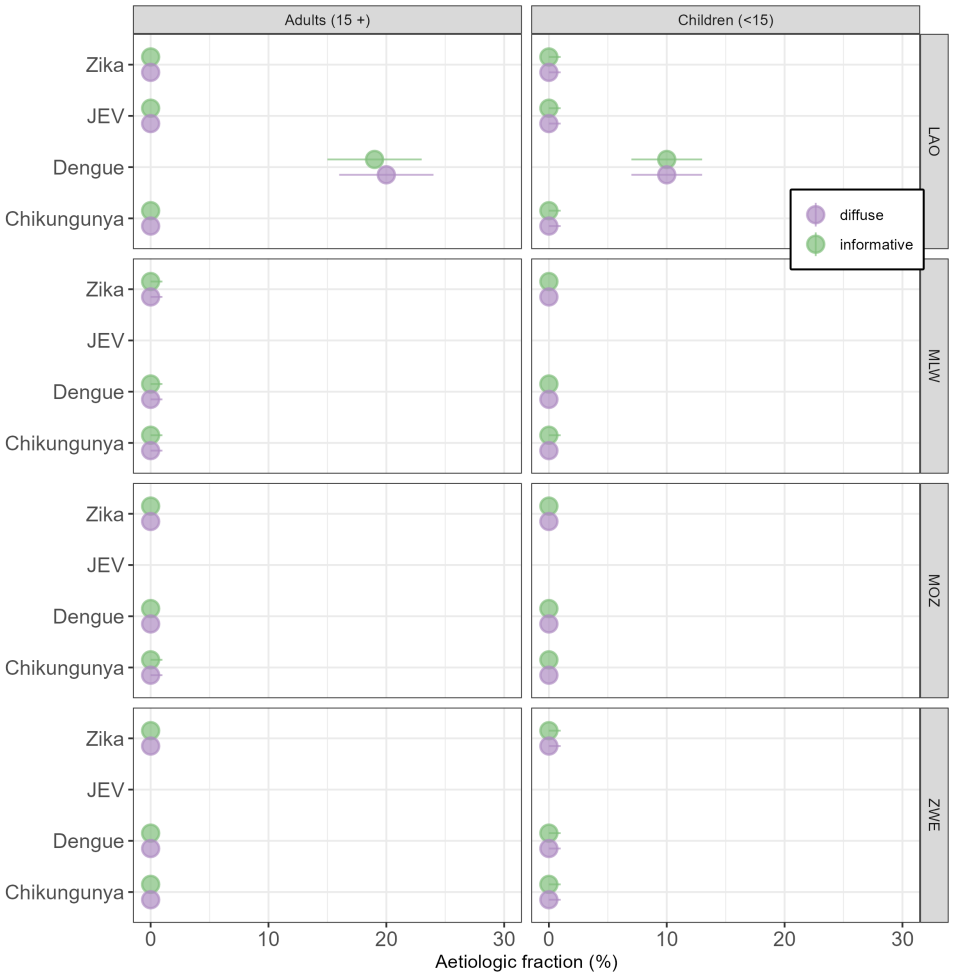


Figure 8.6: Estimated percentage of fever cases attributed to each arbovirus by site and age in Outpatients depending on whether or not an informative or diffuse prior distribution was used for IgM and IgG ELISA.

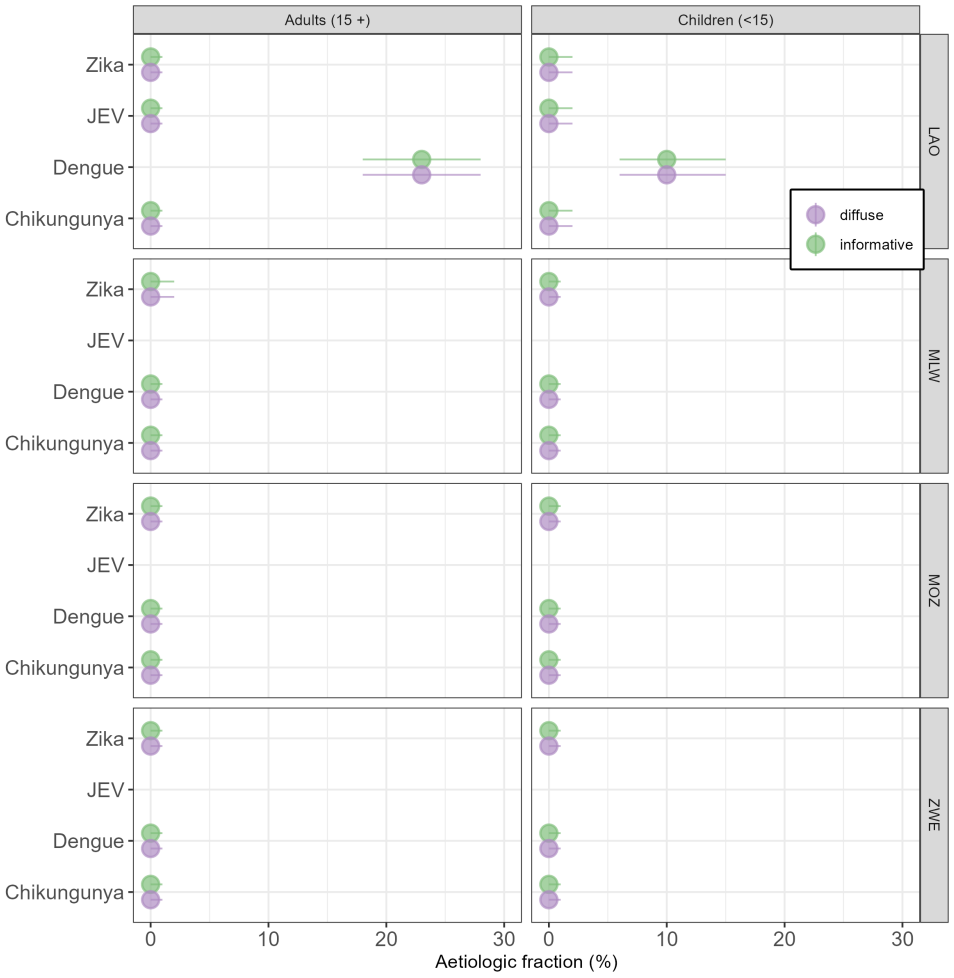
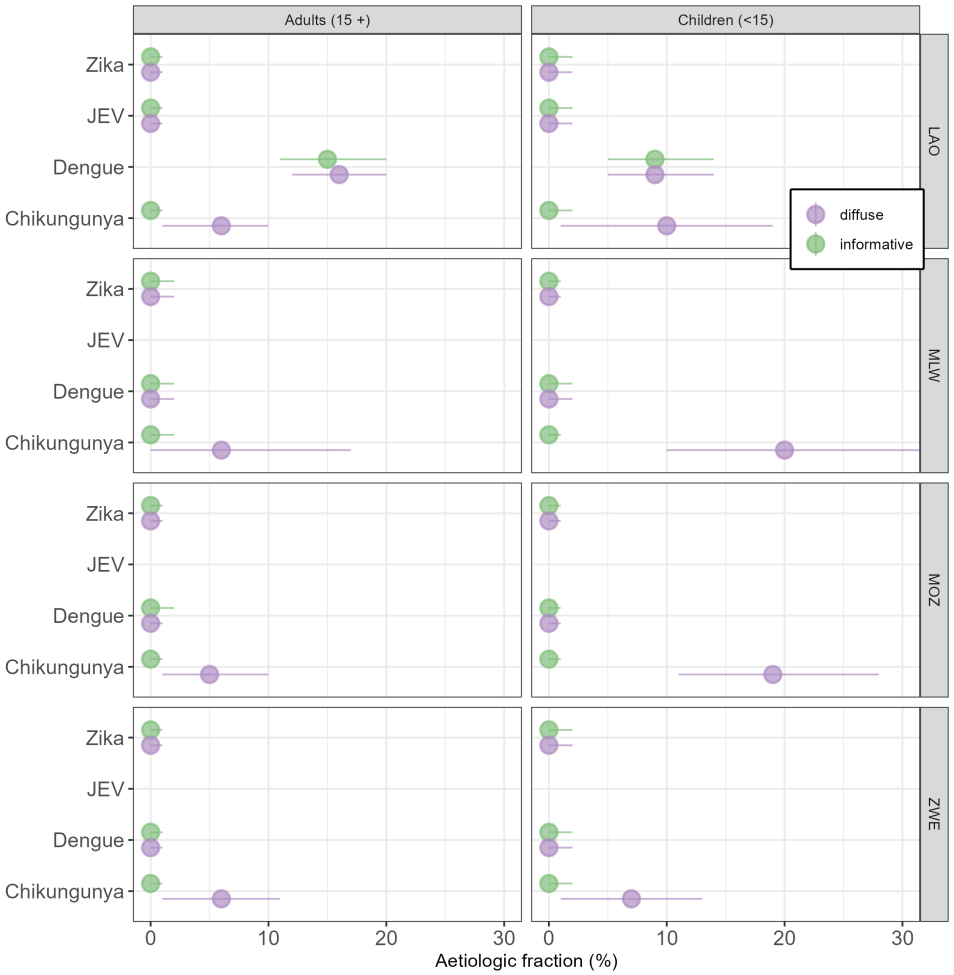


Figure 8.7: Estimated percentage of fever cases attributed to each arbovirus by site and age in Inpatients depending on whether or not an informative or diffuse prior distribution was used for IgM and IgG ELISA.



8.4.5 Leptospirosis

Leptospirosis infection was investigated by four diagnostic methods (Table 8.12). ELISA and PCR were performed on all participant samples. Those participants ELISA positive then had a Microscopic agglutination test (MAT) test. MAT should be performed on a day 0 and day 28 sample with the result inferred by comparing the results on these two days however, only 11% of those with a MAT result on day 0 also have a MAT result on day 28 as a result of loss to follow up. Similar to arboviruses detected by a single sample IgM ELISA, MAT single sample results are not expected to have a high sensitivity.

The single-pathogen model for leptospirosis estimates that between 0 and 1% of fever cases are attributed to leptospirosis across all FIEBRE sites, age groups and participant status. Sensitivity analyses that considered a more diffuse sensitivity prior for MAT paired and PCR had little impact on estimates but did on the whole have greater uncertainty than

those estimates with more informative priors (Figure 8.8).

Table 8.12: A summary of model inputs (observed diagnostic test results and accuracy priors) for leptospirosis

	Test type			
	ELISA	**MAT[1] acute**	**MAT paired**	**PCR**
Test details	Day 0 samples	Day 0 titre > 100 in African sites and ≥800 in Laos	Day 0 and day 28 samples	Day 0 samples
Testing strategy	All	ELISA positive	ELISA positive patients	All
Number tested (%)[2]	8888 (86.7%)	2643 (25.8%)	280 (2.7%)	8639 (84.3%)
Prior sensitivity	55-99%[3]	0-50%	80-99%	80-99%
Prior specificity	0-100%	0-100%	90-99%	0-100%

[1] Microscopic agglutination test

[2] % of enrolled

[3] Sensitivity is assumed to be between 55% and 99% with 95% probability.

Table 8.13: Median (95% credible interval) estimate of the percentage of fevers attributed to leptospirosis

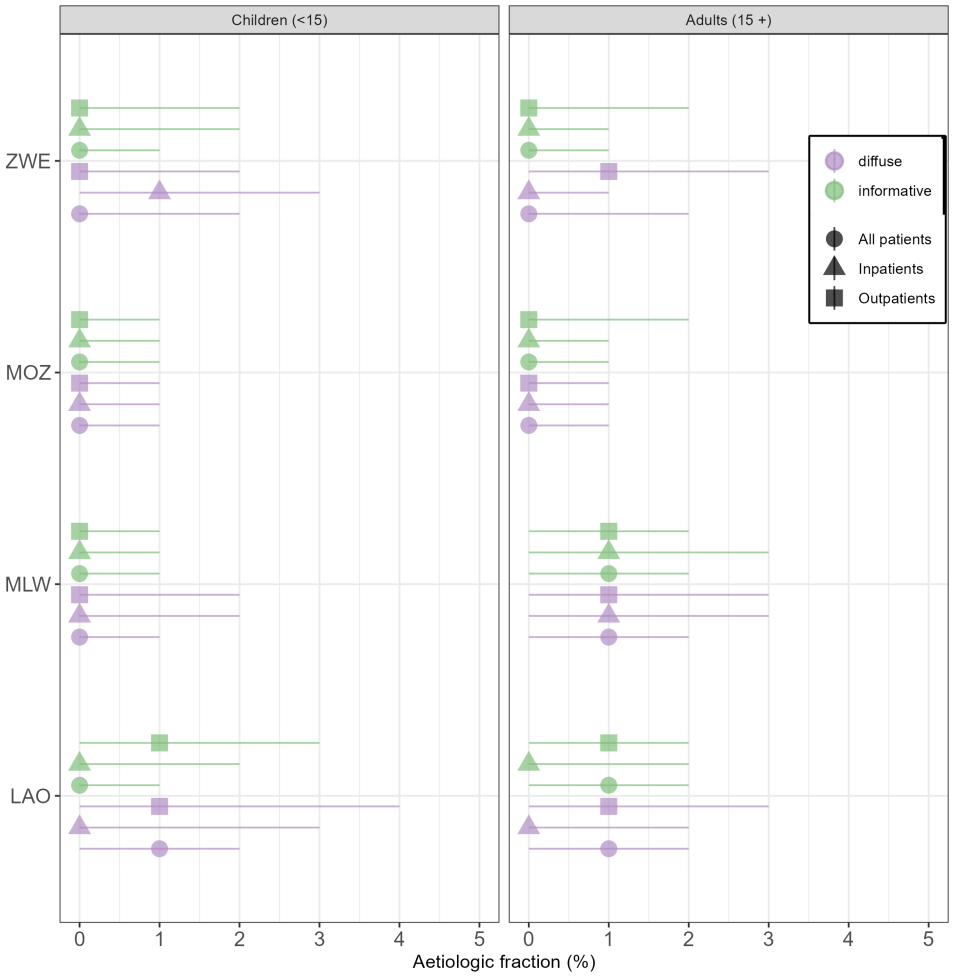
(a) Estimates in children aged less than 15 years

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
All patients	0 (0-1)	0 (0-1)	0 (0-0)	0 (0-1)
Outpatients	1 (0-3)	0 (0-1)	0 (0-1)	0 (0-2)
Inpatients	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)

(b) Estimates in adults aged 15 years or more

	Laos	**Malawi**	**Mozambique**	**Zimbabwe**
All patients	1 (0-2)	1 (0-2)	0 (0-1)	0 (0-1)
Outpatients	1 (0-2)	1 (0-2)	0 (0-1)	0 (0-2)
Inpatients	0 (0-2)	1 (0-3)	0 (0-1)	0 (0-1)

Figure 8.8: A comparison of the estimated percentage of fever cases attributed to leptospirosis by site, age, participant status group and, whether or not an informative or diffuse prior distribution was used for paired MAT and PCR



8.4.6 Blood cultures

On blood cultures, 17 different pathogen groups were found (Table 7.14. For *Talaromyces marneffei*, Streptococcus other than *Streptococcus pneumoniae*, *Acinetobacter baumanii*, *Cryptococcus* species, *Haemophilus influenza*, *Enterococccus faecalis* and other mycobacteria, less than five cases were positive and these are not considered in further analyses. Blood cultures were performed on almost all case's day 0 samples (97.6% Table 7.6). As no controls have test results, a specificity prior is required. For all blood cultures the FIEBRE expert group decided upon a prior that placed 95% probability between 95% and 99.9% for blood culture specificity. Table 8.14 shows the sensitivity prior the expert group decided on for the detection of each pathogen by blood culture. Blood cultures are traditionally thought to have a high specificity but lower sensitivity. This means that we can expect that there are individuals

Table 8.14: A summary of the assumed sensitivity priors for blood cultures. All ranges supplied are assumed with 95% probability

Pathogen	**Sensitivity prior**
Burkholderia pseudomallei	30-50%
Typhoidal *Salmonella*	30-50%
Non-typhoidal *Salmonella*	40-65%
Staphylococcus aureus	40-75%
Escherichia coli	40-75%
Other Enterobacterales	40-75%
Klebsiella pneumoniae	40-75%
Streptococcus pneumoniae	30-50%

who received a negative result for specific pathogens by blood culture but their true infection status is positive (i.e. they had a false negative result).

In children, the pathogen responsible for the largest proportion of fever cases is typhoidal *Salmonella*. Only a small percentage of fever cases are attributed to typhoidal *Salmonella* in Laos, Malawi and Mozambique but in Outpatient children in Zimbabwe the estimated percentage of fever cases attributed to typhoidal *Salmonella* is 31% (95% CrI: 19-47% - Table 8.15). All other pathogens detected via blood culture are estimated to be responsible for less than 5% of fever cases in children across all sites. In adults, typhoidal *Salmonella* remains the pathogen detected by blood culture responsible for the largest proportion of fever cases and in Outpatient adults in Zimbabwe, the percentage is higher than in children at 73% (95% CrI: 53-96% - Table 8.16). Overall, pathogens detected by blood culture account for a greater proportion of fever cases in adults than children but individually, most pathogens detected by blood culture only accounted for a small percentage of fevers.

Sensitivity analyses for blood culture pathogen single-pathogen models considered prior distributions for blood culture sensitivity of plus and minus 5% on the primary analysis 95% credible interval. For example, blood culture was assumed to be 30-50% sensitivity with 95% probability for *Burkholderia pseudomallei* in the primary analysis model. The sensitivity analysis considered a diagnostic test sensitivity of 25% to 55% with 95% probability. Figure 8.9 shows a comparison of the posterior estimates of the percentage of fever cases in Outpatients attributed to each pathogen detected by blood culture. Overall, while point estimates are not too dissimilar, much wider credible intervals highlight the dependence on the priors in these single-pathogen models that rely solely on a single blood culture.

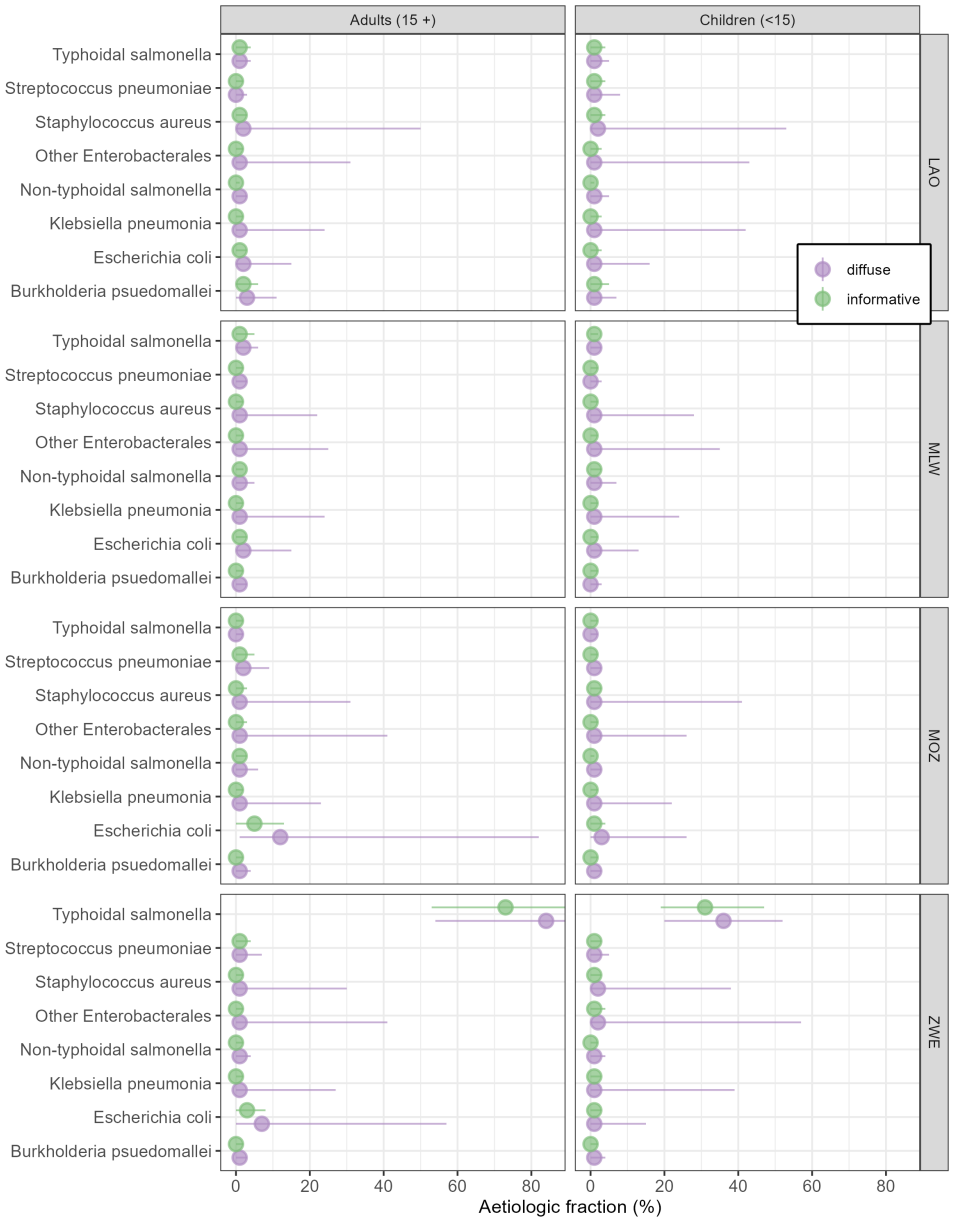
Table 8.15: Median (95% credible interval) estimate of the percentage of fevers attributed to pathogens from blood cultures in all children (All), outpatient children (Out) and inpatient children (In)

		Laos	**Malawi**	**Mozam-bique**	**Zimbabwe**
Burkholderia pseudomallei	All	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)
	Out	1 (0-5)	0 (0-2)	0 (0-2)	0 (0-2)
	In	1 (0-3)	1 (0-3)	0 (0-2)	1 (0-3)
Typhoidal *Salmonella*	All	1 (0-3)	1 (0-2)	0 (0-1)	21 (9-33)
	Out	1 (0-4)	1 (0-2)	0 (0-2)	31 (19-47)
	In	2 (0-6)	1 (0-5)	1 (0-3)	6 (0-15)
Non-typhoidal *Salmonella*	All	0 (0-1)	1 (0-3)	0 (0-1)	0 (0-1)
	Out	0 (0-1)	1 (0-3)	0 (0-1)	0 (0-2)
	In	0 (0-2)	2 (0-6)	0 (0-2)	0 (0-3)
Staphylococcus aureus	All	1 (0-3)	0 (0-2)	2 (0-6)	1 (0-2)
	Out	1 (0-4)	0 (0-2)	1 (0-3)	1 (0-3)
	In	1 (0-4)	1 (0-5)	4 (0-11)	1 (0-5)
Escherichia coli	All	0 (0-1)	0 (0-1)	1 (0-3)	0 (0-2)
	Out	0 (0-3)	0 (0-2)	1 (0-4)	1 (0-3)
	In	0 (0-30	0 (0-3)	1 (0-4)	1 (0-3)
Other Enterobacterales	All	0 (0-1)	0 (0-1)	0 (0-2)	1 (0-3)
	Out	0 (0-3)	0 (0-2)	0 (0-2)	1 (0-4)
	In	0 (0-2)	0 (0-3)	1 (0-4)	1 (0-4)
Klebsiella pneumoniae	All	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
	Out	0 (0-3)	0 (0-2)	0 (0-2)	1 (0-3)
	In	0 (0-3)	1 (0-3)	1 (0-3)	1 (0-4)
Streptococcus pneumoniae	All	1 (0-2)	0 (0-2)	2 (0-4)	0 (0-2)
	Out	1 (0-4)	0 (0-2)	0 (0-2)	1 (0-3
	In	1 (0-4)	1 (0-4)	4 (0-10	1 (0-3)

Table 8.16: Median (95% credible interval) estimate of the percentage of fevers attributed to pathogens from blood cultures in all adults (All), outpatient adults (Out) and inpatient adults (In)

		Laos	**Malawi**	**Mozambique**	**Zimbabwe**
Burkholderia pseudomallei	All	2 (0-5)	0 (0-1)	0 (0-1)	0 (0-1)
	Out	2 (0-6)	0 (0-2)	0 (0-2)	0 (0-2)
	In	3 (2-7)	1 (0-5)	1 (0-3)	0 (0-3)
Typhoidal *Salmonella*	All	1 (0-3)	2 (0-6)	0 (0-2)	43 (29-63)
	Out	1 (0-4)	1 (0-5)	0 (0-2)	73 (53-96)
	In	1 (0-5)	5 (0-14)	1 (0-5)	3 (0-9)
Non-typhoidal *Salmonella*	All	0 (0-1)	0 (0-2)	1 (0-3)	0 (0-1)
	Out	0 (0-1)	1 (0-2)	1 (0-3)	0 (0-2)
	In	0 (0-1)	1 (0-4)	2 (0-5)	0 (0-2)
Staphylococcus aureus	All	0 (0-2)	0 (0-1)	1 (0-5)	0 (0-2)
	Out	1 (0-3)	0 (0-2)	0 (0-3)	0 (0-2)
	In	0 (0-2)	1 (0-5)	4 (0-12)	1 (0-3)
Escherichia coli	All	1 (0-3)	1 (0-3)	5 (0-12)	1 (0-6)
	Out	1 (0-3)	1 (0-3)	5 (0-13)	3 (0-8)
	In	1 (0-5)	3 (0-9)	6 (1-16)	1 (0-5)
Other Enterobacterales	All	1 (0-2)	0 (0-2)	1 (0-3)	0 (0-2)
	Out	0 (0-2)	0 (0-2)	0 (0-3)	0 (0-2)
	In	1 (0-4)	1 (0-6)	2 (0-7)	1 (0-4)
Klebsiella pneumoniae	All	0 (-1)	0 (0-2)	1 (0-2)	0 (0-1)
	Out	0 (0-2)	0 (0-2)	0 (0-2)	0 (0-2)
	In	1 (0-3)	1 (0-6)	2 (0-7)	1 (0-3)
Streptococcus pneumoniae	All	0 (0-1)	0 (0-1)	1 (0-3)	1 (0-2)
	Out	0 (0-2)	0 (0-2)	1 (0-5)	1 (0-4)
	In	0 (0-2)	1 (0-5)	1 (0-4)	1 (0-4)

Figure 8.9: A comparison of the estimated percentage of fever cases attributed to pathogens detected by blood culture in Outpatients using a more informative or diffuse prior for diagnostic test sensitivity by site and age group.



8.4.7 Discussion of single-pathogen models

Single-pathogen model for pathogen j estimates the percentage of fevers attributed to pathogen j assuming each case's fever is either attributed to pathogen j or is attributed to something else. Each pathogen is investigated in isolation and so cases could by this method have a probability of being attributed to a number of pathogens that sums to greater than 1 when the probabilities are combined. In comparison to the percentage of positive test results, this section explores how the crude reflections of pathogen's prevalence reflect true pathogen prevalence through incorpo-

rating the diagnostic accuracy of the diagnostic methods used. Where test sensitivity is less than perfect, it is expected that the percentage of positive test results shown in Chapter 7 will underestimate the percentage of fevers attributed to a particular pathogen.

From these single-pathogen models, malaria is responsible for the largest proportion of fever cases in Malawi and Mozambique. In Laos, the pathogen responsible for the largest proportion of fever cases differs by age. In children, influenza (A or B) is responsible for the most fever cases while in adults it is dengue. In Zimbabwe, typhoidal *Salmonella* is responsible for the largest proportion of fever cases. This alone shows heterogeneity in the predominant cause of fever cases by site with a parasitic primary cause in Malawi and Mozambique, viral in Laos and bacterial in Zimbabwe. In some sites, there are also differences between the primary cause of fever cases in Inpatients and Outpatients reflecting different exposures in fever cases from different locations. For example, while typhoidal *Salmonella* is the primary cause of fever cases (from those pathogens investigated in FIEBRE) in Zimbabwe, this is primarily seen in Outpatients. On the other hand, in Mozambique, malaria is the primary cause of fever in Inpatients.

Sensitivity analyses for single-pathogen models showed little deviation in estimates from the primary analysis model for all pathogens except chikungunya and pathogens detected by blood culture. For chikungunya, the combination of a high proportion of positive test results and relatively diffuse priors has allowed for label switching. That is, that there is an alternative solution of estimates for the aetiologic fraction, sensitivities and specificities that is consistent with the data but less likely given our prior knowledge. In this case, instead of the IgM on acute samples having a poor sensitivity, this has switched to the PCR and IgG having poor sensitivity and the IgM sensitivity is high in turn, leading to higher chikungunya specific case fractions. This is an example, of where incorporating prior knowledge is not only useful but realistic as it is not particularly likely that IgM on acute samples only is more sensitive than PCR or IgG on paired samples. These implausible estimates are most noticeably a problem in the analysis on Inpatients only (Figure 8.7).

In almost all single-pathogen models, sensitivity analyses that considered a more diffuse prior distribution for test sensitivity resulted in greater uncertainty in estimates of the parameter of interest but for pathogens detected by blood culture this was more extreme than other pathogens (Figure 8.9). This is an artefact of only few positive test results combined with only a single diagnostic test of uncertain sensitivity.

8.5 Multi-pathogen model

The multi-pathogen model combines all the single-pathogen models discussed above into a single model. This consists of the observed diagnostic test results for 37 different tests used to identify 25 different pathogens (I exclude JEV from the multi-pathogen model) for 4429 outpatient fever cases (99.9% of enrolled outpatients - Table 7.5) and 3413 inpatient fever cases (99.8% of enrolled inpatients - Table 7.5) where each fever case contributes at least one diagnostic test result. The prior distributions used for each diagnostic tests accuracy are the same as those discussed in the sections above for the single-pathogen primary analysis models. The only exception to this is for malaria, where the model assuming conditional independence between malaria test results is used in the multi-pathogen model due to challenges in implementing the multiple-pathogen model with conditional dependence in disease positive individuals between malaria tests.

In the multi-pathogen model, I estimate the conditional probability that each of the J pathogens is the primary cause of an individuals fever given **all** observed test results for an individual. An additional latent subgroup represents those fever cases attributed to a cause not specified (denoted NOS) i.e. a pathogen not investigated for in the current study. With the addition of the NOS subgroup, there are 26 $(J + 1)$ latent subgroups.

Consider the case when a fever patient has a positive diagnostic test results for two pathogens and the diagnostic test has a perfect specificity. Neither the single-pathogen or the multi-pathogen model can distinguish which pathogen is the most likely cause [185] however, how this impacts the resulting estimates of cause-specific case fractions differs between the models. In the multi-pathogen model the aetiology probability is distributed equally to both pathogens [185]. In comparison, in the single-pathogen model aetiology would be solely attributed to the pathogen under consideration. The result of this, is an underestimation of the percentage of fever cases attributed to the true cause in the multi-pathogen model and may explain differences between the single-pathogen and multi-pathogen model results.

I have presented analyses for All patients, Inpatients only and Outpatients only as in the single-pathogen model results. A secondary analysis considered only those patients confirmed to be HIV positive in one site. HIV infection is known to impact the aetiologic distribution of infections [12] and it was of interest to see if this was also the case in FIEBRE and if so what the differences may mean for clinical practice. In FIEBRE, a

HIV positive diagnosis was defined by a combination of self reported results and local HIV testing. No HIV testing was carried out in Laos [20] and there were small numbers of HIV positive in child strata in Malawi and Zimbabwe. As a result, this secondary analysis was carried out only for HIV positive individuals in Mozambique.

8.5.1 Results

This section presents the results of the multi-pathogen models. Results are presented graphically in the main text but a complete results table for each model can be found in Tables G.1, G.2 and G.3 in Appendix G. All results presented and discussed are from models that exhibited no issues using Stan's diagnostic feature [125] and from visualization of trace plots and rank plots have converged (See Appendix H for trace and rank plots from the final model for Inpatients for the subset of parameters related to children in Laos).

Outpatients

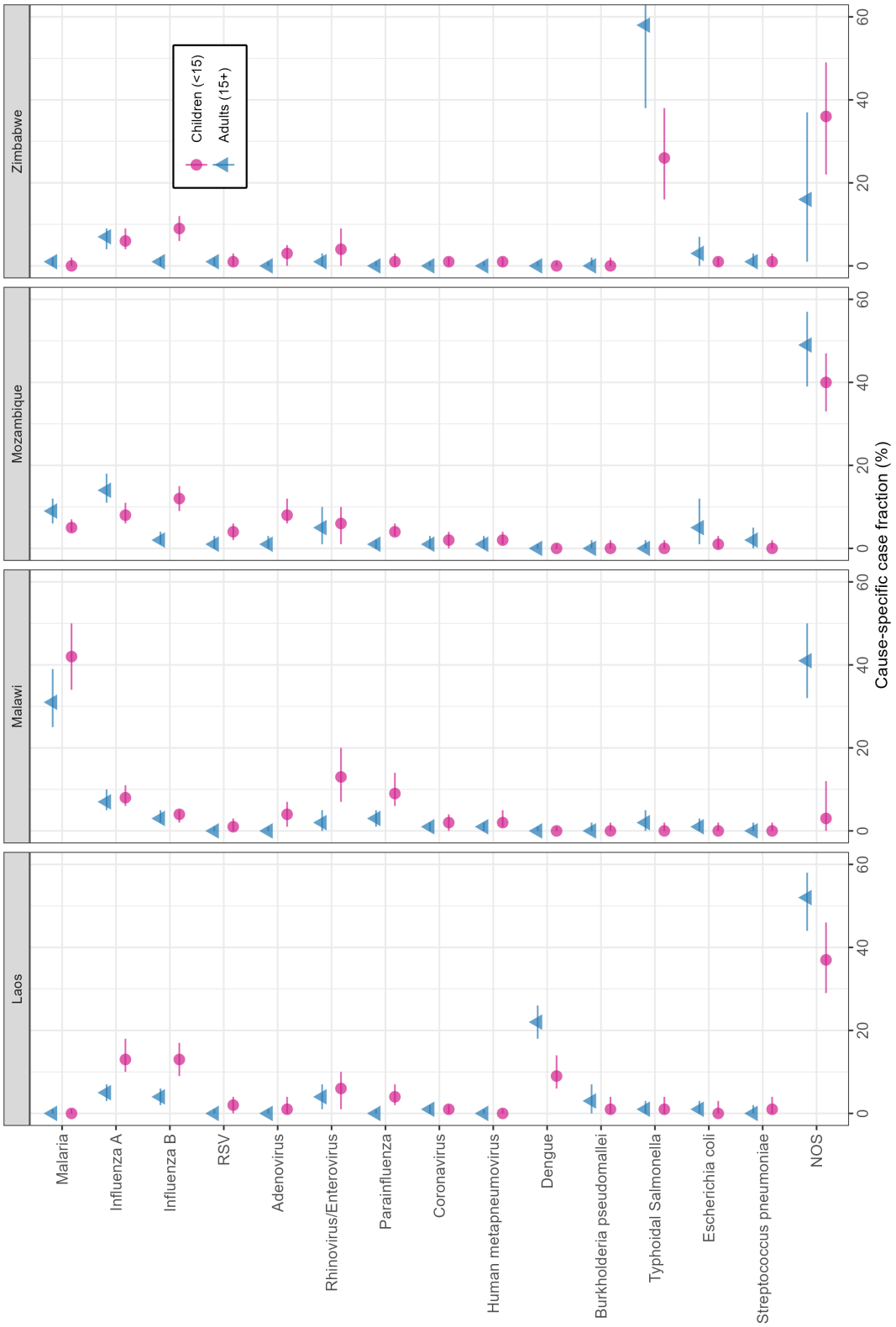
In Outpatients, the pathogen with the largest cause-specific case fraction differed by site (Figure 8.10). In Malawi, malaria has the largest cause-specific case fraction of any pathogen in both children (45%, 95% CrI: 36-52%) and adults (33%, 95% CrI: 26-41%). In Zimbabwe, Typhoidal *Salmonella* was the pathogen with the largest cause-specific case fraction and was responsible for 26% (95% CrI: 16-38%) of fever cases in children and 58% (95% CrI: 38-74%) of fever cases in adults. In Malawi and Zimbabwe the pathogen responsible for the greatest proportion of fever cases in Outpatients was the same in children and adults but this was not true for Laos and Mozambique. In adults from Laos, dengue had the largest cause-specific case fraction (22%, 95% CrI: 18-26%) while in children, it was Influenza A and Influenza B. In Mozambique, Influenza A had the largest estimated cause-specific case fraction at 14% (95% CrI: 11-18%) while in children it was Influenza B (12%, 95% CrI: 9-15%).

In all four sites, fever cases were attributed to a variety of respiratory pathogens in Outpatients. The burden of respiratory pathogens in fever cases was greater in children than adults. Arboviruses were not responsible for any fever cases in the three African sites and very few bacterial pathogens were responsible for Outpatient fever cases across any site with the exception of typhoidal *Salmonella* in Zimbabwe.

Overall, outpatient fever cases in the four FIEBRE study sites were not all attributed to one of the investigated pathogens (i.e. the not specified (NOS) cause-specific case fraction is not equal to 0%). However,

in Malawi children, a large proportion of Outpatient fever cases were attributed to one of the pathogens investigated and just 3% (95% CrI: 0-12%) of fever cases were attributed to the not specified subgroup. In Laos, Malawi and Mozambique, more fever cases were attributed to a pathogen in children than adults however, in Mozambique, similar percentages were attributed to pathogens in both age groups. Adult Outpatients in Laos were the only strata where more than 50% of fever cases were attributed to the not specified category (52%, 95% CrI: 45-58%).

Figure 8.10: Estimated cause-specific case fractions in Outpatient fever cases by site and age



Note. Not specified (NOS) includes those pathogens not tested for in FIEBRE. Pathogens contributing 1% of less to the aetiologic fraction in all stratum are excluded from the figure. These are: human bocavirus, *Chlamydophila pneumonia*, *Mycoplasma pneumonia*, brucella, zika, chikungunya, JEV, leptospirosis, non-typhoidal *Salmonella*, *Staphylococcus aureus*, *Klebsiella pneumonia* and other enterobacterales.

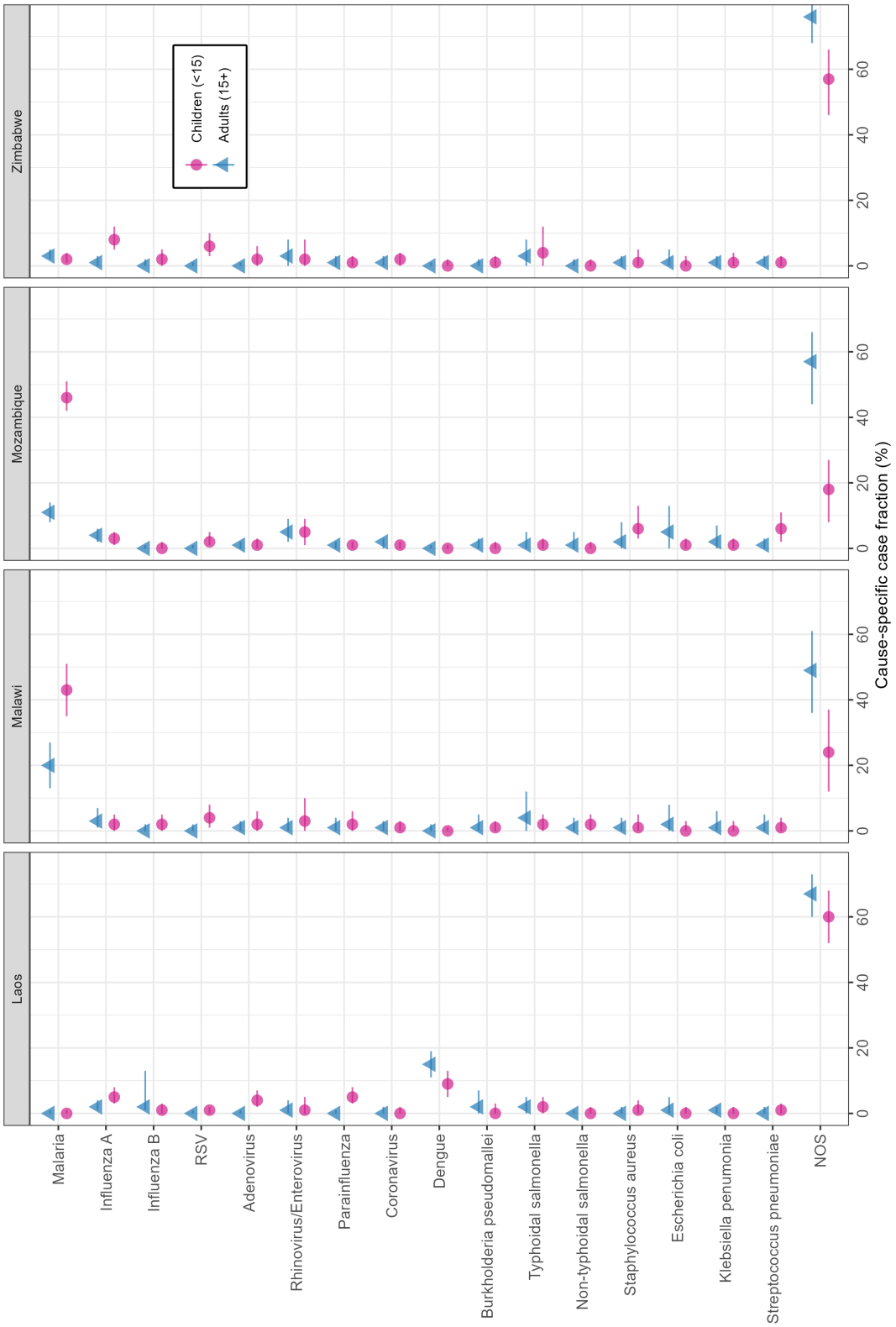
Inpatients

In Inpatients, the pathogen with the largest cause-specific case fraction in Laos, was dengue as in Outpatients however, unlike in Outpatients it was the largest cause-specific case fraction in both adults (15%, 95% CrI 11-19%) and children (9%, 95% CrI 5-13%) (See Figure 8.11). This estimate for Inpatient children is similar to that for Outpatient children but in adults, this estimate suggests that dengue fever is responsible for fewer Inpatient fever cases than Outpatient fever cases (15% vs. 22%). In Malawi, malaria has the largest cause-specific case fraction of any pathogen considered in both children and adults.

In Mozambique, the pathogen with the largest cause-specific case fraction was malaria, 46% (95% CrI: 42-51%) in children and 11% (95% CrI: 8-14%) in adults. In Zimbabwe, there was no pathogen considered in FIEBRE that was responsible for more than 10% of febrile Inpatients. Overall, fevers were attributed to a greater variety of bacterial pathogens in Inpatient cases than Outpatients.

In Inpatients, five out of eight strata had a cause-specific case fraction for the not specific group of over 50%, these were; Laos and Zimbabwe children and adults as well as Mozambique adults. In each of these strata, less than 50% of Inpatient fever cases were attributed to one of the pathogens investigated in FIEBRE.

Figure 8.11: Estimated cause-specific case fractions in Inpatient fever cases by site and age



Note. Not specified (NOS) includes those pathogens not tested for in FIEBRE. Pathogens contributing 1% of less to the aetiologic fraction in all stratum are excluded from the figure. These are: human bocavirus, human metapneuomvirus, *Chlamydophila pneumonia*, *Mycoplasma pneumonia*, brucella, zika, chikungunya, JEV, leptospirosis and other enterobacterales.

All patients

The results for All patients reflect a combination of the Outpatient only and Inpatient only results and are shown in Figure 8.12.

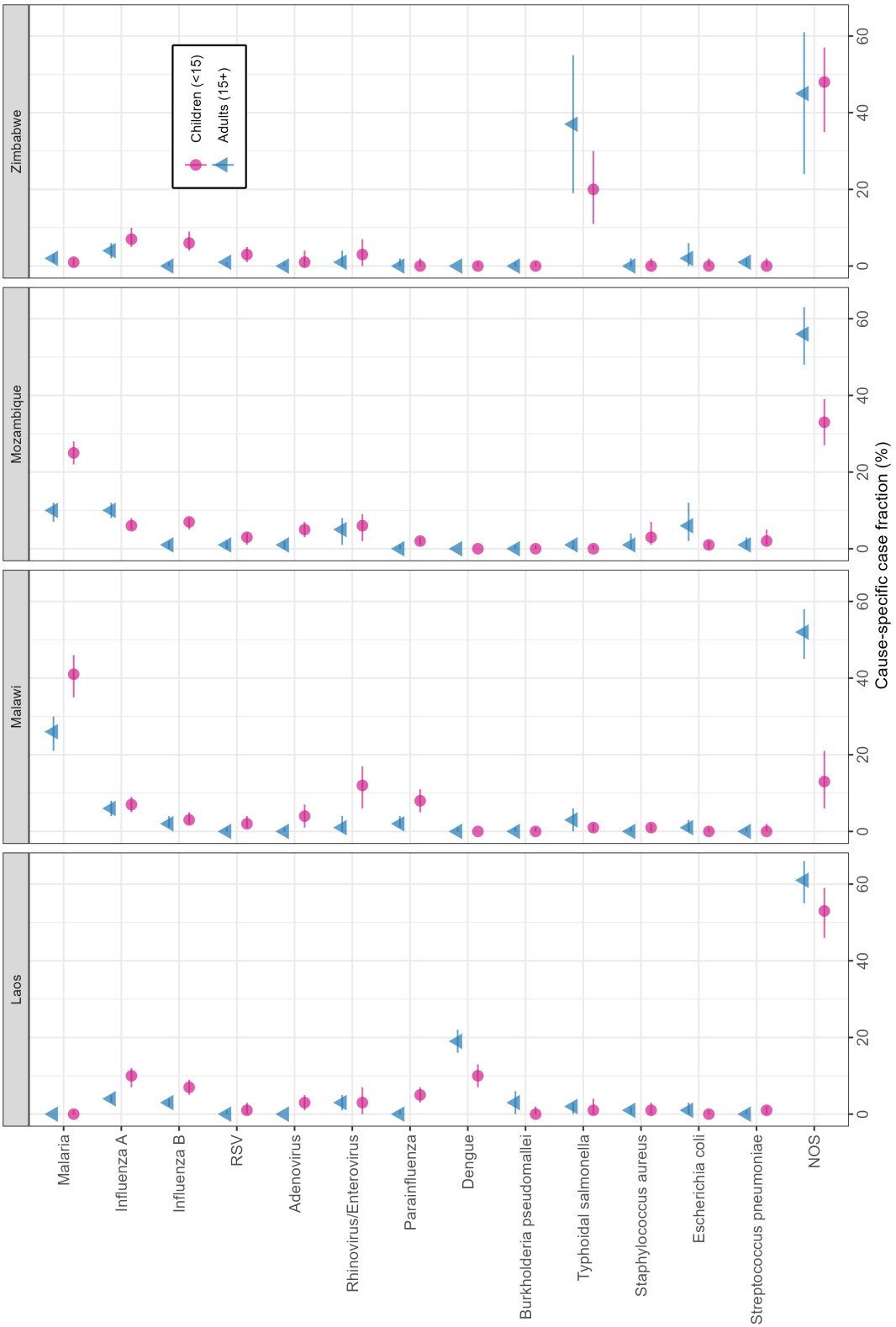
In Laos, dengue fever had a cause-specific case fraction of 19% (95% CrI: 16-22%) in adults and 10% (95% CrI: 7-13%) in children. In Laos children, Influenza A also had a cause-specific case fraction of 10% (95% CrI: 7-12%). Overall, over half of all fever cases (children and adults) were attributed to the not specified category and not attributed to one of the pathogens considered in this analysis.

In Malawi, malaria had the largest cause-specific case fraction 26% (95% CrI: 21-30%) in adults and 41% (95% CrI: 35-46%) in children. In children, rhinovirus/enterovirus also had a cause-specific case fraction over 10% (12%, 95% CrI: 6-17%) and only 13% (95% CrI: 6-21%) of fever cases were attributed to the not specified category. However, in adults 52% (95% CrI: 45-58) of fever cases were attributed to the not specified category.

In Mozambique children, malaria was the pathogen with the largest cause-specific case fraction 25% (95% CrI: 22-28%) and only a third of fever cases (33%, 95% CrI: 27-39%) were attributed to the not specified category. In Adults though, while both malaria and Influenza A had a cause-specific case fraction of 10% overall 56% (95% CrI: 48-63%) of fever cases were attributed to the not specified category.

In Zimbabwe, the only pathogen considered that had an estimated cause-specific case fraction over 10% was Typhoidal *Salmonella*. The cause-specific case fraction for Typhoidal *Salmonella* was estimated at 20% (95% CrI: 11-30%) in children and 37% (95% CrI: 19-55%) in adults. Overall, just under half of fever cases in both children (48%, 95% CrI: 35-57%) and adults (45%, 95% CrI: 24-61%) in Zimbabwe were attributed to the not specified category.

Figure 8.12: Estimated cause-specific case fractions in All patient fever cases by site and age



Note. Not specified (NOS) includes those pathogens not tested for in FIEBRE. Pathogens contributing 1% of less to the aetiologic fraction in all stratum are excluded from the figure. These are: human bocavirus, coronavirus, human metapneumovirus, *Chlamydophila pneumonia*, *Mycoplasma pneumonia*, brucella, zika, chikungunya, JEV, leptospirosis, non-typhoidal *Salmonella*, *Klebsiella pneumonia* and other enterobacterales.

8.5.2 Secondary analyses

Primary analyses considered all fever cases irrespective of HIV status. This section presents the results of a secondary analysis considering only HIV positive cases in Mozambique. Table 8.17 shows the number of cases included in each stratum for this analysis. Just under half (49%) of all adults enrolled in Mozambique in the FIEBRE study were HIV positive with a slightly higher percentage in inpatients (61%) compared to outpatients (42%). Only a small number of HIV positive children are included in this analysis and in both inpatient and outpatients analyses, there are more females included. This is representative of the epidemiology of HIV in Mozambique where the prevalence has been shown to be higher in females than males [186].

Table 8.17: Summary of the number of HIV positive cases in Mozambique, shown overall and by in/out-patient status, age group and sex. Results are number (% of enrolled cases)

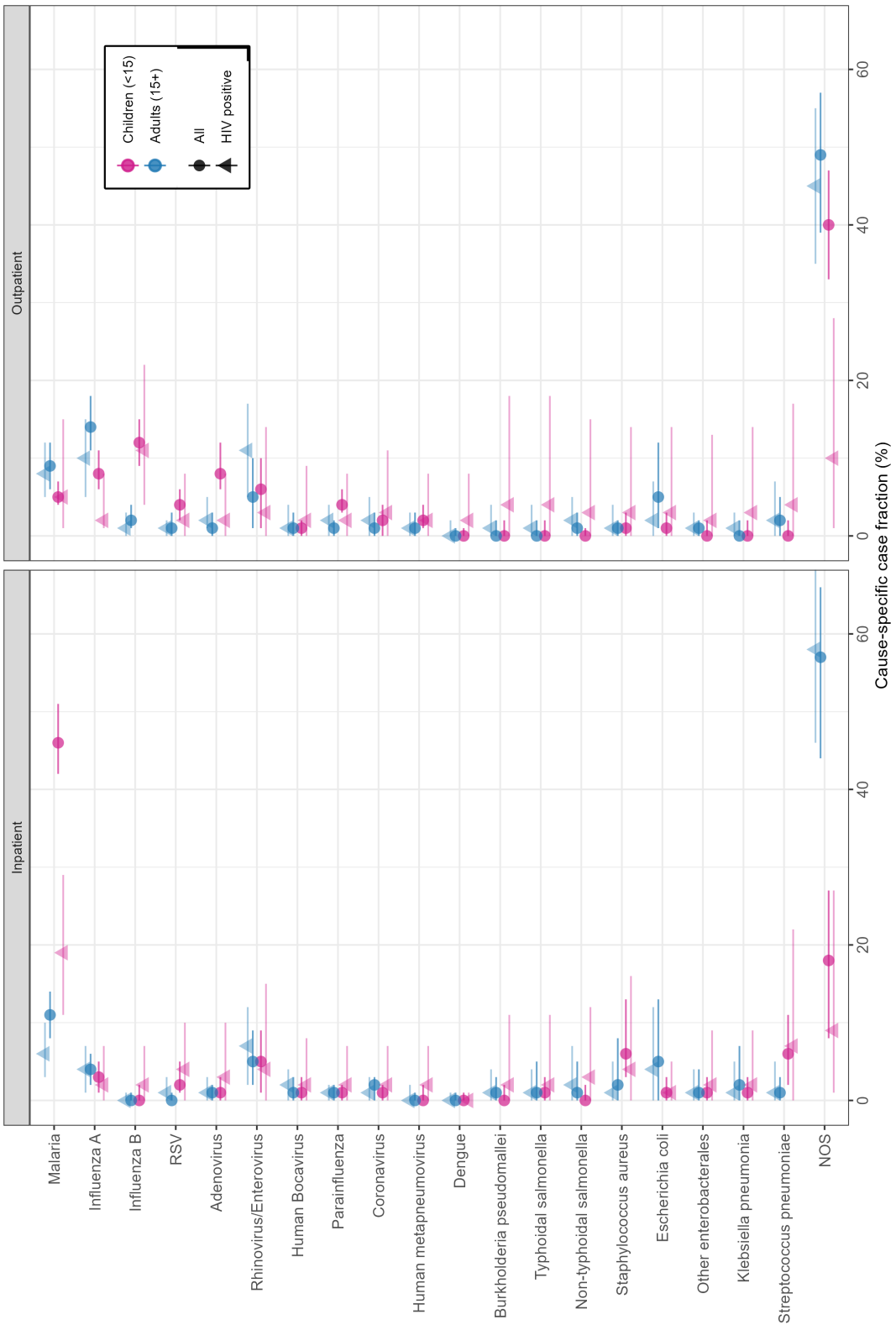
	Mozambique		
	Inpatient	**Outpatient**	**Total**
Total HIV positive	321 (31%)	248 (21%)	569 (26%)
Age			
Age < 15 years	53 (10%)	20 (3%)	73 (6%)
Age ≥15 years	268 (61%)	228 (42%)	496 (49%)
Sex			
Female	172 (35%)	190 (27%)	362 (30%)
Male	149 (28%)	58 (13%)	207 (21%)

The estimated cause-specific case fractions are similar between a model including all cases irrespective of HIV status and a model including HIV positive cases only (Figure 8.13). A notable difference in inpatients is the estimated cause-specific case fraction for Malaria. In all inpatients 46% (95% CrI: 42-51%) of fever cases were attributed to malaria in children however, when only the HIV positive inpatients are considered, this estimate fell to 19% (95% CrI: 11-29%) (See Table G.4 for a complete results table). In adults the median percentage of fevers attributed to malaria also fell but the credible intervals remained overlapping.

In the outpatients, the estimated cause-specific case fraction was higher for rhinovirus/enterovirus in HIV positive adult cases compared to all adult cases and was lower for Influenza A and adenovirus in children. In total, the estimated cause-specific case fraction not attributed to one of the pathogens in FIEBRE was lower in the HIV positive cohort of children compared to all outpatient children. Overall, analyses for inpatient

and outpatient HIV positive cases only had wider uncertainty intervals reflecting the smaller number of individuals included in each analysis.

Figure 8.13: A comparison of the estimated cause-specific case fractions in outpatients and inpatients from Mozambique by HIV status and age group



Not specified (NOS) includes those pathogens not tested for in FIEBRE. Pathogens contributing 1% of less to the aetiologic fraction in all stratum are excluded from the figure. These are: *Chlamydophila pneumonia*, *Mycoplasma pneumonia*, brucella, zika, chikungunya, leptospirosis, non-typhoidal *Salmonella*, *Staphylococcus aureus*, *Klebsiella pneumonia* and other Enterobacterales.

8.6 Discussion

This chapter has presented the statistical methods used to estimate the aetiologies of fever in the FIEBRE study from multivariate binary imperfect diagnostic test data. The approach used was adapted from those methods developed to analyse a case-control study investigating a different syndrome: pneumonia [37, 168, 185]. This section discusses the findings of the multi-pathogen model in comparison to other studies of fever aetiology before reviewing the strengths and limitations of the statistical approach taken in light of these findings.

8.6.1 Fever aetiology results in context

Laos

In Laos, the pathogen attributed to the largest proportion of fever cases (inpatient and outpatient) in adults was dengue (19%, 95% CrI: 16-22%) and in children was both dengue (10%, 95% CrI: 7-13%) and Influenza A (10%, 95% CrI: 7-12%).

Laos, is one of two FIEBRE study sites where previous fever aetiology studies have been carried out and to which estimates can be compared. A single-study analysis of fever cases in Laos also found dengue to be responsible for the largest single proportion of fever cases (8%) from any pathogen considered [187]. However, in contrast to the findings in this thesis, they also found that leptospirosis and scrub typhus are important and treatable causes of fever in rural Laos [187]. There are multiple reasons why different results may have been found. Firstly, Mayxay *et al.* [187] recruited individuals from northern and southern regions not included in FIEBRE recruitment. It is possible, that these regions experience different causes of fever to the region from which fever cases in FIEBRE arise. Secondly, the cases from this study were recruited over 10 years prior to FIEBRE with possible changes in fever aetiology over time. Lastly, in 2019 a dengue epidemic was reported in Laos which overlapped with FIEBRE recruitment and may explain the larger proportion of fever cases attributed to dengue in this thesis than other studies [188].

Another study, this time carried out in the same region as FIEBRE, also recognised rickettsial infections (such as scrub typhus) as an under recognised cause of fever in Laos [189]. Rickettsial infections were not included in the analyses in this thesis as less than 75% of enrolled participants were tested. Despite this, Laos was the site with the greatest proportion of positive test results out of those individuals tested at 8.2% in inpatient adults and 4.8% in outpatient adults. It is likely, that a pro-

portion of fever cases would have been attributed to rickettsial infections if they had been included in analyses, which would have made results between these studies more congruent.

Overall, estimates from both previous studies [187, 190] assumed diagnostic tests used were perfect. In comparison, the methods used in this thesis did not make this same assumption which may also reflect differences between the study findings. The two previous studies, as well as a further study [189] all found dengue to be the pathogen responsible for the largest proportion of fever cases in Laos. However, these other studies did not explore whether this was true for fever cases of all ages. In this thesis, I show that dengue was responsible for the largest proportion of fever cases in adults but in children respiratory pathogens are perhaps more important. This is contrary to findings from a study of dengue in Thailand which found dengue was predominant in younger age groups [191]. Unfortunately, similar to these other studies, approximately 50% of fever cases in Laos remain without a cause. This suggests there are other important causes of fever in Laos that FIEBRE did not test for or that diagnostic tests are even worse than we think.

Malawi

Of all strata considered, Malawi children (inpatient and outpatient) had the lowest proportion of fever cases with no cause assigned (13%, 95% CrI: 6-21%). However this was not the case for adults in Malawi, with 52% (95% CrI: 45-58%) of fever cases attributed to the category not specified. In Malawi, the primary cause of fever identified was malaria in both adults (26%, 95% CrI: 21-30%) and children (41%, 95% CrI: 35-46%) and, similar to one of the few studies investigating causes of fever in Malawi, [192] respiratory pathogens were the next most common cause. While respiratory pathogens were important causes of fever across all sites, in Malawi, rhinovirus/enterovirus and parainfluenza viruses were the most commonly attributed respiratory pathogens. Batlzell *et al.* [192] found the importance of respiratory pathogens in fever cases increased in the rainy season compared to the dry season which is an interesting element of fever aetiology not explored in this thesis.

A surprising result from the analyses presented here is the lack of blood stream infections as causes of fever in contrast to results from another study in Malawi [193]. However, this study was published over twenty years ago and may not represent current fever aetiology. A further study that investigated causes of fever in hospitalised patients only found that blood stream infections were a primary cause of fever [159].

In this thesis, blood stream infections in Malawi were responsible for more fever cases in inpatients compared to outpatients, in line with [159, 193] and suggesting different fever aetiologies in inpatients compared to outpatients.

Mozambique

In Mozambique, the pathogen responsible for the largest proportion of fever cases in all children was malaria (25%, 95% CrI: 22-28%). In adults, both malaria and Influenza A were estimated to be responsible for 10% of fever cases. Unique to Mozambique, *E.coli* was the bacterial pathogen responsible for the largest proportion of fever cases (6%, 95% CrI: 2-12%).

In the secondary analysis, I have shown that HIV positive fever cases from Mozambique have a similar aetiology distribution of respiratory and bacterial pathogens as the wider population. The results shown here suggest that malaria is responsible for fewer fever cases in the HIV positive cohort than in the whole cohort but it is not clear what pathogen/s have replaced malaria. In this analysis, the diagnostic test results from CrAg were excluded as only a very small subset of individuals were tested however, CrAg is a likely cause of more severe fever cases in HIV positive individuals [194].

At the time of this thesis, no previous fever aetiology studies had been published for Mozambique. This means that the results from this thesis cannot be compared to other work and also that the results from the thesis will be the first of their kind for Mozambique.

Zimbabwe

The primary cause of fever in outpatients in Zimbabwe identified from this analysis was Typhoidal *Salmonella*, estimated to be responsible for 20% (95% CrI: 11-30%) of child fever cases and 37% (95% CrI: 19-55%) of adult fever cases. However, an outbreak of Typhoid fever coinciding with FIEBRE recruitment means that these results can't necessarily be thought of as typical for the area [195].

Typhoidal *Salmonella* includes serovars *Typhi* and *Paratyphi*. The large cause-specific case fraction may be a consequence of Typhoidal *Salmonella*, notably *Salmonella typhi* showing reduced susceptibility or resistance since 2009 to the primary treatment ciprofloxacin [196]. The potential burden from Typhoidal *Salmonella* has been recognised in Zimbabwe and a mass typhoid vaccination campaign was undertaken between February and March 2019 overlapping the FIEBRE recruitment time frame. The vaccination campaign was targeted at children only and

may explain the difference between estimates of fever cases attributed to Typhoidal *Salmonella* in adults and children. Highlighting Typhoidal *Salmonella* as an important cause of fever in Zimbabwe is important because its diagnosis requires diagnostic and laboratory capacity not available in all lower resource settings. However, one of the first steps to increasing capacity to test for this pathogen are reliable estimates quantifying the burden. Interestingly, the fraction of fever cases attributed to Typhoidal *Salmonella* in inpatients was low compared to outpatients.

Overall

Despite a declining burden of malaria in many countries, malaria remains an important cause of health-care seeking due to a fever in Malawi, particularly in children. Since the start of recruitment for FIEBRE, the RTS-S malaria vaccine has begun its roll out and Malawi was one of the first three countries to begin providing the vaccine to children [197]. Assuming the malaria vaccine is effective the burden of fever cases attributed to malaria should decline as a result.

Across all sites, respiratory pathogens were important causes of fever in children. In particular, the burden of respiratory pathogens was greatest in outpatients (Figure 8.10) compared to inpatients (Figure 8.11). This is an important finding as global estimates of the burden of respiratory pathogens focus almost entirely on inpatients. These findings show that there is a potentially important and unaccounted burden of fever attributed to respiratory pathogens in the community settings as well. This was also highlighted in a recent aetiology study from Guinea-Bissau published in 2024[198]. If this is the case, highlighting this may help to reduce unnecessary antibiotic prescriptions.

Furthermore, respiratory pathogens have been shown to have a higher burden in children under 5 years of age and the elderly over 70 years of age [199]. The analysis in this chapter only considered age categorised into two: under 15 years and 15 years and over. Had I considered further age groups, I may well have found respiratory pathogens were responsible for more fever cases in the youngest and oldest age groups. RSV in particular, is one of the most common viruses to infect children worldwide and has increasingly been recognised as an important pathogen in adults [200]. However, in FIEBRE, RSV was not found to be responsible for more than 5% of fever cases in All patients (Figure 8.12). This may reflect changes in the epidemiology of RSV in response to the COVID-19 pandemic [201].

A greater burden of respiratory infections in outpatients compared

to inpatients may also reflect differences in severity of illness. Inpatients were hospitalised and so this could be seen as a marker of more severe infection. It follows, that respiratory infections may be a key cause of fever in less severe cases but there are other infections that are more important in the most severe fever cases. The results in this chapter indicate that a greater fraction of inpatient fever cases in Mozambique were attributed to bacterial pathogens compared to outpatients (See Tables G.1 and G.2). This trend of a different fever aetiology based on severity can also be seen from two studies in Tanzania. In the first study in outpatients, acute respiratory infections were attributed to the largest proportion of fever cases [11] while in the second study in inpatients blood stream infections and bacterial zoonoses accounted for over half of the fever cases [18].

In most FIEBRE sites, there was still a substantial proportion of fever cases not attributed to a pathogen, though more fever cases have been attributed to a cause in children than adults these results highlight gaps in our understanding of the causes of fever in these locations. This could be due to pathogens not tested for but it could also be due to pathogens not yet discovered. Either way, more research is needed to understand the causes of fever.

The data used in this analysis are not error free. Any errors in the data could lead to erroneous findings in our analyses however, errors are at least expected to be constant between cases and controls. Possible sources of error include; specimen mislabelling resulting in a specimen being associated with the wrong individual, data entry at the reference laboratories and, challenges related to the storage of samples shipped around the world which affect the result of a diagnostic test. Furthermore, particular pathogens exhibited high levels of missing data, where less than 75% of all participants had been tested and as a result were excluded from analyses.

The results presented here rely to some extent on the prior distributions assumed for the diagnostic test sensitivities. While attempts were made to ensure the estimates are robust via sensitivity analyses, there are important factors in diagnostic test accuracy not considered here. For example, the days post onset of fever that an individual was recruited and specimens for diagnostic testing taken. Depending on the diagnostic test target, the days post onset of fever is fundamental to the test's accuracy and is data that could be utilised to refine the accuracy estimates used in this analysis. The prior distributions assumed for test accuracy are based on the assumption of an acute fever, typically defined as less than

seven days. This assumption is valid in FIEBRE as the median number of days of fever upon presentation was four days (Interquartile range 2 to 4 days). However, 4% of cases reported that their fever had in fact lasted longer than seven days. This variable is dependent on the memory of the respondent and is subject to recall bias. Given the limitations of this type of variable, incorporating this information to improve the prior distributions for test accuracy is not straightforward.

8.6.2 Statistical analysis considerations

I considered both single- and multi- pathogen models. These models are similar in many ways with many of the same assumptions made for example, the assumed prior distributions of test sensitivity. In this section, I focus on the assumptions made in the multi-pathogen model but most of these also apply to the single-pathogen model as well and I will indicate when this is not the case.

The multi-pathogen model presented in this thesis relies on the assumption of a single cause of fever. As a consequence, currently, when two pathogens are both equally likely to be the cause of a fever, this probability is split between the pathogen subgroups (or latent states). When the true cause of fever is only one of these two causes the true cause's aetiologic fraction is underestimated and the other aetiologic fraction is overestimated. From preliminary investigations into co-infections, it was not obvious that co-infections play an important role in fever causes in the FIEBRE study and so no co-infections were considered. However, if a well understood co-infection was present, this could be incorporated into the current model framework with an additional latent subgroup representing each co-infection to be considered [168].

Another assumption made in these analyses is that we can borrow the specificity estimates gleaned from controls for the cases. That is, we assume a common test accuracy between cases and controls within each site/age strata. This assumption helps with model identifiability as the specificity parameters are then identifiable however, a test may exhibit a different specificity in cases and controls rendering this assumption invalid. For example, in FIEBRE controls, who are apparently 'healthier' than the cases as they did not seek health care, controls may be less likely to have any infection than a case as a result the test may appear highly specific. However, in a population of fever cases (our population of interest) with various infections, the test may not be as specific due to cross-reaction not visible when specificity is estimated from controls without fever. To mitigate this, FIEBRE did not exclude controls with

fever so long as they do not meet the criteria for a case [167]. If controls had been excluded based on fever status, biased results would be likely when a common specificity is assumed[202].

Also related to the selection of controls, in FIEBRE controls were matched to outpatient cases only. Inpatients may come from a different location with different exposures and different background levels of certain pathogens. As a result, assuming a common specificity between outpatients and inpatients in this case may not be valid and hence the results may be biased if the specificity of that test was different in the population from which inpatients arose. A consequence of this is estimation bias in our inpatient only and all patient results. Further, if the control group is not representative of the population from which the inpatient cases have arisen, then any results found are not applicable beyond the particular group studied [145]. As a result, in general, diagnostic test accuracy should be estimated among the population in which the test is intended to be used.

The multi-pathogen model assumes independence between test results for different pathogens conditional on the true pathogen subgroup. However, there are cases where this assumption may not hold. For example, test results for different pathogens could be positively correlated if they cross-react for their respective targets. This is a well documented phenomenon for some of the FIEBRE pathogens including leishmaniasis and malaria [203]. An example of a correlation among negative test results, given the true disease status, is when laboratory factors such as poor specimen collection results in negative results for all pathogens measured. To relax the conditional independence assumption, [173] propose the nested partially latent class model in which local dependence is induced by nesting latent subgroups within each true infection class. Not accounting for dependence conditional on the true infection class may lead to biased estimates with the size of bias dependent on the strength of correlation between results given an infection subgroup. However, previous work has shown that identification of the most prevalent cause is generally robust to this assumption [173]. This assumption of conditional independence is different to the conditional independence between tests for the same pathogen given their true infection status. This assumption is not relevant to the single-pathogen models as only a single pathogen is considered.

The assumption that all cause-specific case fractions must sum to 100% also has potential consequences. If any of the test sensitivity priors are incorrect, leading to a biased estimate of a cause's aetiologic fraction,

because of the assumption that all aetiologic fractions must sum to 100%, a biased estimate for one cause will indirectly impact estimates for all other causes. The choice of sensitivity priors is therefore paramount as well as sensitivity analyses for each of these. The sensitivity analyses presented in this chapter show that almost all of our results are robust to the choice of sensitivity prior and are instead driven by the data.

Other statistical analysis considerations that may impact the estimates found include the use of binary data when more granular data was available. In this thesis, only binary diagnostic test results were used. However, for more than one infection diagnostic test results were ordered categorical or continuous and transformed to a binary variable. Making use of this more granular data could improve estimates in a similar way to the addition of gold-standard data has been shown to give more precise estimates of aetiologic fractions [168].

Within FIEBRE, controls were matched to cases on age, sex and location of residence however, there may be other confounding factors that were not considered. HIV status is a potentially important factor to consider which may alter the aetiologic distribution of fever causes. HIV could have been included in the presented analysis through additional strata however some strata would have very small numbers and so this approach was not taken. Recent advances to the statistical approach taken here add additional explanatory variables such as HIV or seasonality via a regression modelling framework [151]. In this thesis, HIV status was considered in a small secondary analysis and estimated aetiologic fractions did not differ much from the primary analysis. Another explanatory variable is seasonality as many infections have a seasonal profile. Participants in FIEBRE were recruited over multiple years and so it is hoped that infections with seasonality have not been missed even if their importance may change throughout a year. Having said this, FIEBRE recruitment was only over a two year period and this coincided with the COVID-19 pandemic which may have changed fever aetiology as a result of a significant behaviour changes [201].

Two of the most common limitations of case-control studies arise from selection bias and recall bias [204]. Selection bias often occurs due to high rates of non-participation in controls that render the control population non-representative of the population at risk. While selection bias is a concern, particularly given that fewer controls were recruited than planned, all attempts were made to ensure the controls were representative of the population at risk, including the inclusion of controls with fever [167]. For the analysis in this thesis, recall bias is not a concern as the only

data used were the results of diagnostic tests which were not subject to the memory of a participant.

Chapter 9

Comparison with a frequentist approach to estimating fever aetiology

9.1 Preamble

In this chapter, I investigate a different statistical approach for estimating fever aetiology from the FIEBRE case-control data and compare these with the approach used in Chapter 8. I explore the strengths and limitations of each approach through an application to five infections. For each infection considered, I present the percentage of fever cases attributed to each infection (the PAF) from two different methodological approaches.

Chapter 8 presented the results of a Bayesian analysis of a case-control study to estimate the causes of fever, hereafter referred to as the Bayesian approach. The approach accounted for imperfect diagnostic test accuracy by incorporating prior information on diagnostic test sensitivity in a Bayesian latent class model. An alternative method, that is a simpler approach and requires more assumptions does not incorporate prior information on diagnostic test sensitivity. Instead, this approach assumes the sensitivity of each diagnostic test is perfect and is hereafter referred to as the frequentist approach. While I refer to these two approaches as frequentist and Bayesian the key difference I am comparing between these approaches is their assumption regarding diagnostic test sensitivity as opposed to a comparison of frequentist and Bayesian statistics more generally.

The frequentist approach discussed in this chapter was implemented by FIEBRE statistician Christian Bottomley in Stata. I wish to thank Dr Bottomley for sharing these results. For the application of the frequentist approach, three age strata were considered, under 5 years, 5 to 14 years

and 15 years and over. To be able to compare the results from this approach with those from the Bayesian approach used in Chapter 8, I re-ran the Bayesian analyses for the five infections considered with three age strata instead of the two presented in chapters 7 and 8.

9.2 Frequentist approach model specification

The frequentist approach I will present is similar to that discussed in section 6.3 applied to the GEMS study [164]. The estimand of interest is the same as in Chapter 8 but will be referred to as population attributable fraction (PAF) as in Levine *et al.* [160] and to distinguish the results from the frequentist approach to those from the Bayesian approach. The PAF used here, is defined as the proportion of fever cases caused by a particular pathogen and that would be theoretically eliminated if that pathogen were not present in the population. The PAF is calculated using Miettinen's formula [146] as in 6.2.2. The odds ratio (OR_j) for a pathogen j denotes the odds ratio comparing the odds of infection by pathogen j between fever cases (inpatients and outpatients combined) denoted $Y = 1$ and controls ($Y = 0$). Assuming the risk ratio can be approximated by the odds ratio [145], then the population attributable fraction for each pathogen is estimated by:

$$PAF_j = \eta_j(1 - 1/OR_j) \tag{9.2.1}$$

where, η_j denotes the proportion of fever patients positive for pathogen j (i.e. $\Pr(j = j|Y = 1)$). The odds ratio for infection by a pathogen between cases and controls for each pathogen is estimated by logistic regression. If the only covariate included is the test result denoting the presence or absence of pathogen j, then, to estimate the odds ratio, the fitted model is of the form:

$$\log\left(\frac{\Pr(Y = 1|X_j)}{1 - \Pr(Y = 1|X_j)}\right) = \alpha + \beta_j X_j \tag{9.2.2}$$

where, α is the intercept, X represents the presence or absence of pathogen j from the diagnostic test and β is the corresponding coefficient. In the examples to follow, each pathogen is considered as a dichotomous variable (either an individual is infected or not) so the estimated coefficients from the logistic regression model are an estimate of the odds ratio (i.e. $\exp(\beta_j) = OR$) [164].

To account for the matched case-control study design and adjust for confounding, logistic regression was adjusted for sex, year of recruitment (2018,2019,2020/2021), season of recruitment (quarter), geographical location (from GPS data using 2 or 3 clusters of GPS coordinates identified in each country) [20], and additional potential confounders (HIV status). Let the number of additional covariates considered be denoted W, then the fitted model including these additional independent covariates is now:

$$\log\left(\frac{\Pr(Y=1|X_j)}{1-\Pr(Y=1|X_j)}\right) = \alpha + \beta_j X_j + \beta_{w=1} X_{w=1}, ..., \beta_W X_W \quad (9.2.3)$$

where, $\exp(\beta_j)$ still denotes the adjusted odds ratio comparing the odds of infection between fever cases and controls. When the odds ratio is adjusted then it is also the case that the estimated PAF is also adjusted. ORs were calculated separately for each site (Laos, Malawi, Mozambique and Zimbabwe) and age group. The proportion of fever cases positive for a pathogen η_j was weighted to account for variation in the proportion of recruited fever cases during months with high case numbers. The weights were calculated by dividing the total monthly count of fever cases at a site by the number of recruited cases. Confidence intervals for PAFs were calculated via non-parametric bootstrap using the percentile method.

With this approach, only a single diagnostic test can be used for each infection and diagnostic test sensitivity is assumed to be 100%. No odds ratio and no PAF was estimated when there were no positive test results in cases or controls (as was the case for malaria in Laos) and it is assumed that in this case, this pathogen does not contribute to the aetiology of fever cases. Further details of the differences between the two methods are given in Table 9.1.

9.3 Results

In this section, I present the results of both the frequentist and Bayesian approaches for five pathogens. The five pathogens considered in this comparative analysis are malaria, influenza A, influenza B, rhinovirus/enterovirus and adenovirus. For malaria, in the frequentist approach, the diagnostic test used to determine presence or absence of malaria was the RDT and for the respiratory infections the Luminex RPP. For these infections in particular, there is an argument for stratifying by three age groups instead of just two as these infections commonly present with different frequencies in infants to older children [185, 205]. Figure 9.1 shows the results for all patients (inpatients and outpatients combined)

Table 9.1: A comparison of the assumptions and definitions of key principles between the two statistical approaches

Component	**Approach**	
	frequentist	**Bayesian**
Estimand	Population attributable fraction (PAF), defined as the proportion of fever cases caused by a particular pathogen and that would be theoretically eliminated if that pathogen were not present in the population.	The same as in the frequentist approach however I have referred to this as the cause-specific case fraction
Uncertainty interval	The 95% confidence interval: defined as the proportion (expressed as a percentage) of intervals (a pair of values defining a lower and upper limit), obtained from repeated sampling in the long run, which contain the true population value for a parameter	The 95% credible interval: defined as there is a 95% probability that the true value lies within the interval
Cases and controls	Cases are assumed to be infected and controls are assumed to be not infected	The same as in the frequentist approach
Diagnostic test data	A single diagnostic test for each pathogen and tests must be applied to both cases and controls	More than one diagnostic test per pathogen can be included and test results from cases only can also be utilised
Diagnostic test accuracy	Test sensitivity is assumed to be 100%, test specificity is adjusted by the observed proportion positive among controls	Test sensitivity is assumed imperfect and described by a prior probability distribution, test specificity is estimated from the observed test positives among controls
Data requirements for estimation	At lease one positive test result among cases	The estimand can be estimated with no positives and results are driven solely by the priors
Constraints on estimand	No constraints are applied to the PAF, the total combined PAF could total greater than 100%	The total combined aetiologic fractions must sum to 100%
Ascribing unknown aetiology	Assumed that the difference between the combined PAF of each pathogen and 100% represents that proportion of fever cases with unknown aetiology	Directly incorporated into the model and represented by its own latent state

stratified by site and age group for malaria and Figure 9.2 the same for the four respiratory pathogens considered.

In this section I compare the estimates from the two different statistical models however, the uncertainty intervals estimated do have different interpretations (See Table 9.1). In the frequentist approach a 95% confidence interval (CI) is calculated while in the Bayesian approach I estimate the 95% credible interval (CrI). Further, in the frequentist approach the odds ratio is adjusted for other covariates meaning that while the estimates may be comparable something slightly different has been estimated in each approach.

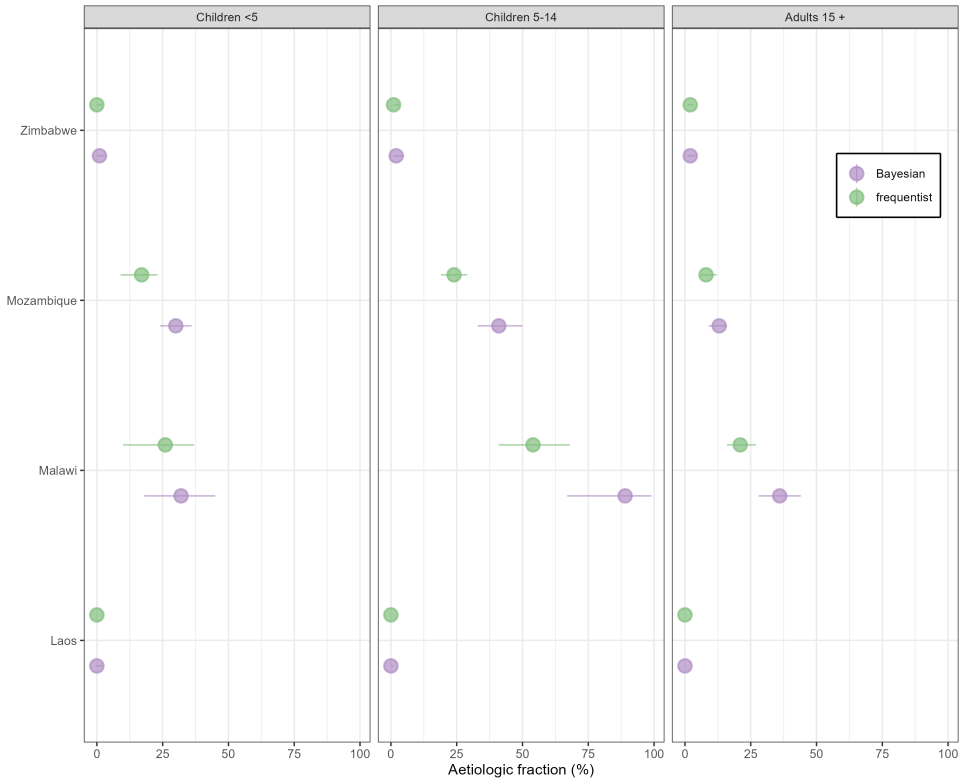
In Figure 9.1, the estimated percentage of fever cases attributed to malaria in Laos is 0% using both statistical approaches. However, while in the frequentist approach there is no confidence interval as there were no positives in cases or controls and so no OR or PAF was estimated and zero aetiology was assumed, in the Bayesian approach malaria was included in the model for Laos and so there is some uncertainty in the estimand. This is due to the use of prior distributions that assume a range of plausible values. As a result, if a test is imperfect the Bayesian approach will always estimate an uncertainty interval for the aetiologic fraction. In Malawi, there is a stark difference in the estimated percentage of fever cases due to malaria by age with both approaches estimating the largest proportion of fever cases from malaria in children 5 to 14 years. The Bayesian approach attributed 89% (95% CrI: 67-99%) of fever cases in this age group to malaria and the frequentist approach attributed 54% (95% CI: 41-68%). However, although both statistical approaches attribute the largest proportion of fever cases to malaria in the 5-14 year age group, the estimates from these two approaches differ with a higher estimate from the Bayesian approach. A similar trend in age is seen in Mozambique with both statistical approaches estimating the largest proportion of fever cases attributed to malaria in the 5 to 14 years age group but with a higher estimate from the Bayesian approach with uncertainty intervals from both approaches that do not overlap.

Across all sites excluding Laos, where the estimated percentages were 0%, the median estimate is higher from the Bayesian statistical approach that accounted for imperfect diagnostic test sensitivity. That is, more fevers were attributed to malaria under this assumption. The estimated sensitivity of the malaria RDT from the Bayesian approach was 77% (95% CrI: 68-87%). If this is true, and the RDT is imperfect, then the estimates from the frequentist approach would be an underestimate of the true percentage of fevers attributed to malaria as this estimation does not

account for those individuals with a false negative result. This explains why the point estimates are lower with the frequentist approach. The magnitude of bias is larger in those sites where there is a larger proportion of fever cases attributed to malaria than not. For example, in Malawi and compared to Zimbabwe.

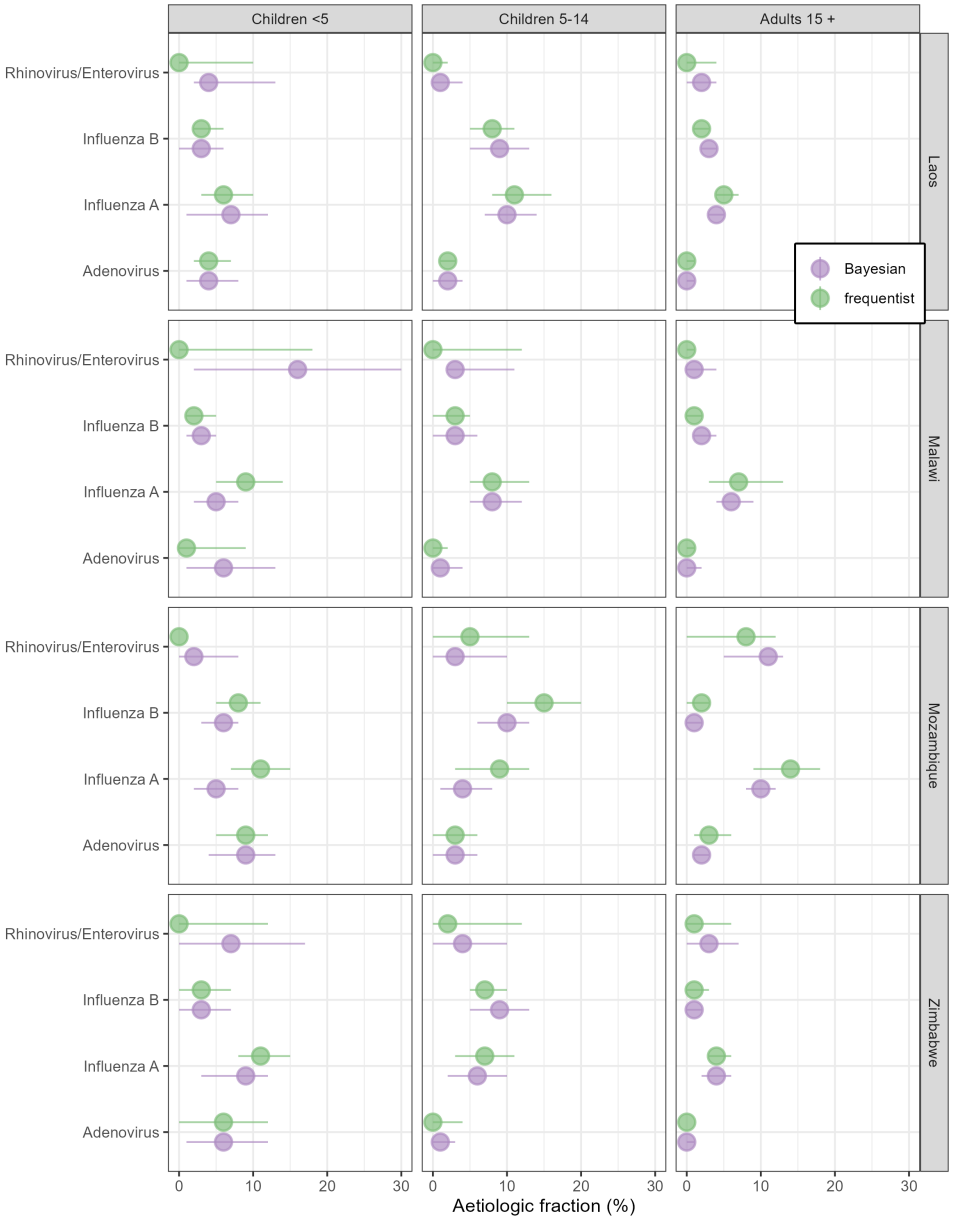
Another difference in the estimates of aetiologic fractions for malaria between these two approaches is the width of the uncertainty interval (confidence or credible interval). Accounting for the diagnostic test sensitivity in the Bayesian approach inherently comes with greater uncertainty as we do not assume a know sensitivity of the malaria RDT. As a result, in comparison to the frequentist approach that assumes a fixed sensitivity of 100%, there is greater uncertainty in the estimates from the Bayesian approach.

Figure 9.1: A comparison of the estimated percentage of fever cases attributed to malaria by statistical approach



Similar to the findings from the malaria comparison, an overall trend in the difference between estimates from the two statistical approaches for the respiratory infections is a wider uncertainty interval for those estimates from the Bayesian approach, which accounts for diagnostic test uncertainty (Figure 9.2). However, in contrast to the malaria example, accounting for the imperfect diagnostic test accuracy of the Luminex RPP does not always result in a higher point estimate. A higher point estimate under the Bayesian approach would be expected if the test was

Figure 9.2: A comparison of the estimated percentage of fever cases attributed to respiratory pathogens by statistical approach



not 100% sensitive. In fact, for influenza A in all stratum except children under 5 years in Laos, the opposite is found: a higher point estimate is found from the frequentist model that assumes the Luminex RPP is 100% sensitive. For all respiratory infections, the estimated sensitivity of the Luminex RPP was 95% (95% CrI: 88-99%). Given the high but imperfect sensitivity, this means that there are some false negatives (but not as many as with the malaria RDT that was less sensitive). Consequently, while the estimate of the percentage of fevers attributed to each respiratory pathogen is still routinely underestimated, this underestimation is much smaller when the diagnostic test sensitivity is higher and the uncertainty intervals for all estimates are overlapping.

Overall, estimates for respiratory pathogens are similar irrespective of the statistical approach taken with confidence and credible intervals overlapping in all instances. Rhinovirus/enterovirus in children under 5 years is the respiratory pathogen where there appears the least concordance between the two methods, notably in Malawi and Mozambique. In Mozambique, the specificity of the diagnostic test used (Luminex) in children 0 to 4 years was estimated at 70% (95% CrI: 66-74%), in children 5 to 14 years 83% (95% CrI: 79-88%) and in adults it was estimated to be 93% (95% CrI: 89-96%). A similar trend was seen across all sites where the specificity was lowest in the youngest age strata and this was the strata that displayed the least concordance between estimates from the two statistical approaches. This may explain why the estimates between methods are more dissimilar as the magnitude of both bias and variance has been shown to increase as specificity decreases with the frequentist approach [37].

In the frequentist approach, rhinovirus/enterovirus had an odds ratio of less than 1 in the youngest age stratum for all sites. An odds ratio of less than one indicates that the odds of fever when the pathogen is absent was greater than the odds of fever when the pathogen is present (i.e there was a large proportion of controls positive for the pathogen). This is in line with the estimated specificity of the diagnostic test in this age group from the Bayesian approach.

9.4 Discussion

There is more than one statistical approach that can be used to estimate fever aetiology from imperfect diagnostic test data arising from a case-control study design. In this chapter I compared two approaches. The first, presented in chapter 8, does not assume diagnostic test sensitivity is perfect and instead incorporates available information on the diagnostic test sensitivity in the form of prior distributions. The second, frequentist approach assumes sensitivity is fixed at 100%. The two approaches are presented and discussed through application to five different pathogens investigated as causes of fever cases in the FIEBRE study.

A key limitation of what has been called the frequentist approach, is that only a single diagnostic test can be used to infer infection presence or absence. While for the respiratory pathogens only a single diagnostic test was used, for malaria we have observed results from two diagnostic tests and in FIEBRE more broadly, many pathogens were diagnosed by multiple diagnostic tests. In the results presented in Figure 9.1, I compare

the Bayesian single-pathogen model results with the estimates from the frequentist model. The Bayesian single-pathogen model can integrate the results from both the RDT and microscopy tests used to diagnose malaria however the frequentist model can only use a single test. Because of these fundamental differences in the approaches the input data for each model differs which further helps to explain the large differences in some of the estimated aetiologic fractions.

Table 9.2 shows the estimated percentage of fever cases attributed to malaria from each statistical approach as well as estimates from a Bayesian model that assumes imperfect diagnostic test sensitivity for the RDT but does not use the additional microscopy data. Estimates of the percentage of fever cases attributed to malaria from the two statistical approaches just using the RDT data (Table 9.2) are more similar. In all strata, estimates from the frequentist model, that assumes the RDT is perfect, attributed the fewest fever cases to malaria as a result of bias from incorrectly assuming a perfect diagnostic test. The Bayesian model that assumes the RDT is imperfect but only uses the RDT test data then has higher or the same estimates as the frequentist model and finally, the Bayesian model that utilises both the RDT and microscopy attributes the greatest proportion of fever cases to malaria. Herein lies an advantage of the latter approach, combining two tests, one with a large sample, carried out in almost all cases and controls, but has a lower specificity (RDT) and another with a smaller sample size, carried out in only a subset of cases and controls but has a high specificity (microscopy) (See Table 8.1). By utilising both sets of test results, the credible intervals from the Bayesian models in table 9.2 are narrower than only using the RDT demonstrating the value of combining evidence from multiple diagnostic tests which is only possible with the Bayesian approach. The result that combining data from more than one diagnostic test for the same pathogen, leads to estimates with greater certainty was also shown in a simulation study from the PERCH research group [37].

In this relatively small example, the other situation in which disparate results were found from the two statistical approaches was when there was a test with poorer specificity and the odds ratio approached 1. Under the frequentist approach, estimated odds ratios of less than one result in no aetiology assigned. However, this does not take into account the statistical uncertainty in the estimate of the odds ratio itself. While this is a limitation of the frequentist method there are strengths. In the results presented in this chapter, odds ratios were estimated from logistic regression that not only adjusted for the matching variables (sex, age, location

Table 9.2: Estimates of percentage of fever cases attributed to malaria from three different models by site and age group

Age group	Approach	Included data	Laos	Malawi	Mozambique	Zimbabwe
< 5	frequentist	RDT	0	26 (10-37)	17 (9-23)	0 (0-2)
	Bayesian	RDT	0 (0-2)	27 (14-40)	27 (21-35)	1 (0-2)
	Bayesian	RDT+microscopy	0 (0-2)	32 (18-45)	30 (24-36)	1 (0-3)
5-14	frequentist	RDT	0	54 (41-68)	24 (19-29)	1 (1-3)
	Bayesian	RDT	0 (0-1)	74 (55-97)	37 (29-47)	1 (0-4)
	Bayesian	RDT+microscopy	0 (0-0)	89 (67-99)	41 (33-50)	2 (1-5)
15+	frequentist	RDT	0	21 (16-27)	8 (6-12)	2 (1-3)
	Bayesian	RDT	0 (0-1)	31 (24-41)	11 (8-15)	2 (0-3)
	Bayesian	RDT+microscopy	0 (0-0)	36 (28-44)	13 (9-16)	2 (1-4)

and season of recruitment) but also the potential confounder HIV status. Estimating odds ratios from logistic regression models adjusted for other variables is a relatively straightforward process but implementing the same regression adjustments within the latent class Bayesian approach is not as simple. In the Bayesian model presented in this thesis, there is no adjustment for sex, season, recruitment or HIV status though there are extensions that could allow for this [151].

These fundamental differences in model formulation make a direct comparison between the results challenging. This difference may also explain why point estimates for certain respiratory pathogens from the frequentist model are higher than from the Bayesian model. This finding is in contradiction to the presumption that PAFs will be underestimated when diagnostic test sensitivity is imperfect compared to the Bayesian model. Having said this, higher point estimates may also be possible in the frequentist model compared to the Bayesian model by chance as opposed to differences in the models.

The estimates from both statistical approaches discussed here are from models that estimate aetiologic fraction of each cause (or risk factor) in isolation. Both methods can be extended to account for multiple possible causes. In the Bayesian approach, this was discussed in section 8.5 and for the frequentist approach this has been discussed in Blackwelder *et al.* [164]. Briefly, to estimate the population attributable fraction for malaria in the presence of influenza A (fluA), a multiple logistic regression model is fit that includes the interaction between the two infections.

In its simplest form, for a single strata, this is:

$$\log\left(\frac{\Pr(Y=1|\mathbf{X})}{1-\Pr(Y=1|\mathbf{X})}\right) = \alpha+\beta_{malaria}X_{malaria}+\beta_{fluA}X_{fluA}+\gamma X_{malaria}X_{fluA} \tag{9.4.1}$$

where, $\mathbf{X}$ denotes both exposures (malaria and Influenza A) and the coefficient γ represents the interaction between malaria and influenza A. Then, the population attributable fraction in 9.2.1 can be extended to the adjusted population attributable fraction for malaria, defined as:

$$PAF_a = \\ \eta_{malaria}(1-1/exp(\beta_{malaria})) + \eta_{malaria/fluA}(1-1/exp(\beta_{malaria}+\gamma)) \tag{9.4.2}$$

where, $\eta_{malaria}$ denotes the prevalence of malaria in the cases and $\eta_{malaria/fluA}$ denotes the combined prevalence of having both malaria and Influenza A in the cases. Under both statistical approaches, the assumption of each cause being treated in isolation can lead to under- or over-estimates of aetiologic fractions with the magnitude of bias a function of the strength of association between two or more causes [206].

Out of the 26 different pathogens considered in Chapter 8, only five have been considered here. For many pathogens in FIEBRE, multiple diagnostic tests were used and, for many of these tests not all cases and controls have an observed result. The frequentist approach is only applicable when we have observed results from both cases and controls from a single diagnostic test. For pathogens with multiple diagnostic tests, a single diagnostic test could be used and the frequentist approach can proceed as normal however this may also have limitations in comparison to utilising all the available data as seen in the malaria example. However, the frequentist approach could not estimate a PAF from diagnostic tests not carried out on both cases and controls. Altogether, the frequentist approach has a limited applicability with the complex dataset that has arisen from the FIEBRE study. If the frequentist approach were to be utilised, it would have to be done so in combination with other analytical approaches [165].

Overall, when diagnostic test sensitivity and specificity are both high or when the odds ratio is large, the frequentist approach approximates the Bayesian approach. However, if sensitivity is less than perfect, the frequentist method is biased and underestimates aetiology with the magnitude of bias depending on how imperfect the test is. Further, in those cases where the odds ratio is less than 1, the magnitude of bias and variance increases as specificity decreases. Having said this, accounting

for imperfect sensitivity if prior distributions are used that are incorrect, estimates of fever aetiology from this approach could also be biased. In conclusion, if there is a suggestion that a diagnostic test is imperfect, the approach presented in chapter 8 should be used to estimate aetiologic fractions.

Part III

Thesis Summary

Chapter 10

Summary of findings and future directions

10.1 Overview

The overall aim of this thesis, was to develop the application of Bayesian latent class models for investigating attribution of a syndrome to particular pathogens. The methods were then used to provide estimates of the extents to which fever-related illness could be attributed to different fever-causing pathogens in four countries using diagnostic test data from fever cases and controls. This chapter outlines the main results as well as the implications of these findings in the context of each of the following objectives:

1. Review how diagnostic test accuracy can be estimated in the presence of imperfect reference tests using Bayesian latent class analyses (Chapter 2)
2. Explore the impact of key latent class model assumptions on estimates of diagnostic test sensitivity and specificity through a simulation study (Chapter 3)
3. Estimate the sensitivity and specificity of diagnostic tests used in the FIEBRE study using Bayesian random-effect latent class meta-analyses (Chapters 4 and 5)
4. Review aetiology research methodologies in case-control studies (Chapter 6)
5. Estimate the aetiology of fever in the FIEBRE dataset using Bayesian partial latent class analysis (Chapters 7 and 8)
6. Objective 6: Compare the strengths and limitations of two different statistical approaches to estimating fever aetiology

Finally, I outline some considerations for future aetiology research.

10.2 Key findings

10.2.1 Objective 1: Review how diagnostic test accuracy can be estimated in the presence of imperfect reference tests using Bayesian latent class analyses

Diagnostic tests are a mainstay of clinical decision-making, particularly for a syndrome like fever that often lacks obvious signs or symptoms to pinpoint aetiology. However, for many infections diagnostic tests are imperfect [41]. With imperfect diagnostic tests, it can be unclear which tests should be recommended for use in clinical practice and clinicians may be uncertain how to interpret imperfect diagnostic test results. This may impact the care a patient receives, whether this be length of time before a diagnosis or in the appropriateness of the treatment provided. Consequently, understanding the accuracy of a diagnostic test is paramount and for this, there must be robust statistical methods to estimate diagnostic test accuracy.

In Chapter 2, I highlighted that there are various methods available to estimate the accuracy of diagnostic tests. However, when you have an imperfect reference test, it is recommended to use a latent class analysis. Latent class analysis can remove biases that arise from alternative approaches to estimating diagnostic test accuracy in the presence of imperfect reference tests such as composite reference standards and discrepancy resolution [25]. I described the steps involved in latent class analysis for diagnostic test accuracy estimation and the associated challenges which include identifiability and whether or not to assume conditional independence between test results on an individual given the true disease status of that individual.

10.2.2 Objective 2: Explore the impact of key latent class model assumptions on estimates of diagnostic test sensitivity and specificity through a simulation study

In chapter 2, I introduced the conditional independence assumption, which is the assumption that the result of one test from an individual provides no information about the result of another test on the same

individual for the same pathogen (conditional on the true disease status of the individual). It is clear, that in practice this assumption is unlikely to hold.

In chapter 3, I evaluated the impact of not accounting for conditional dependence and accounting for conditional dependence in one or both disease states on estimates of diagnostic test sensitivity and specificity through a simulation study and the re-analysis of a published study [22]. This work found that bias and poor coverage were symptoms of misspecifying the conditional dependence structure.

When evaluating the predictive performance of the models considered in the re-analysis of a published study, this research highlighted a lack of consensus with regards to model evaluation of latent class models used for diagnostic test accuracy. Previous works have used a combination of Akaike Information criterion, Deviance Information Criterion and Bayesian p-values [75, 89]. In this thesis, I used a relatively novel approach called leave-one-out cross validation using Pareto smoothed importance sampling. This approach was chosen as it uses the entire posterior distribution to calculate the log predictive density and in doing so, more completely takes account of the uncertainty in our parameter estimates than any other approach.

From the findings of this simulation study, I have recommended that conditional dependence should be accounted for between diagnostic test results of an individual given their true disease status. This approach should be used even if researchers are unsure of the presence of conditional dependence or if dependence is only expected at minimal levels. The reason for this recommendation is due to the minimal loss in precision seen by using the more complex model even if the truth is that the test results from an individual are independent, given the individual's true disease status.

10.2.3 Objective 3: Estimate the sensitivity and specificity of diagnostic tests used in the FIEBRE study using Bayesian random-effect latent class meta-analyses

Most diagnostic test accuracy meta-analyses published to date have not accounted for an imperfect reference test [90, 91]. However, many of these diagnostic tests are imperfect and assuming a perfect reference test, when in truth the reference test is imperfect, results in biased estimates of test accuracy [50]. As a result, inferences from meta-analyses that assume a

perfect reference standard should be made with caution.

In this thesis, I carried out multiple diagnostic test accuracy meta-analyses to estimate the sensitivity and specificity of tests using an extension to the Hierarchical Summary Received Operating Curve model [89]. These results have provided estimates of the accuracy of diagnostic tests, some for the first time, and can help users of these tests interpret the results and guide policy makers to ensure the most useful diagnostic test is being recommended for use. The variety of diagnostic tests and pathogens explored via meta-analyses highlight the broad applicability of this method.

The predicted test accuracy from the meta-analyses presented in this thesis highlight that heterogeneity is common in diagnostic test accuracy meta-analyses. These findings were not a surprise as there are many reasons why a diagnostic test may be more sensitive or specific in a particular study including different laboratories with different technical staff, different background prevalence's and different cohorts, for example, young versus old or symptomatic versus healthy. For a diagnostic test accuracy meta-analysis to be most informative, studies must report all details of the study design so that a meta-analysis can incorporate this information to account for the expected heterogeneity.

10.2.4 Objective 4: Review aetiology research methodologies in case-control studies

In chapter 6, I discussed previous examples of aetiology research in case-control studies where the goal was to estimate the attributable fraction of specific exposures to an outcome. Since the mainstay case-control studies of Doll and Hill [137, 207] odds ratios have been used in epidemiology research to measure the association between exposure and outcome. The use of the odds ratio from case-control studies in the population attributable fraction came almost twenty years later [146] but there has been little change in the methodological approach to attributable fractions since. The PERCH study was the first to estimate the attributable fraction of various exposures to an outcome with a different approach [37].

The partial latent class analysis model developed to estimate the attributable fraction of exposures to an outcome was used in the PERCH study of pneumonia. This thesis is the first piece of work to apply this method outside of the pneumonia setting.

10.2.5 Objective 5: Estimate the aetiology of fever in the FIEBRE dataset using Bayesian partial latent class analysis

In chapter 8, I outline the Bayesian partial latent class analysis, extended from Wu *et al.* [168] to the FIEBRE study to estimate the cause-specific case fractions for each pathogen considered in FIEBRE. Extensions included exploring single-pathogen versus multi-pathogen model results, not assuming any test has 100% specificity and utilising estimates of diagnostic tests accuracy from meta-analyses to derive prior distributions for diagnostic test sensitivity. Results from this analysis are in line with previous findings. I found that malaria remains a leading cause of fever in Malawi and dengue in Laos. There was no single pathogen that dominated fever cases in Mozambique and in Zimbabwe our results were affected by an outbreak of typhoid fever.

A recommendation from Chapter 3 was that conditional dependence should be accounted for between diagnostic test results (for the same pathogen) of an individual given their true disease status. However, in practice this was difficult to implement. I attempted to include conditional dependence between the two malaria tests in those individuals with malaria in the multi-pathogen model but I found that this introduced convergence issues without reparameterization or additional prior information. Future work would benefit from investigating this so that conditional dependence can be accounted for between tests for the same pathogen given an individual's true disease status.

Overall though, these results suggest that key causes of fever across all sites have not been tested for in the FIEBRE study. In Laos children and adults, Malawi adults and Mozambique adults, less than 50% of fever cases were attributed to a pathogen and in Zimbabwe children and adults, it was just over 50%. These findings indicate that there are either other important causes of fever that were not tested for in FIEBRE or that the diagnostic tests used are less sensitive than previously thought.

From the results presented in this thesis, I have shown that the partial latent class analysis approach is applicable to syndromes other than pneumonia and is a novel approach that offers great potential for other large aetiology studies with complex data.

10.2.6 Objective 6: Compare the strengths and limitations of two different statistical approaches to estimating fever aetiology

In chapter 9, I compare the cause-specific case fractions or 'PAFs' estimated from two different models for five pathogens. The first method was the Bayesian, partial latent class approach and the second was the traditional frequentist approach. The latter estimates attributable fractions using the odds ratio comparing the odds of infection with each pathogen between fever cases and controls estimated from unconditional logistic regression adjusted for the matching variables.

Where only a single diagnostic test was used to detect infection with a pathogen and where this test was highly sensitive and specific, results from the two approaches were similar (Figure 9.2). However, when more than one diagnostic test was used and/or where the diagnostic test/s used are imperfect, estimates from the frequentist approach suffered from biased estimates of the cause-specific case fraction due to incorrectly assuming the diagnostic test is 100% sensitive (Figure 9.1). Furthermore, when diagnostic tests used in a study are imperfect, the ability to incorporate additional diagnostic test data for the same pathogen is particularly beneficial but not possible with the traditional frequentist approach discussed here.

Overall, I recommend using the partial latent class analysis approach when more than one diagnostic test is used to detect each pathogen and when the diagnostic tests used are known to be less than 100% sensitive.

10.3 Future work

10.3.1 Diagnostic test accuracy estimation in the presence of imperfect reference tests

In low resource settings diagnostic tests that are often seen as the 'gold-standard' for a particular infection are typically not used due to poor accessibility and high associated costs. Instead, imperfect diagnostic tests are used. Traditionally, diagnostic test accuracy has been evaluated against perfect reference tests however, given their poor availability in many settings, this also means that diagnostic tests are often evaluated from individuals in high-resource settings, where the test can be easily evaluated against a gold-standard reference test. These accuracy estimates may not be applicable in different settings and so it would be useful to estimate the accuracy of tests using data from different settings.

For this, it is paramount that there are statistical methods to estimate diagnostic test accuracy in the presence of imperfect reference tests, which are more likely to be used in practice in many settings.

A new application called *MetaBayesDTA* [118] was released in May 2023. This web application allows researchers to run both the bivariate and HSROC models (Section 4.3) without having to write their own code. Tools like this increase the accessibility of these advanced statistical methods to a wide variety of researchers which is of great benefit. However, while the application can run both bivariate and HSROC models most of the additional functionalities, like the ability to carry out subgroup analyses, are only available for the bivariate model. Unfortunately, their bivariate model assumes a perfect reference standard. This assumption leads to biased estimates of test sensitivity and specificity if in fact the reference test is imperfect [22] which, as I have discussed, is highly likely. The work presented in this thesis used R and Stan and the model code is available via github so that future researchers could adapt this to their own research.

Also in 2023, an updated version of the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [86] was released which includes new and updated guidance on all stages of meta-analysis for diagnostic test accuracy. This update highlights the continued importance of meta-analysis research and the importance of the statistical methods that underpin the analysis. Meta-analyses remain a mainstay method in medical research and so continued methodological improvements will only serve to improve the outputs of this relied upon tool. Further developments to the Cochrane Handbook should include guidance on model evaluation in the diagnostic test accuracy meta-analysis setting which is currently lacking.

Recent criticisms of latent class models for diagnostic test accuracy estimation include a study that showed estimates of sensitivity and specificity are biased by assuming a 2-state latent class model if in fact, the truth is a 3-state scenario [72]. However, this paper also assumed conditional independence. Further simulations should explore whether estimates of sensitivity and specificity are biased by assuming a 2-state latent class model with conditional dependence [72]. It would also be interesting to compare the estimates from a latent class model that uses dichotomoized data on test results with a latent class model that uses the raw continuous data (examples include counts of malaria parasites or the optical density recorded from an ELISA for leishmaniasis). A latent class analysis to estimate test accuracy using continuous data is used in the

diagnosis of *Schistosoma haematobium* by haematuria-based diagnostic techniques. This work utilised the continuous data of observed counts of parasites [208] which could then be compared with the dichotomized version of presence of any parasites versus absence.

10.3.2 Fever aetiology estimation

In chapter 8, I described and implemented a partial latent class model to estimate the cause-specific case fraction for 26 different pathogens. This analysis approach was first developed for a study investigating the causes of another syndrome: pneumonia [37]. To my knowledge, this is the first attempt to use this method in a different aetiology application, this time looking at the causes of fever.

In this thesis, the partial latent class model was implemented in Stan, a Bayesian statistical modelling software. The PERCH study that initially developed the model implemented this in JAGS and wrote an R package called *baker* [209]. While in theory having a publicly available R package should make it easier for other researchers to run these models, I found I could not get the R package to work for FIEBRE data even after communication with the package creators. Implementing these models is not straightforward and so advancing the R package would be of great benefit for other aetiology studies. However, given the time and money required for such large aetiology studies one can see the cost benefit to spending more time improving the R package. Given the computing time required for these models, it may also make sense to switch from an R package reliant on JAGS to a software like Stan with much quicker computation time.

The FIEBRE study performed tests for a large number of pathogens, meaning that it was perhaps surprising that this study was not able to attribute a larger proportion of fever cases to one of the sought pathogens and begs the question what pathogens are causing health-care seeking fevers in these settings. Future studies that encompass different pathogens or currently unknown pathogens would be of interest. It is also unfortunate that there was an outbreak of typhoidal *Salmonella* in Zimbabwe during the study period meaning the results are not necessarily representative of what normally causes fevers in the study population in Zimbabwe.

Overall, the results of this study are important for two reasons. The first is at the individual level. If these results are used to update local clinical guidelines, when a patient with undifferentiated fever comes to a clinic, the clinician can use the guidelines to aid in their decision-making

of how best to treat a patient given the probability of infection with a particular pathogen. The second is a broader public health goal helping in the planning of prevention and treatment programs and highlighting where improved diagnostics are needed. Respiratory pathogens were a key cause of fever particularly in children across all FIEBRE study sites. If a point-of-care rapid multiplex respiratory pathogen panel was available this would definitely be useful in these settings. The advantages of such a test and the identification of viral respiratory pathogens would be the exclusion of bacterial pathogens and the reduced use of unnecessary antibiotics, which precipitate antimicrobial resistance.

10.4 Concluding remarks

Health care workers want to provide appropriate and timely care for patients with fever but unpicking the true cause of a fever is difficult. Not only can a fever be caused by many different pathogens but many of these pathogens present as undifferentiated febrile illness. As a result, health care workers are reliant on the diagnostic tests available at the point-of-care and the results of those diagnostic tests being informative as to the true cause of a patient's fever. When limited diagnostic tests are available and diagnostic tests are imperfect, this presents a challenge.

In many settings, there remains limited access to diagnostic tests, despite recognition at the highest levels that diagnostics are fundamental to delivering quality health care [47]. To improve the understanding of fever aetiology improving access to diagnostics must be tackled.

In this thesis, I have investigated latent class models for use in aetiology research of fever. I have combined multivariate imperfect binary diagnostic test data from cases and controls or in some instances from cases alone, to estimate attribution of fever cases to different pathogens across four sites in Africa and Asia.

It is my hope that the work in this thesis will be useful to other researchers and so it was a key endeavour to ensure resources were made publicly available. In line with this, I have made all model code publicly available on github and published manuscripts in open access journals to ensure these resources are available to any who may wish to use them.

Appendices

Appendix A

Influenza and RSV diagnostic test accuracy meta-analysis

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1702488	**Title**	Miss
First Name(s)	Suzanne		
Surname/Family Name	Keddie		
Thesis Title	Latent class models for diagnostic test accuracy with application to fever aetiology		
Primary Supervisor	John Bradley		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMC Infectious Diseases		
When was the work published?	13/10/2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

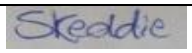
Where is the work intended to be published?	

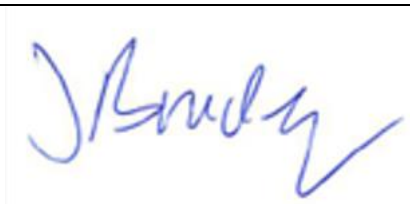
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am joint first author of the paper. I developed, coded and carried out the statistical analysis with support from my supervisors. I also created data visualizations of the results for the manuscript and provided considerable input into the drafting of the manuscript.

SECTION E

Student Signature	
Date	06/09/2023

Supervisor Signature	
Date	25/03/2024

BMC Infectious Diseases

RESEARCH

Open Access

Diagnostic accuracy of multiplex respiratory pathogen panels for influenza or respiratory syncytial virus infections: systematic review and meta-analysis

Sophie Jullien[1*†], Felicity Fitzgerald[2†], Suzanne Keddie[3†], Oliver Baerenbold[2], Quique Bassat[1,4,5,6,7], John Bradley[3], Jane Falconer[3], Colin Fink[8,9], Ruth Keogh[3], Heidi Hopkins[3†] and Marie Voice[8†]

Abstract

Respiratory syncytial virus (RSV) and influenza viruses are important global causes of morbidity and mortality. We evaluated the diagnostic accuracy of the Luminex NxTAG respiratory pathogen panels (RPPs)™ (index) against other RPPs (comparator) for detection of RSV and influenza viruses. Studies comparing human clinical respiratory samples tested with the index and at least one comparator test were included. A random-effect latent class meta-analysis was performed to assess the specificity and sensitivity of the index test for RSV and influenza. Risk of bias was assessed using the QUADAS-2 tool and certainty of evidence using GRADE. Ten studies were included. For RSV, predicted sensitivity was 99% (95% credible interval [CrI] 96–100%) and specificity 100% (95% CrI 98–100%). For influenza A and B, predicted sensitivity was 97% (95% CrI 89–100) and 98% (95% CrI 88–100) respectively; specificity 100% (95% CrI 99–100) and 100% (95% CrI 99–100), respectively. Evidence was low certainty. Although index sensitivity and specificity were excellent, comparators' performance varied. Further research with clear patient recruitment strategies could ascertain performance across different populations.

Protocol Registration: Prospero CRD42021272062.

Keywords: Respiratory syncytial virus, Influenza virus, Diagnosis, Molecular diagnostics, Respiratory infection

Background

Respiratory syncytial virus (RSV) and influenza viruses are important causes of global morbidity and mortality. An estimated ~ 33 million episodes of RSV occur annually in children under five years of age, causing at least 3.2 million hospital admissions and 59,600 in-hospital deaths, with an even higher unmeasured community burden in low-resource settings [1–3]. Influenza is estimated to cause up to 650,000 deaths per year, the majority in low-resource settings and in people over 75 years old [4]. However, current estimates suggest that up to 100,000 deaths from influenza occur annually in children under five years old [4, 5]. Post-mortem studies in childhood deaths under the age of five also show a higher than expected burden of these two pathogens [6]. Given that RSV also causes significant mortality in the elderly, these two viruses pose significant health risks throughout the human lifespan [7].

†Sophie Jullien, Felicity Fitzgerald and Suzanne Keddie contributed equally to the work and should share first authorship

†Heidi Hopkins and Marie Voice contributed equally to the work and should share senior authorship

*Correspondence: sophjullien@gmail.com

[1] Barcelona Institute for Global Health (ISGlobal), Hospital Clínic - Universitat de Barcelona, Carrer Rosselló 132, 08036 Barcelona, Spain
Full list of author information is available at the end of the article

The need for rapid, accurate diagnostics for these pathogens is threefold. Firstly, for the purposes of antimicrobial stewardship: the symptoms of respiratory tract infections are non-specific, and empiric antibiotics are frequently commenced to cover possible bacterial pneumonia [8]. Rapid viral tests can therefore reduce the unnecessary prescription of antibiotics, although viral detection does not exclude bacterial co-infection [9–11]. Secondly, to confirm the specific viral cause of illness and guide commencement (or cessation) of specific antiviral therapy [12]. Finally, rapid diagnostic tests have a crucial role in determining need for infection control prevention.

In recent years there has been a rise in the use of proprietary multiplex respiratory pathogen panels (RPP) in routine clinical setting, using a range of technologies, which have increasingly replaced in-house individual real-time polymerase chain reaction (RT-PCR) assays for clinical diagnostics [13]. This follows improved turn-around time, reduced number of manual steps in the laboratory, and the multiplexing of several pathogens within a single panel, alongside a continuous evolution of regional regulatory standards [14, 15]. Luminex NxTAG RPP™ is one such panel, offering high throughput of up to a hundred samples per run with the potential to test for up to 21 viral and bacterial pathogen genes simultaneously in each sample, improving both turn-around time and cost-effectiveness [16].

The Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) is a prospective observational study of the infectious causes of fever at four sites in Africa and Asia, collecting data and samples from inpatients, outpatients and community controls [17]. FIEBRE focuses on illnesses deemed preventable or treatable; respiratory pathogens of interest include RSV and influenza viruses. The Luminex NxTAG RPP™ on respiratory samples was chosen as the reference standard for detecting these infections in the FIEBRE study. Firstly, it is CE marked for in vitro diagnostic (IVD) use and is internally verified by the assigned United Kingdom Accreditation Service accredited laboratory [18]. Secondly, its high-throughput platform allows for 96 samples to be analysed per run, multiplexing 21 genes (hence testing for up to 21 pathogens at once). In this systematic review and meta-analysis, we aimed to evaluate the diagnostic accuracy of the Luminex NxTAG RPP™ in comparison to other RPP for the detection of RSV and influenza viruses in respiratory samples. This systematic review is part of a series conducted by the FIEBRE research team, with the purpose of determining the accuracy of reference tests used to diagnose infectious causes of fever.

Methods

Inclusion criteria

We included observational and interventional studies that reported findings of the Luminex NxTAG RPP™ assay performed to detect influenza A/B viruses and RSV in respiratory samples from children (aged 2 months and older) and adults attending healthcare settings. We first intended to include studies testing for the Luminex NxTAG RPP™ assay in patients with reported or documented fever, but we found no such study. We broadened our inclusion criteria, therefore, to studies when patients were tested with the Luminex NxTAG RPP™ assay (index test I) and at least one other RPP as comparator (C). We excluded studies describing in vitro identification of viruses as opposed to detection in clinical samples and studies that did not provide data from which we could extract a binary classification table (I+/C+, I−/C+, I−/C− and I+/C−).

Search methods

An experienced library information specialist (JF) compiled a search strategy in the OvidSP Medline database. The search strategy included strings of terms, synonyms and controlled vocabulary terms (where available) to reflect two concepts: respiratory tract infections, specifically RSV or influenza, and Luminex NxTAG RPP™. The search was limited to papers published from January 2015, when Luminex NxTAG RPP™ assay was commercialized. No other search filters or limits were added. The agreed OvidSP Medline search was adapted for each database to incorporate database-specific syntax and controlled vocabularies (Additional file 1: Annex S1). We searched the following databases on 22 September 2020: OvidSP Medline, OvidSP Embase, OvidSP Global Health, Wiley Cochrane Central Register of Controlled Trials, Clarivate Analytics Web of Science, Elsevier Scopus, Ebesco Africa-Wide Information, WHO LILACS and WHO Global Index Medicus (Additional file 1: Annex S2). We imported all citations identified by our searches into EndNote X9 software and identified and removed duplicates [19]. To identify additional eligible studies, we hand-searched the reference lists of relevant manuscripts and contacted the Luminex manufacturer.

Study selection

Two reviewers (SJ, FF) selected studies independently and in duplicate using the online tool CADIMA [20]. We performed the initial eligibility assessment of titles and abstracts identified by the search strategy, using the pre-determined eligibility criteria. We retrieved full-text copies of potentially eligible reports and contacted

researchers for further information when needed. We resolved disagreements through discussion and excluded reports not meeting criteria.

Data collection and methodological quality assessment

We piloted the data extraction form and quality assessment on two studies. For each study, using the finalized data extraction form, two reviewers (SJ, FF) independently extracted data including study design, methodology, participant and comparator test characteristics, and flow and timing of sample analysis. We contacted study investigators when data reported were unclear or insufficient to produce 2×2 tables for I+/C+, I+/C−, I−/C+ and I−/C−.

Two independent reviewers (SJ, FF) evaluated the quality of each study using the quality assessment tool for diagnostic accuracy studies (QUADAS-2), which assesses both the risk of bias and applicability to the review question for four domains: patient selection, index test, reference standard (renamed as comparator test for this review) and the flow and timing of patients through the study [21]. We resolved disagreements by discussion.

Statistical analysis and data synthesis

We extracted for each study the performance results for the Luminex NxTAG RPP™ test and the comparator test into a 2×2 table. Where a study used multiple comparator tests, we created a 2×2 table for each comparator. Within the statistical analyses, test results from discrepancy resolution (results from a third test when results from the index and comparator tests differed) were not included [22].

We implemented a Bayesian random-effect latent class meta-analysis, which is an extension to the hierarchical summary receiver operating characteristic (HSROC) Model [23], to estimate the sensitivity and specificity of Luminex NxTAG RPP™. This approach takes into account within- and between-study variation as well as accounting for multiple imperfect comparator tests. The model allows us to relax the assumption that, conditional on disease status, tests on the same individual are independent. Inference is done on the estimated mean sensitivity and specificity across studies, i.e. pooled sensitivity/specificity, and the predicted diagnostic accuracy in an out-of-sample study, i.e. predicted sensitivity/specificity. For RSV and influenza separately, we present modelled estimates of the Luminex NxTAG RPP™ test sensitivity and specificity within each study along with a single pooled estimate. By assessing the variability within the studies included in the present meta-analysis we are able to predict the sensitivity and specificity of the Luminex NxTAG RPP™ test if it were applied to a future similar population. We present these predicted estimates of Luminex NxTAG RPP™ for RSV and influenza viruses as summary ROC curves, plotting the 95% credible region. The meta-analyses were implemented using Stan in R [24]. A full model specification including the choice of prior distributions and sensitivity analyses can be found in Additional file 1: Annex S3.

We fit separate meta-analyses for RSV and influenza. Within the influenza model we explored heterogeneity between influenza A and influenza B viruses and present pooled estimates by influenza type.

Assessment of the certainty of the evidence

We assessed the certainty of the evidence using GRADE and GRADEpro GDT software [25–27]. We rated certainty as high, moderate, low, or very low across four domains (risk of bias, indirectness, inconsistency and imprecision). We assessed risk of bias and indirectness by using the QUADAS-2 tool [21]. We explored inconsistency by investigating potential sources of heterogeneity. For imprecision, we considered the width of the Bayesian credible intervals (CrI). We calculated I+/C+, I+/C−, I−/C+ and I−/C−, with ranges for these values based on the CrI of the predicted estimates of sensitivity and specificity for prevalences of 5% and 20% of RSV or influenza viruses, and we made judgements on imprecision using these calculations.

The protocol, developed prior to conducting the review, is accessible online (Prospero CRD42021272062) [28].

Results

We identified 610 potentially eligible studies (Additional file 1: Annex S4). Of these, ten met our selection criteria and were included in the review and meta-analysis (Additional file 1: Fig. S1).

Study description

The ten studies included are described in Additional file 1: Tables S1–S10 and their key findings in Table 1.

The studies included data from 4329 samples. Samples were collected from children and adults in three studies [29–31], children only in two [32, 33], adults only in one [34], and age was not specified in the remaining studies [35–38]. For seven studies, participants were recruited if they presented with symptoms suggestive of acute lower respiratory infection; this was not clearly stated for the remaining three studies [29, 36, 37]. No study specified fever as an inclusion criterion, nor reported the proportion of participants with fever. Patients were selected from a paediatric intensive care unit in one study [33]. In another, samples were collected from mostly immunocompromised patients with underlying chronic lung conditions [34]. Luminex NxTAG RPP™ was performed on 2132 nasopharyngeal washings [31], 1194

Table 1 Summary of characteristics of the studies included in this review

Study ID	Setting	Participants	Samples	n	Comparator tests
Beckmann 2016	Switzerland	Children and adults Symptoms and fever not reported	NPS (199), BAL (76), others (7)	282	RespiFinder-221[a]
Brotons 2016	Spain	< 18 years with ALRI Fever not reported	NPA	320	Anyplex II RV16 assay[a,b,c]
Chan 2017	China	Children and adults with ARI Fever not reported	NPA	133	RT-PCR **AND** DFA[a]
Chen 2016	China	Patients with ARI Fever not reported	NPS	284	FilmArray RP[a,d,e]
Esposito 2016	Italy	Children with ARI in PICU and children with pneumonia by *M. pneumoniae*	NPS	185	xTAG RVP FAST v2[a,c] RT-PCR
Gonsalves 2019	USA, Canada	Children and adults with ARI Fever not reported	NPW	2132	xTAG RVP[a,c] **OR** bidirectional sequencing
Lee 2017	Singapore	Not reported	Respiratory samples	142	xTAG RVP FAST v2[a,c]
Locher 2019	Canada	Adults, mostly immunocompromised and with underlying chronic lung conditions, with ARI	Bronchoscopy collected samples	133	FilmArray RP[a,d,e]
Sails 2017	United Kingdom	"Symptomatic" Other characteristics not reported	NPS (122), throat swabs (53), endotracheal (47), BAL (17), others (122)	314	In-house multiplex RT-PCR panel
Tang 2016	USA	Patients with respiratory symptoms Fever not reported	NPS	404	FilmArray RP[a,d,e]

ALRI acute lower respiratory infection, *ARI* acute respiratory infection, *BAL* broncho-alveolar lavage, *DFA* direct immunofluorescence, *n* sample size, *NPA* nasopharyngeal aspirate, *NPS* nasopharyngeal swabs, *RP* respiratory panel; *RVP* respiratory virus panel

[a] Complies with CE-IVD regulations

[b] Complies with Canadian Department of Health regulations

[c] Complies with Korea Food and Drug Administration regulations

[d] Complies with the United States Food and Drug Administration regulations

[e] Complies with Therapeutic Goods Administration regulations

nasopharyngeal swabs [29, 33, 35, 37, 38], 453 nasopharyngeal aspirates [30, 32] and other respiratory samples [29, 34, 36, 37].

The studies compared Luminex NxTAG RPP™ with various comparator tests including BioFire FilmArray Respiratory Panel [34, 35, 38], xTAG Respiratory Virus Panel Fast Assay v2 [33, 36], xTAG Respiratory Virus Panel [31], RespiFinder-221 [29], Anyplex II RV16 assay [32], RT-PCR [33, 37], and bidirectional sequencing [31]. In one study, the comparator consisted of the combination of RT-PCR and direct immunofluorescence [30]. Another used either xTAG Respiratory Virus Panel or RT-PCR as comparator, without providing disaggregated data [31].

Methodological quality of included studies

See Additional file 1: Tables S1–S10 for the assessment of the methodological quality of each study included. Figure 1 summarizes the risk of bias and applicability concerns, describing our judgements about each domain for each included study.

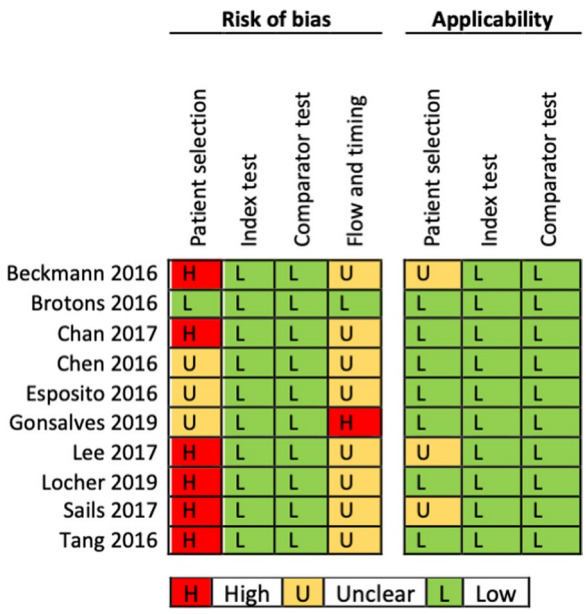


Fig. 1 Risk of bias and applicability concerns in the accuracy of Luminex NxTAGG RPP™ for the detection of RSV and influenza viruses in respiratory samples

In the patient selection domain, we judged six studies at high risk of bias, because recruitment of participants was not consecutive or random but planned after the test was performed on patients selected by physician (e.g. respiratory symptoms) with no clear inclusion and exclusion criteria [29, 30, 34, 36–38]. Regarding applicability, we rated three studies as 'unclear' as inclusion criteria were not recorded by the investigators [29, 36, 37] and considered the remaining seven studies to match the review question.

In the index test domain, we considered all ten studies at low risk of bias, because Luminex NxTAG RPP™ was interpreted without the knowledge of the results of the comparator, and because we judged that knowing the result of the comparator was at very low risk of introducing bias due to the test characteristics.

In the comparator test domain, we judged all the studies to be at low risk of bias because knowing the finding of the index test is at low risk of introducing bias in the interpretation of the comparator tests, due to their intrinsic characteristics.

In the flow and timing domain, we considered eight studies at unclear risk of bias because samples were stored for a long or unclear duration between the performance of the index and comparator tests [29, 30, 33–38]. We considered one study at high risk of bias because investigators did not use the same comparator for all the samples [31], and one study at low risk of bias because index and comparator tests were performed on the same sample collected prospectively [32].

Findings

Luminex NxTAG RPP™ for detection of RSV

The ten studies reported findings of Luminex NxTAG RPP™ and at least one other RPP for detecting RSV (Additional file 1: Table S11). Two studies reported data separately for RSV-A and RSV-B such that it was not possible to pool data for RSV-A or RSV-B, as possible co-infection was not reported [30, 32]. Six studies provided disaggregated findings for RSV-A and RSV-B (Additional file 1: Table S12) [29–33, 37].

The studies included had estimated mean sensitivities ranging from 99 to 100% and specificities of 100% (Additional file 1: Table S11, Fig. 2).

The pooled sensitivity of Luminex NxTAG RPP™ was 100% (95% CrI 99–100) and pooled specificity was 100% (95% CrI 99–100). Predicted sensitivity was 99% (95% CrI 96–100, 8 studies, 527 samples; low certainty evidence) and predicted specificity was 100% (95% CrI 98–100, 8 studies, 5601 samples; low certainty evidence) (Table 2; Fig. 3).

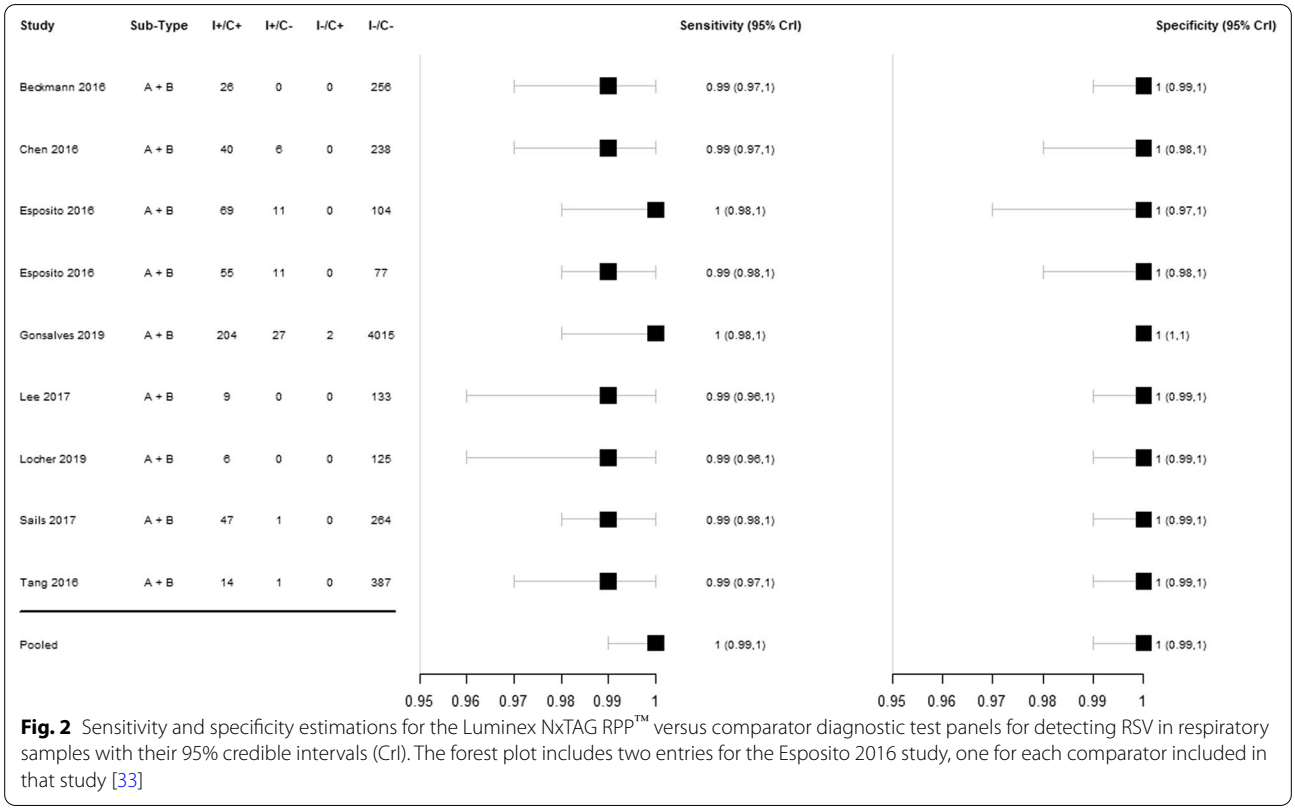


Fig. 2 Sensitivity and specificity estimations for the Luminex NxTAG RPP™ versus comparator diagnostic test panels for detecting RSV in respiratory samples with their 95% credible intervals (CrI). The forest plot includes two entries for the Esposito 2016 study, one for each comparator included in that study [33]

Table 2 Summary of findings for diagnostic accuracy of Luminex NxTAG RPP™ for the diagnosis of RSV

Outcome	Effect per 1000 patients tested		№ of studies (№ of samples)	Test accuracy certainty of evidence
	Pre-test probability of 5%	Pre-test probability of 20%		
Index and comparator tests positive (I+/C+) (patients with RSV infection)	50 (48 to 50)	198 (192 to 200)	8 studies 527 samples	⊕⊕⊖⊖[a] LOW
Index test negative, comparator positive (I−/C+) (patients incorrectly classified as not having RSV infection)	0 (0 to 2)	2 (0 to 8)		
Index and comparator tests negative (I−/C−) (patients without RSV infection)	950 (931 to 950)	800 (784 to 800)	8 studies 5601 samples	⊕⊕⊖⊖[a] LOW
Index test positive, comparator negative (I+/C−) (patients incorrectly classified as having RSV infection)	0 (0 to 19)	0 (0 to 16)		

C comparator, *CrI* credible interval, *I* index test, *RPP* respiratory pathogen panel

Patient or population: adults and children with symptoms of acute lower respiratory infection

Setting: worldwide

Index test: Luminex NxTAG RPP™

Comparator tests: other RPP

Predicted sensitivity: 0.99 (95% CrI: 0.96 to 1.00) | Predicted specificity: 1.00 (95% CrI: 0.98 to 1.00)

[a] Downgraded two levels for risk of bias: there is high or unclear risk of bias on the patient selection and flow and timing domains for all included studies. Six studies were planned after the test was performed on patients selected by physician (e.g. respiratory symptoms) with no clear inclusion and exclusion criteria, which is at high risk of introducing bias for evaluating diagnostic test accuracy.

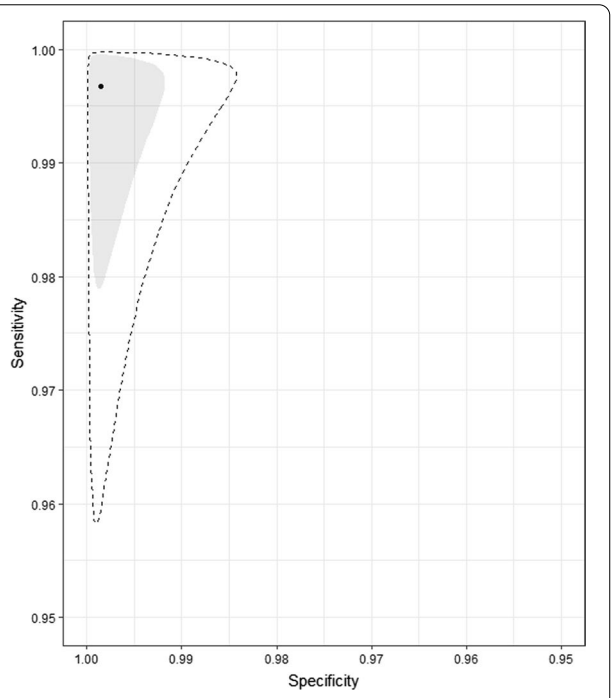


Fig. 3 Pooled (shaded) and predicted (dashed) credible regions of Luminex NxTAG RPP™ for detecting RSV in respiratory samples

The performance of Luminex NxTAG RPP™ appeared similar for RSV-A and RSV-B (Additional file 1: Fig. S2 and S3) [30, 37].

Luminex NxTAG RPP™ for detection of influenza viruses

All except one study [33] reported findings of Luminex NxTAG RPP™ and at least one other RPP for detecting influenza A and influenza B (Additional file 1: Table S13). Seven studies presented disaggregated data for influenza subtypes AH1 and AH3 [29–31, 35–38]. For three of these, we inferred the 2×2 table for influenza A from the subtypes data assuming no co-infection [29, 37, 38].

For detection of influenza A virus, mean sensitivity estimates were 96–98% and mean specificity estimates were 100% in all studies (Additional file 1: Table S13, Fig. 4).

Pooled sensitivity of Luminex NxTAG RPP™ against comparator tests was 98% (95% CrI 95–100) and pooled specificity was 100% (95% CrI 99–100). Predicted sensitivity was 97% (95% CrI 89–100, 9 studies, 460 samples; low certainty evidence) and predicted specificity was 100% (95% CrI 99–100, 9 studies, 3677 samples; low certainty evidence) (Table 3; Fig. 5).

For detection of influenza B virus, mean sensitivity estimates ranged from 97 to 98% and mean specificity was 100% in all studies [31] (Table 4, Fig. 4). Pooled sensitivity was 98% (95% CrI 95–100) and pooled specificity was 100% (95% CrI 100–100). Predicted sensitivity was 98% (95% CrI 88–100, 9 studies, 164 participants; low certainty evidence) and predicted specificity was 100% (95% CrI 99–100, 9 studies, 3965 participants; low certainty evidence) (Table 4; Fig. 5).

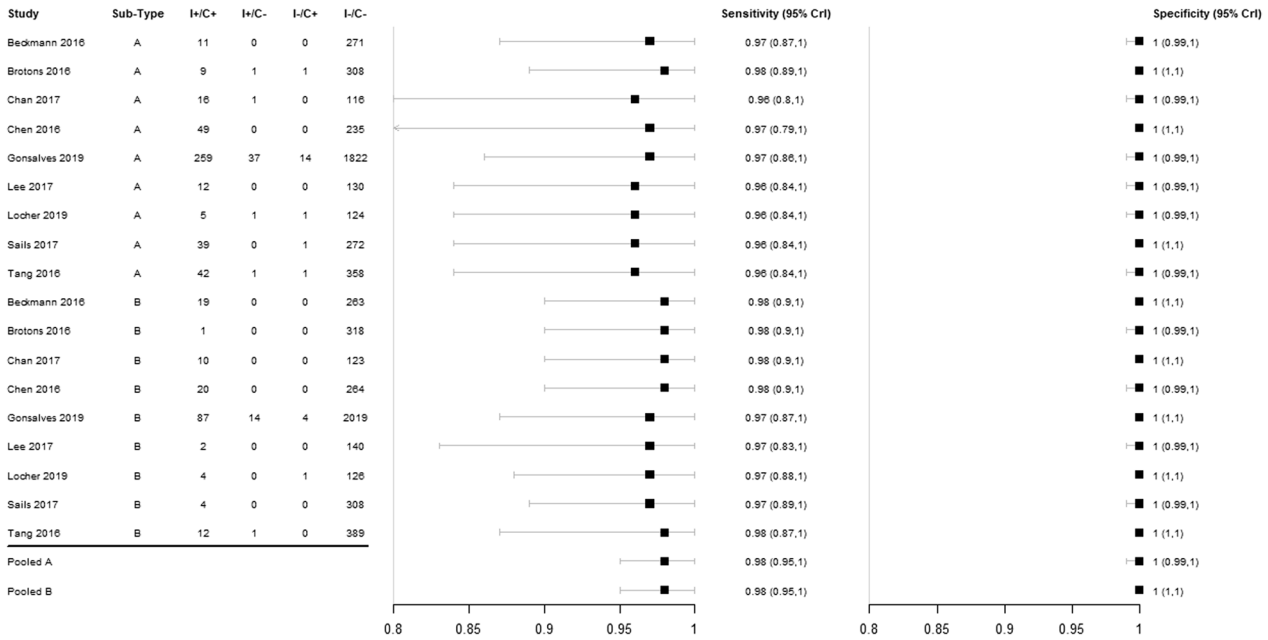


Fig. 4 Forest plot of the sensitivity and specificity of Luminex NxTAG RPP™ versus comparator RPP for detecting influenza virus in respiratory samples, with their 95% credible intervals (CrI)

Table 3 Summary of findings for diagnostic accuracy of Luminex NxTAG RPP™ for the diagnosis of influenza A virus

Outcome	Effect per 1000 patients tested		Nº of studies (Nº of samples)	Test accuracy certainty of evidence
	Pre-test probability of 5%	Pre-test probability of 20%		
Index and comparator tests positive (I+/C+) (patients with influenza A infection)	49 (45 to 50)	194 (178 to 200)	9 studies 460 samples	⊕⊕⊖⊖[a] LOW
Index test negative, comparator positive (I−/C+) (patients incorrectly classified as not having influenza A infection)	1 (0 to 5)	6 (0 to 22)		
Index and comparator tests negative (I−/C−) (patients without influenza A infection)	950 (941 to 950)	800 (792 to 800)	9 studies 3677 samples	⊕⊕⊖⊖[a] LOW
Index test positive, comparator negative (I+/C−) (patients incorrectly classified as having influenza A infection)	0 (0 to 9)	0 (0 to 8)		

C comparator, *CrI* credible interval, *I* index test, *RPP* respiratory pathogen panel

Patient or population: adults and children with symptoms of acute lower respiratory infection

Setting: worldwide

Index test: Luminex NxTAG RPP™

Comparator tests: other RPP

Predicted sensitivity: 0.97 (95% CrI: 0.89 to 1.00) | Predicted specificity: 1.00 (95% CrI: 0.99 to 1.00)

[a] Downgraded two levels for risk of bias: there is high or unclear risk of bias on the patient selection and flow and timing domains for all included studies. Six studies were planned after the test was performed on patients selected by physician (e.g. respiratory symptoms) with no clear inclusion and exclusion criteria, which is at high risk of introducing bias for evaluating diagnostic test accuracy

Although seven studies provided data for influenza subtypes (Additional file 1: Table S13), we did not perform subgroup meta-analysis at this level due to the scarcity of data.

Discussion

In this systematic review and meta-analysis of ten studies, including results from 4329 patient samples, we found that Luminex NxTAG RPP™ had a predicted

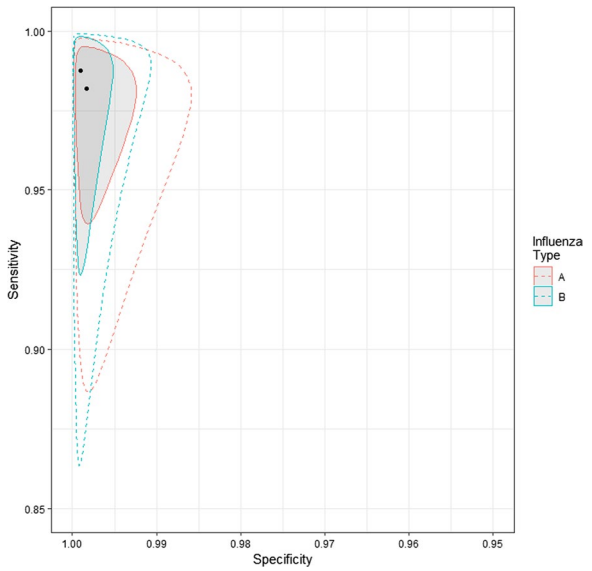


Fig. 5 Pooled (shaded) and predicted (dashed) credible regions of Luminex NxTAG RPP™ for detecting influenza viruses in respiratory samples

mean sensitivity and specificity of 99% and 100% for detecting RSV, 97% and 100% for influenza A, and 98% and 100% for influenza B. If Luminex NxTAG RPP™ were used in a hypothetical population of 1000 persons with acute lower respiratory symptoms where 50 actually were infected with RSV-A or RSV-B (pre-test probability of 5%), we estimated that the test would correctly detect RSV in 50 people (50 I+/C+, 95% CrI 48–50), would not miss any infection (0 I−/C+, 95% CrI 0–2), and would not detect RSV in people in discordance with the comparator tests (0 I+/C-,−95% CrI 0–19) (Table 2). Similar results were seen with influenza A and B: with a pre-test probability of 5%, we would anticipate one I−/C+ case and no I+/C− case (Tables 3 and 4).

However, these results must be treated with caution. We found a high risk of bias in most studies, particularly as regards patient selection, and a lack of clarity in many studies as to sample flow and timing. In several studies Luminex NxTAG RPP™ was performed on stored frozen respiratory samples with unclear storage duration. While the data generated by these studies is important for assay validation, it is more complex to generalise their results to other patient populations.

Respiratory pathogens including RSV and influenza viruses are often in the differential diagnosis for patients presenting with febrile illness. Consequently, RPP may be used clinically for diagnostic testing in undifferentiated fever cases. We did not find any studies matching our specific initial inclusion criterion of febrile patients. It may well be that Luminex NxTAG RPP™ performs equally well in patients with undifferentiated fever, but wider evaluation with prospective recruitment and clear inclusion criteria (symptomatic with fever and/or respiratory symptoms) should be conducted.

Table 4 Summary of findings for diagnostic accuracy of Luminex NxTAG RPP™ for the diagnosis of influenza B virus

Outcome	Effect per 1000 patients tested		№ of studies (№ of samples)	Test accuracy certainty of evidence
	Pre-test probability of 5%	Pre-test probability of 20%		
Index and comparator tests positive (I+/C+) (patients with influenza B infection)	49 (45 to 50)	194 (178 to 200)	9 studies 164 samples	⊕⊕⊖⊖[a] LOW
Index test negative, comparator positive (I−/C+) (patients incorrectly classified as not having influenza B infection)	1 (0 to 5)	6 (0 to 22)		
Index and comparator tests negative (I−/C−) (patients without influenza B infection)	950 (941 to 950)	800 (792 to 800)	9 studies 3965 samples	⊕⊕⊖⊖[a] LOW
Index test positive, comparator negative (I+/C−) (patients incorrectly classified as having influenza B infection)	0 (0 to 9)	0 (0 to 8)		

C Comparator, *CrI* Credible intervall, *I* Index test, *RPP* Respiratory pathogen panel

Patient or population: adults and children with symptoms of acute lower respiratory infection

Setting: worldwide

Index test: Luminex NxTAG RPP™

Comparator tests: other RPP

Predicted sensitivity: 0.98 (95% CrI: 0.88 to 1.00) | Predicted specificity: 1.00 (95% CrI: 0.99 to 1.00)

aDowngraded two levels for risk of bias: there is high or unclear risk of bias on the patient selection and flow and timing domains for all included studies. Six studies were planed after the test was performed on patients selected by physician (e.g. respiratory symptoms) with no clear inclusion and exclusion criteria, which is at high risk of introducing bias for evaluating diagnostic test accuracy

The uncertainty in this review is compounded by the wide range of comparator tests used. Alternative reference tests to RT-PCR include culture (the classic gold standard but time consuming and laborious); direct fluorescent antibody testing (requiring technical expertise and potentially subjective); serology (in general too slow to be of acute clinical relevance); and rapid immunoassays such as lateral flow tests, which may lack sensitivity [39]. Furthermore, in-house RT-PCRs are all likely to be unique in the first place, with different probe combinations and thus varying sensitivity and specificity. The wide range of reference tests is not isolated to our review—a previous meta-analysis of multiplex PCRs for diagnosis of respiratory infections showed similar findings [13]. This apparent lack of a single 'gold standard' may be explained by a reliance on national regulatory bodies to rigorously assess commercial tests to ensure quality and performance (e.g. Food and Drug Administration approval in the United States, CE-IVD marking in Europe) as opposed to large scale clinical studies evaluating each test against a 'gold standard'. Indeed, under changing IVD regulation in Europe, laboratories are likely to need to justify the use of in-house tests over and above those that are commercially available. This lack of a gold standard might appear concerning, but with regulatory bodies ensuring baseline performance conformity, the broad range of test kits available means laboratories have the freedom to choose test kits that fit best with local demographics and individual laboratory logistics. What then becomes most important is ongoing quality assurance, in particular external quality assurance such as inter-laboratory exchange schemes.

In terms of the limitations of our review, we set out to review the diagnostic accuracy of Luminex NxTAG RPP™ for detecting RSV, influenza A and influenza B in febrile patients, to match FIEBRE study objectives [17], but we found no studies including participants enrolled on the basis of fever. We therefore expanded our review to include any study where clinical samples were evaluated with both Luminex NxTAG RPP™ and another assay, with obvious consequences in the applicability of our findings to patients with the common syndrome of febrile illness. Strengths of this review include a comprehensive literature search and a robust methodology with independent duplicate review and adherence to QUADAS-2 and GRADE methodology, and PRISMA guidelines. Furthermore, by using an extension to the HSROC model we have not assumed that any one test is a gold standard, but that all tests are imperfect measures of an underlying not directly observable (true disease) status or class [23]. This statistical method lends itself well to analysing the multiple comparator tests used in studies identified for this review and the inherent heterogeneity this brings as well as mitigating against the lack of a true gold standard reference test in this context.

Conclusion

We found excellent sensitivity and specificity for the Luminex NxTAG RPP™ assay for RSV and influenza A and B, but within studies that were either limited to patients with respiratory symptoms, or with an unclear participant enrolment strategy. Further research is merited to ascertain whether Luminex NxTAG RPP™ will perform equally well among patients with febrile illness.

Abbreviations

CrI: Credible interval; FIEBRE: Febrile Illness Evaluation in a Broad Range of Endemicities; HSROC: Hierarchical summary receiver operating characteristic; IVD: In vitro diagnostic; QUADAS-2: Quality assessment tool for diagnostic accuracy studies; RPPs: Respiratory pathogen panels; RSV: Respiratory syncytial virus; RT-PCR: Real-time polymerase chain reaction.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-022-07766-9.

Additional file 1: Annex S1. Search methodology. **Annex S2.** Databases. **Annex S3.** Statistical model. **Annex S4.** Literature search results. **Additional tables and figures.**

Acknowledgements

Not applicable.

Author contributions

SJ, FF, HH and QB conceived the study. SJ and FF assessed the eligibility of the studies, extracted the data, and assessed the methodological quality of the included studies. JF developed the search strategy and conducted the literature search. SK carried out the statistical analysis. JB, RK and OB advised on the statistical analysis. SJ and FF prepared the original draft of the manuscript, with considerable input from SK, MV, QB, CF and HH. All authors contributed to the interpretations of results. All authors read and approved the final manuscript.

Funding

ISGlobal receives support from the Spanish Ministry of Science and Innovation through the "Centro de Excelencia Severo Ochoa 2019–2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. CISM is supported by the Government of Mozambique and the Spanish Agency for International Development (AECID). Dr F. Fitzgerald is supported by the Academy of Medical Sciences and the funders of the Starter Grants for Clinical Lecturers scheme. Dr F. Fitzgerald is further supported by the National Institute for Health Research Great Ormond Street Hospital Biomedical Research Centre. Dr H Hopkins' salary is covered through the FIEBRE study, which is funded by UK aid from the UK government; the views expressed, however, do not necessarily reflect the UK government's official policies. S. Keddie is supported by the Medical Research Council London Intercollegiate Doctoral Training Partnership Studentship (MR/N013638/1). The funders had no role in study design, data collection and analysis, or preparation of this report.

Availability of data and materials

Code used is publicly available at: https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/RSV_Influenza. Data included in meta-analyses can be found in Additional file 1: Tables S11 and S12.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Barcelona Institute for Global Health (ISGlobal), Hospital Clínic - Universitat de Barcelona, Carrer Rosselló 132, 08036 Barcelona, Spain. [2]Imperial College London, London, UK. [3]London School of Hygiene & Tropical Medicine, London, UK. [4]Centro de Investigação Em Saúde de Manhiça (CISM), Maputo, Mozambique. [5]ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain. [6]Pediatrics Department, Hospital Sant Joan de Déu, Universitat de Barcelona, Esplugues, Barcelona, Spain. [7]Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [8]Micropathology Ltd., University of Warwick Science Park, Coventry, UK. [9]Department of Life Sciences, University of Warwick, Coventry, UK.

Received: 13 May 2022 Accepted: 30 September 2022
Published online: 13 October 2022

References

1. Shi T, McAllister D, O'Brien K, Simoes E, Madhi S, Gessner B, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet. 2017;390(10098):946–58.
2. Srikantiah P, Vora P, Klugman KP. Assessing the full burden of respiratory syncytial virus in young infants in low- and middle-income countries: The importance of community mortality studies. Clin Infect Dis. 2021;73(S3):177–9.
3. Mazur N, Löwensteyn Y, Willemsen J, Gill C, Forman L, Mwananyanda L, et al. Global respiratory syncytial virus-related infant community deaths. Clin Infect Dis. 2021;73(Suppl 3):S229–37.
4. Iuliano A, Roguski K, Chang H, Muscatello D, Palekar R, Tempia S, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet. 2018;391(10127):1285–300.
5. Wang X, Li Y, O'Brien K, Madhi S, Widdowson M, Byass P, et al. Global burden of respiratory infections associated with seasonal influenza in children under 5 years in 2018: a systematic review and modelling study. Lancet Glob Heal. 2020;8(4):e497-510.
6. Blau D, Baillie V, Els T, Mahtab S, Mutevedzi P, Keita A, et al. Deaths attributed to respiratory syncytial virus in young children in high-mortality rate settings: Report from Child Health and Mortality Prevention Surveillance (CHAMPS). Clin Infect Dis. 2021;73(Suppl 3):S218–28.
7. Smithgall M, Maykowski P, Zachariah P, Oberhardt M, Vargas CY, Reed C, et al. Epidemiology, clinical features, and resource utilization associated with respiratory syncytial virus in the community and hospital. Influenza Other Respi Viruses. 2020;14(3):247–56.
8. Forum of International Respiratory Societies. The global impact of respiratory disease – second edition. Sheffi eld, European Respiratory Society, 2017.
9. Chavada R, Clifford L, Weisback O. Clinical impact of rapid diagnostics using Xpert Flu/RSV™ PCR on antimicrobial stewardship initiatives during influenza season. Infect Dis Heal. 2020;25(2):71–6.
10. Tillekeratne LG, Bodinayake CK, Nagahawatte A, Vidanagama D, Devasiri V, Arachchi WK, et al. Use of rapid influenza testing to reduce antibiotic prescriptions among outpatients with influenza-like illness in southern Sri Lanka. Am J Trop Med Hyg. 2015;93(5):1031–7.
11. O'Brien KL, Levine OS, Deloria Knoll M, Feikin DR, DeLuca AN, Driscoll AJ, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. Lancet. 2019;394:757–79.
12. Fowlkes AL, Steffens A, Reed C, Temte JL, Campbell AP, Rubino H. Influenza antiviral prescribing practices and the influence of rapid testing among primary care providers in the US, 2009–2016. Open Forum Infect Dis. 2019;6(6):2009–16.
13. Huang HS, Tsai CL, Chang J, Hsu TC, Lin S, Lee CC. Multiplex PCR system for the rapid diagnosis of respiratory virus infection: systematic review and meta-analysis. Clin Microbiol Infect. 2018;24(10):1055–63.
14. The United States Food and Drug Administration. Center for Devices and Radiological Health (CDRH) Proposed guidance development [Internet]. 2021. https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/cdrh-proposed-guidance-development#proposed. Accessed 21 Dec 2021.
15. EUR-Lex. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnotic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. 2017.
16. Luminex. NxTAG® Respiratory Pathogen Panel [Internet]. 2021. https://www.luminexcorp.com/nxtag-respiratory-pathogen-panel/. Accessed 21 Dec 2021.
17. Hopkins H, Bassat Q, Chandler CI, Crump JA, Feasey NA, Ferrand RA, et al. Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE): protocol for a multisite prospective observational study of the causes of fever in Africa and Asia. BMJ Open. 2020;10(7): e035632.
18. Micropathology Ltd. Verification of the Luminex NxTAG® Respiratory Pathogen Panel (RPP) and Luminex MAGPIX® analyser for the detection of respiratory pathogens, S-1134. Internal company document. 2017.
19. Falconer J. Removing duplicates from an EndNote Library. Library & Archives Service Blog [Internet]. London School of Hygiene & Tropical Medicine. 2018 [cited 2021 Oct 4]. Available from: https://blogs.lshtm.ac.uk/library/2018/12/07/removing-duplicates-from-an-endnote-library/
20. Kohl C, McIntosh E, Unger S, Haddaway N, Kecke S, Schiemann J, et al. Oyynline tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Env Evid. 2018;7:8. https://doi.org/10.1186/s13750-018-0115-5.
21. Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
22. Miller WC. Bias in discrepant analysis: When two wrongs don't make a right. J Clin Epidemiol. 1998;51(3):219–31.
23. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics. 2012;68(4):1285–93.
24. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.28 [Internet]. 2019. https://mc-stan.org. Accesse 12 Dec 2021.
25. Schünemann H, Mustafa R, Brozek J, Steingart K, Leeflang M, Murad M. GRADE guidelines: 21 part 2. Inconsistency, Imprecision, publication bias and other domains for rating the certainty of evidence for test accuracy and presenting it in evidence profiles and summary of findings tables. J Clin Epidemiol. 2020;122:142–52.
26. McMaster University (developed by Evidence Prime). GRADEpro GDT. 2015. gradepro.org.
27. Schünemann H, Brożek J, Guyatt G, Oxman A, editor(s). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach [Internet]. GRADE Working Group. 2013. gdt.guidelinedevelopment.org/app/handbook/handbook.html. Accessed 26 Sep 2021.
28. Jullien S, Fitzgerald F, Hopkins H, Bassat Q, Voice M, Fink C, et al. Luminex NxTAG Respiratory Pathogen Panel™ in respiratory samples for the diagnosis of infections with influenza or respiratory syncytial virus compared with standard real-time polymerase chain reactions, a systematic review [Internet]. PROSPERO CRD42021272062. 2021. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021272062. Accessed 20 Oct 2021.
29. Beckmann C, Hirsch HH. Comparing luminex NxTAG-respiratory pathogen panel and respifinder-22 for multiplex detection of respiratory pathogens. J Med Virol. 2016;88(8):1319–24.
30. Chan KH, To KKW, Li PTW, Wong TL, Zhang R, Chik KKH, et al. Evaluation of NxTAG respiratory pathogen panel and comparison with xTAG respiratory viral panel fast v2 and film array respiratory panel for detecting

respiratory pathogens in nasopharyngeal aspirates and swine/avian-origin influenza a subtypes in culture. Adv Virol. 2017;2017:1324276.
31. Gonsalves S, Mahony J, Rao A, Dunbar S, Juretschko S. Multiplexed detection and identification of respiratory pathogens using the NxTAG ® respiratory pathogen panel. Methods. 2019;158:61–8.
32. Brotons P, Henares D, Latorre I, Cepillo A, Launes C, Muñoz-Almagro C. Comparison of NxTAG Respiratory Pathogen Panel and Anyplex II RV16 tests for multiplex detection of respiratory pathogens in hospitalized children. J Clin Microbiol. 2016;54(12):2900–4.
33. Esposito S, Scala A, Bianchini S, Presicce ML, Mori A, Sciarrabba CS, et al. Partial comparison of the NxTAG Respiratory pathogen panel assay with the luminex xTAG respiratory panel fast assay v2 and singleplex real-time polymerase chain reaction for detection of respiratory pathogens. Diagn Microbiol Infect Dis. 2016;86(1):53–7. https://doi.org/10.1016/j.diagmicrobio.2016.06.018.
34. Locher K, Roscoe D, Jassem A, Wong T, Hoang LMN, Charles M, et al. FilmArray respiratory panel assay: an effective method for detecting viral and atypical bacterial pathogens in bronchoscopy specimens. Diagn Microbiol Infect Dis. 2019;95(4):114880. https://doi.org/10.1016/j.diagmicrobio.2019.114880.
35. Chen JHK, Lam HY, Yip CCY, Wong SCY, Chan JFW, Edmond SK, et al. Clinical evaluation of the new high-throughput luminex NxTAG respiratory pathogen panel assay for multiplex respiratory pathogen detection. J Clin Microbiol. 2016;54(7):1820–5.
36. Lee CK, Lee HK, Ng CWS, Chiu L, Tang JWT, Loh TP, et al. Comparison of luminex NxTAG respiratory pathogen panel and xTAG respiratory viral panel FAST version 2 for the detection of respiratory viruses. Ann Lab Med. 2017;37(3):267–71.
37. Sails AD, Eltringham G, Valappil M, Waugh S, Saunders D. Comparison of the Luminex NxTAG respiratory pathogen panel and a multiplex in-house real-time PCR panel for the detection of respiratory viruses in symptomatic patients. J Med Microbiol. 2017;66(9):1291–6.
38. Tang YW, Gonsalves S, Sun JY, Stiles J, Gilhuley KA, Mikhlina A, et al. Clinical evaluation of the Luminex NxTAG Respiratory Pathogen Panel. J Clin Microbiol. 2016;54(7):1912–4.
39. Das S, Dunbar S, Tang YW. Laboratory diagnosis of respiratory tract infections in children—the state of the art. Front Microbiol. 2018;9:1–9.

Publisher's Note

Appendix B

CrAg diagnostic test accuracy meta-analysis

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1702488	**Title**	Miss
First Name(s)	Suzanne		
Surname/Family Name	Keddie		
Thesis Title	Latent class models for diagnostic test accuracy with application to fever aetiology		
Primary Supervisor	John Bradley		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMC Infectious Diseases		
When was the work published?	06/04/2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

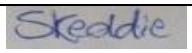
Where is the work intended to be published?	

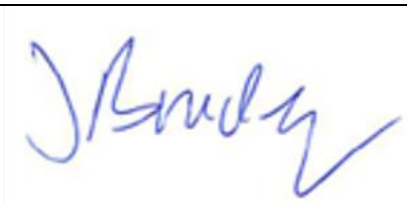
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I helped conceive the study as well as carrying out the statistical analysis and helping prepare the original draft of the manuscript.

SECTION E

Student Signature	
Date	06/09/2023

Supervisor Signature	
Date	25/03/2024

BMC Infectious Diseases

RESEARCH

Open Access

Diagnostic performance of the IMMY cryptococcal antigen lateral flow assay on serum and cerebrospinal fluid for diagnosis of cryptococcosis in HIV-negative patients: a systematic review

Catriona Macrae[1*], Jayne Ellis[2,3], Suzanne H. Keddie[2], Jane Falconer[2], John Bradley[2], Ruth Keogh[2], Oliver Baerenbold[2], Heidi Hopkins[2] and Joseph N. Jarvis[2,4]

Abstract

Background The incidence of cryptococcosis amongst HIV-negative persons is increasing. Whilst the excellent performance of the CrAg testing in people living with HIV is well described, the diagnostic performance of the CrAg LFA has not been systematically evaluated in HIV-negative cohorts on serum or cerebrospinal fluid.

Methods We performed a systematic review to characterise the diagnostic performance of IMMY CrAg® LFA in HIV-negative populations on serum and cerebrospinal fluid. A systematic electronic search was performed using Medline, Embase, Global Health, CENTRAL, WoS Science Citation Index, SCOPUS, Africa-Wide Information, LILACS and WHO Global Health Library. Studies were screened and data extracted from eligible studies by two independent reviewers. A fixed effect meta-analysis was used to estimate the diagnostic sensitivity and specificity.

Results Of 447 records assessed for eligibility, nine studies met our inclusion criteria, including 528 participants overall. Amongst eight studies that evaluated the diagnostic performance of the IMMY CrAg® LFA on serum, the pooled median sensitivity was 96% (95% Credible Interval (CrI) 68–100%) with a pooled specificity estimate of 96% (95%CrI 84–100%). Amongst six studies which evaluated the diagnostic performance of IMMY CrAg® LFA on CSF, the pooled median sensitivity was 99% (95%CrI 95–100%) with a pooled specificity median of 99% (95%CrI 95–100%).

Conclusions This review demonstrates a high pooled sensitivity and specificity for the IMMY CrAg® LFA in HIV-negative populations, in keeping with findings in HIV-positive individuals. The review was limited by the small number of studies. Further studies using IMMY CrAg® LFA in HIV-negative populations would help to better determine the diagnostic value of this test.

Keywords Cryptococcosis, HIV-negative, CrAg, Cryptococcal antigen, Lateral flow assay, Diagnosis, Diagnostic, Diagnostic performance, Serum, CSF

*Correspondence:
Catriona Macrae
catriona.macrae3@nhs.scot
Full list of author information is available at the end of the article

Background

Cryptococcosis is a fungal infection caused by the pathogenic *Cryptococcus* species, of which there are seven recognised species: *C. neoformans* variety *grubii*, *C. neoformans* variety *neoformans* and five species within *C. gatti* [1]. Infection occurs following inhalation of fungal cells which may lead to either asymptomatic colonisation or pulmonary cryptococcal disease [2] presenting with cough, fever, shortness of breath and/or pulmonary nodules on chest radiographs [3]. *Cryptococcus spp.* may disseminate to cause cryptococcal antigenaemia, with or without progression to multi-organ disease. Dissemination to the central nervous system causes cryptococcal meningitis, which typically presents with fever, headache, neck stiffness, altered mental status and visual disturbance [2]. Other body sites such as liver, spleen skin and bone are less commonly affected [3].

Cryptococcal infection most often occurs in people living with HIV (PLWH); however, the proportion of cases in HIV-negative patients is increasing in high income countries [4, 5], in part due to increasing use of immunosuppressive therapies for cancer chemotherapy and organ transplantation [4, 6].

In addition to immunosuppressive therapy or solid organ transplantation, hematopoietic and other malignancies, innate immune defects, advanced renal or liver disease, diabetes mellitus, rheumatologic diseases and sarcoidosis increase the risk of cryptococcal infection [3, 7, 8]. Clinical cases of cryptococcal disease have also been reported in apparently immunocompetent individuals [2, 3, 9].

Cryptococcal antigen (CrAg) is a biomarker of cryptococcosis, and detection of CrAg in cerebrospinal fluid (CSF), serum, plasma or whole blood either by lateral flow assay (LFA), latex agglutination (LA) or enzyme-linked immunosorbent assays (ELISA) is the cornerstone in diagnosing cryptococcosis. Other diagnostic modalities include basic CSF analysis (white cell count, protein, glucose), India ink staining, cryptococcal culture on Sabouraud's dextrose agar, and histology. Multiplex polymerase chain (PCR) platforms including *Cryptococcus spp.* as a target pathogen have also been evaluated as a diagnostic tool for cryptococcosis; and matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (MALDI-TOF) has also been reported to detect *Cryptococcus spp.* in clinical specimens [10]. The World Health Organization (WHO) recommends rapid Ag-detection assays for diagnosis of cryptococcal disease in PLWH [11].

The IMMY CrAg® LFA (Norman, Oklahoma, USA), approved by the U.S. Food and Drug Administration (FDA) in 2011, is an immunochromatographic dipstick assay that detects antigen with qualitative or semiquantitative results. The IMMY CrAg® LFA is currently the most sensitive commercially available cryptococcal diagnostic test, with superior sensitivity to India ink microscopy on CSF, CSF cryptococcal culture, Meridian Cryptococcal Antigen Latex Agglutination System (CALAS®), the Meridian EIA assay, and the BioFire® FilmArray® Meningitis/Encephalitis (ME) panel [12–16]. The IMMY CrAg® LFA was therefore employed as part of The Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE) study; a prospective observational study to investigate the infectious causes of fever at four sites in Africa and Asia, collecting data and samples from PLWH and HIV-negative inpatients and outpatients [17]. FIEBRE focused on illnesses deemed preventable or treatable, of which cryptococcosis is an important example. Lumbar punctures were not routinely conducted as part of the FIEBRE diagnostic package, so the IMMY CrAg® LFA performed on serum samples was chosen as the diagnostic strategy for all FIEBRE participants.

The performance of CrAg testing for the diagnosis of cryptococcosis in HIV-negative populations has not previously been systematically reviewed. This review aims to assess diagnostic performance of the IMMY CrAg® LFA compared to other cryptococcal diagnostic tests for the diagnosis of cryptococcosis in HIV-negative persons.

Methods

This systematic review was registered at PROSPERO (www.crd.york.ac.uk/PROSPERO) as CRD42022314040 on 02/03/2022 and is reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for the reporting of systematic reviews and meta-analyses [18].

Literature search strategy

The following searches were conducted with an aim of identifying all studies reporting on the diagnostic performance of the IMMY CrAg® LFA for the diagnosis of cryptococcosis in HIV-negative populations. The study population was HIV-negative adults and children. The index test was the IMMY CrAg® LFA and comparator tests were any alternative cryptococcal diagnostic test/s, including clinical composite end-points.

A systematic electronic search was conducted using Medline, Embase, Global Health, CENTRAL, WoS Science Citation Index, SCOPUS, Africa-Wide Information, LILACS and WHO Global Health Library. A draft search strategy was compiled in the OvidSP Medline database by an experienced information specialist (JF). The search strategy included strings of terms, synonyms and

controlled vocabulary terms (where available) to reflect two concepts: *Cryptococcus spp.* and IMMY lateral flow assay. Further information on the search methodology is available in Additional file 1.

Information management

All citations identified were imported into EndNote™ X9 software (Pennsylvania, PA, USA). Duplicates were identified and removed using the method described on the London School of Hygiene & Tropical Medicine Library & Archives Service blog [19].

The OvidSP MEDLINE search was adapted for each of the bibliographic databases. The search period was 2009–July 2021, as the IMMY CrAg® LFA was introduced in 2009.

Study selection

A two-stage screening process was employed: (1) at title and abstract and (2) at full-text level according to eligibility criteria as detailed below. Screening was performed in duplicate independently by two reviewers (CM, JE), and any disagreements were resolved by discussion. Reports not meeting the eligibility criteria were excluded. Reference and citation checking were conducted for included articles.

Studies were eligible for inclusion if they reported on the use of the IMMY CrAg® LFA tested on serum and/or CSF, in HIV-negative persons, compared to any other test/s or composite used to diagnose cryptococcal disease. Studies including asymptomatic and/or symptomatic persons were included. We included all study types, irrespective of country, region, continent, or level of care (primary, secondary, or tertiary). Studies that did not have disaggregate data for HIV-negative participants were excluded.

Study selection criteria

1) Paper written in English,
2) Studies from 2009 onwards,
3) Study reports use of IMMY CrAg® LFA on serum or CSF,
4) Samples tested are from HIV-negative persons (adults andchildren) only, or if HIV-positive persons included disaggregated data ispresented,
5) Paired data: The same samples tested with IMMY CrAg® LFA compared to any referencestandard,
6) Not case study or case report (and participants testedn>5),
7) Full peer-reviewed published text available.

Data extraction and synthesis

For all eligible studies two reviewers (CM, JE) independently extracted data using an Excel spreadsheet (Microsoft Corp., Redmond, WA, USA) including sample size, study design, participant characteristics, sample characteristics, flow and timing of sample analysis and comparator test characteristics. For each study the performance results for the CrAg LFA test (Index, "I") and the comparator test ("C") were extracted into 2×2 tables. In studies using multiple comparator tests a 2×2 table was generated for each comparator.

Quality assessment

Two reviewers (CM, JE) used the QUADAS-2 (quality assessment of diagnostic-accuracy studies-2) tool for quality assessment to evaluate the risk of bias and applicability of all included studies [20]. Disagreements were resolved by discussion.

Statistical analysis and data synthesis

The original analysis plan included random effect meta-analysis. A random-effect meta-analysis accounting for between-study heterogeneity would usually be the model of choice in this scenario because we would not expect the sensitivity and specificity of the diagnostic test to be the same in each study. In this systematic review however, there were a limited number of studies and only one study that actively sought to investigate the specificity [21] of the diagnostic test of interest, making a random-effects model inappropriate [22]. A fixed-effect model was therefore used in this instance.

Modelled estimates for the sensitivity and specificity of the IMMY CrAg® LFA were calculated for each study, as well as a single pooled estimate. The studies were subdivided by sample type; the estimates were calculated for use of the CrAg LFA on both serum and CSF. For the analysis similar comparator tests, for example, different latex agglutination tests, were grouped to represent a single comparator test.

We chose to fit a Hierarchical Summary Receiver Operating Characteristic (HSROC) model [23] with fixed accuracy and threshold parameters. This model still ensures that sensitivity and specificity are jointly estimated as well as accounting for imperfect reference tests [24] while also allowing for asymmetry in the SROC curve. This model can be seen as a simplification of the random effects model fit in Jullien et al. [25] where the variances of the random effects are zero ($i.e. \sigma_\theta = 0$, and $\sigma_\alpha = 0$, such that θ_j and α_j are equal to Θ and Λ, respectively).

All analyses were conducted in R with stan [26]. For model code see: https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/CrAg.

Results

Our searches yielded 447 potentially eligible articles. After removal of duplicates (n = 12), screening of titles and abstracts (n = 435) and review of the full texts (n = 41), nine articles met our eligibility criteria for inclusion (Fig. 1 PRISMA diagram).

Study description

The nine articles included and their key characteristics are summarised in Tables 1 and 2. In total, the included studies evaluated the diagnostic performance of the IMMY CrAg® LFA in 528 HIV-negative persons, across three continents. The reports were published between 2015 and 2021. The mean number of participants per study was 59, with a median of 37 participants per study. The age of study participants ranged from eight [27] to 88 years. [28, 29] The majority of participants were male, with the percentage of female participants ranging from 27 to 50% [28, 30]. All studies used cross sectional study design.

In seven of the nine included studies, a proportion (12–55%) of the participants were reported to be immunosuppressed. Where documented, immunosuppression included long-term immunosuppressive therapy (1–20%) [21, 27, 28, 31], solid organ transplant (3–19%) [21, 27, 28, 31], malignancy (3–11%) [21, 28, 29, 31–33], innate immune defects (3–39%) [27, 32], liver disease (1–19%) [21, 27, 32], renal disease (3%) [32], diabetes (3–14%) [27–29, 32, 33] and rheumatological disease (1–14%) [21,

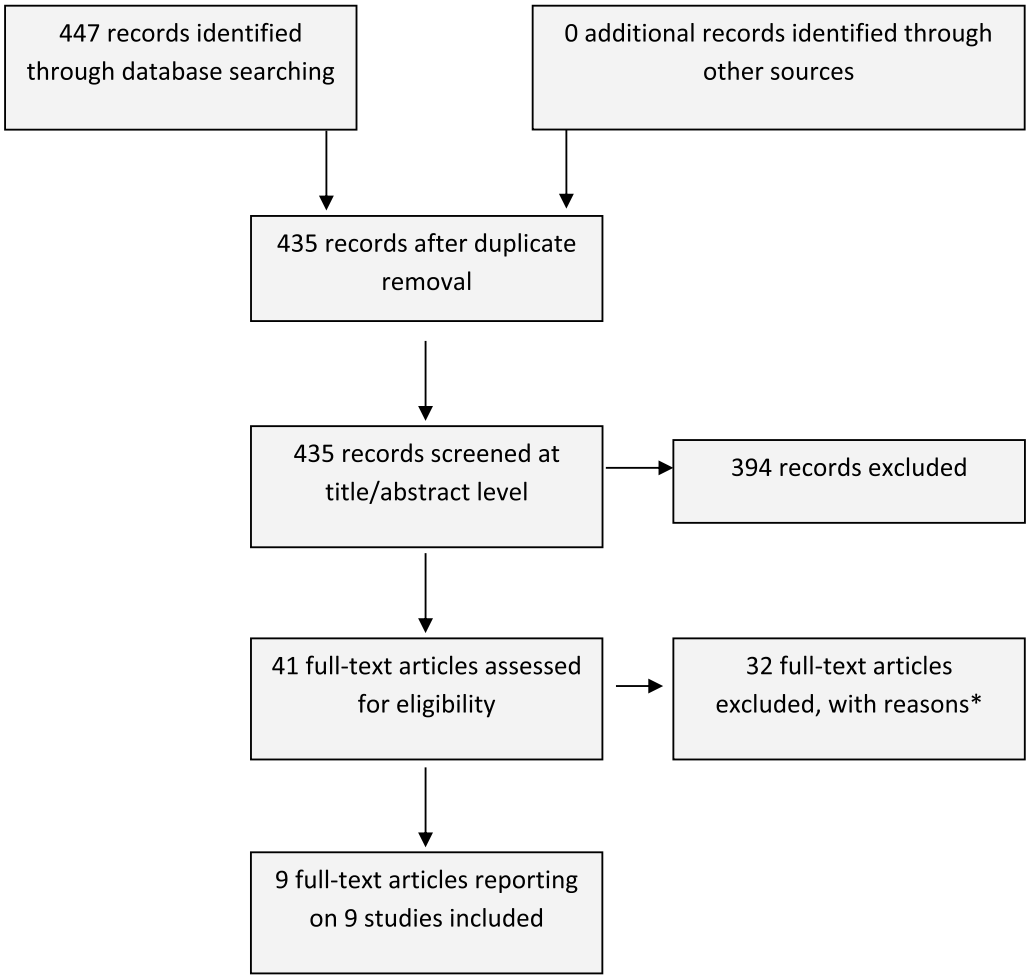


Fig. 1 PRISMA diagram showing selection of studies for a systematic review of the diagnostic performance of the IMMY cryptococcal antigen lateral flow assay on serum and cerebrospinal fluid in HIV-negative patients. *No full text available (6), paper not written in English (2), samples tested not from HIV-negative persons (adults or children) – or no disaggregate data for HIV-negative persons (13), study reports on < 4 cases (8), does not report use of IMMY CrAg® LFA on serum or CSF (3)

Table 1 Summary of characteristics of studies reporting findings of IMMY CrAg® lateral flow assay on serum and comparator test(s)

	Author, date of publication	Setting	n	Study population	Immunosuppressed n (%)	Cryptococcal disease phenotype	Samples tested (n)	Comparator test(s)	Reported sensitivity (95% CI)	Reported specificity
1	Dubbels 2017	USA	37	Not reported	15 (40.5%)	Cryptococcal meningitis, pulmonary cryptococcosis, other	36	Culture, histology, LA, composite[c]	Not calculated	66%
2	Harrington 2021	USA	96	Asymptomatic inpatients and outpatients	43 (45%)	No disease	35	LA	Not calculated	Not calculated
3	Hevey 2020	USA	34	Symptomatic inpatients	Not reported	Pulmonary cryptococcosis, other[b]	34	Composite clinical and laboratory end point[d]	Overall 85.3%, localised pulmonary 90.9% (58.7–99.8%), disseminated 82.6% (61.2–95.1%)	Not calculated
4	Jitmuang 2015	USA	31	Symptomatic inpatients	17 (55%)	Cryptococcal antigenemia,[a] Cryptococcal meningitis, pulmonary cryptococcosis, other[b]	53	LA, EIA	100% (92–100%)	Not calculated
5	Min 2020	China	78	Symptomatic inpatients	17 (22%)	Pulmonary cryptococcosis	78	Lung biopsy (histopathology)	69.2% overall, immunocompetent 80.3%, immunocompromised 29.4%	Not calculated
6	Tintelnot 2015	Germany	8	Not reported[g]	0	Cryptococcal antigenemia,[a] cryptococcal meningitis	9	LA	Not calculated	Not calculated
7	Wang 2020	China	149	Symptomatic inpatients	55 (37%)	Cryptococcal antigenemia,[a] cryptococcal meningitis, pulmonary cryptococcosis	136	Composite clinical and laboratory end point[e]	Titre 1:10 39.6% (29.7–50.1%), Titre 1:5 72.9% (62.9–81.5%)	Titre 1:10 100%(69.2–100%), Titre 1:5 70.0% (34.8–93.3%)
8	Wu 2020	China	37	Symptomatic and asymptomatic inpatients	15 (41%)	Pulmonary cryptococcosis	25	Composite clinical and laboratory end point[f]	Not calculated	Not calculated

Cryptococcal disease phenotypes

[a] Cryptococcal antigenaemia: isolated cryptococcal antigenaemia without evidence of disease

[b] Other: any other cryptococcal disease including disseminated disease

Definition of composites used as comparators

[c] (i) a Cryptococcus species was recovered in culture from any specimen source, (ii) a Cryptococcus species was histopathologically identified in any specimen, or (iii) the patient responded to targeted antifungal therapy with concomitant decreases in serial CrAg LFA titers [28]

[d] Positive serum or CSF CrAg, isolation of Cryptococcus neoformans in culture, or identification by the International Classification of Diseases (ICD) 9th (117.5, 321.0) or 10th (B45.1-B45.9) editions [34]

[e] Cryptococcal infections defined as: either "proven", "probable", "possible" or "non-crypto- coccosis", as described for other invasive fungal diseases, with some modifications in patients with low CrAg LFA titers as per De Pauw et al. [36]

[f] Proven pulmonary cryptococcosis based on the following criteria: histopathological, cytopathological or direct microscopic examination of a specimen obtained by a needle aspiration or biopsy from a normally sterile site (other than mucous membranes) showing encapsulated budding yeasts; or probable PC if each of the three elements of host factor, clinical features and mycological evidence were present [33]

[g] Testing of stored laboratory samples

Table 2 Summary of characteristics of studies reporting findings of IMMY CrAg® lateral flow assay on cerebrospinal fluid (CSF) and comparator test(s)

	Author, date of publication	Setting	n	Study population	Immunosuppressed n (%)	Cryptococcal disease phenotype	Samples tested (n)	Comparator test(s)	Reported sensitivity (95% CI)	Reported specificity
1	Chen 2016	China	58	Symptomatic inpatients	7 (12%)	Cryptococcal meningitis	85	India ink, culture, LAMP, qPCR	97.6%, (91.8–99.7%)	Not calculated
2	Dubbels 2017	USA	37	Not reported	15 (40.5%)	Cryptococcal meningitis, pulmonary cryptococcosis, other	12	Culture, histology, LA, composite[c]	Not calculated	66%
3	Harrington 2021	USA	96	Asymptomatic inpatients and outpatients	43 (45%)	No disease	79	LA	Not calculated	Not calculated
4	Jitmuang 2015	USA	31	Symptomatic inpatients	17 (55%)	Cryptococcal antigenemia,[a] Cryptococcal meningitis, pulmonary cryptococcosis, other[b]	11	LA, EIA	100% (66–100%)	Not calculated
5	Tintelnot 2015	Germany	8	Not reported[e]	0	Cryptococcal antigenemia,[a] cryptococcal meningitis	2	LA	Not calculated	Not calculated
6	Wang 2020	China	149	Symptomatic inpatients	55 (37%)	Cryptococcal antigenemia,[a] cryptococcal meningitis, pulmonary cryptococcosis	22	Composite clinical and laboratory end point[d]	CSF titre 1:10 50.0% (21.1–78.9%), CSF titre 1:5 66.7% (34.9–90.1%)	Not calculated

Cryptococcal disease phenotypes

[a] Cryptococcal antigenaemia: isolated cryptococcal antigenaemia without evidence of disease

[b] Other: any other cryptococcal disease including disseminated disease

Definition of composites used as comparators

[c] (i) a Cryptococcus species was recovered in culture from any specimen source, (ii) a Cryptococcus species was histopathologically identified in any specimen, or (iii) the patient responded to targeted antifungal therapy with concomitant decreases in serial CrAg LFA titers [28]

[d] Cryptococcal infections defined as: either "proven", "probable", "possible" or "non-cryptococcosis", as described for other invasive fungal diseases, with some modifications in patients with low CrAg LFA titers as per De Pauw et al. [42]

[e] Testing of stored laboratory samples

27, 33]. Other forms of reported immunosuppression included tuberculosis [27] and myasthenia gravis [32].

The majority of studies (six of seven) reported on the diagnostic performance of the IMMY LFA amongst symptomatic inpatients. This included patients with a range of cryptococcosis clinical phenotypes including cryptococcal antigenaemia (n = 56), cryptococcal meningitis (n = 103), pulmonary cryptococcosis (n = 233), and other cryptococcal disease including unspecified disseminated cryptococcosis (n = 39). There was significant heterogeneity between study cohorts with one study looking at cryptococcal meningitis only [27], two looking at only pulmonary cryptococcosis [31, 33], and the others including a combination of cryptococcal meningitis, pulmonary cryptococcosis and cryptococcal antigenaemia in varying proportions [28–30, 32, 34].

The diagnostic performance of the IMMY CrAg® LFA was compared to a wide range of comparators. Across all studies the results of IMMY CrAg® LFA testing on serum were compared with eight different cryptococcal diagnostic tests/composites: IMMY LA (n = 1), Meridian LA (n = 3), Biorad LA (n = 1) and Remel LA (n = 1) [21, 30, 32], Meridian EIA (n = 1) [32], culture of any site (n = 1) [28], histopathology (n = 2) [28, 31], and composites (n = 3) [29, 33, 34]. IMMY CrAg® LFA testing on CSF was compared with 10 different cryptococcal diagnostic tests/composites: IMMY LA (n = 1), Meridian LA (n = 3), Biorad LA (n = 1) and Remel LA (n = 1) [21, 30, 32], Meridian EIA (n = 1) [32], culture (n = 2) [27, 28], microscopy performed on India ink-stained samples (n = 1) [27], composites (n = 2) [28, 29], LAMP (n = 1) [27] and qPCR (n = 1) [27]. A total of three clinical composite end-point definitions were used, as described in the footnotes of Tables 1 and 2 [28, 33, 34].

Findings

Amongst eight studies which used the IMMY CrAg® LFA on serum to detect cryptococcal disease, the pooled sensitivity estimate, as compared to all comparator tests, was calculated as 96% (95%CrI 68–100%) and the pooled specificity estimate was calculated as 96% (95%CrI 84–100%). Amongst six studies which evaluated the diagnostic performance of IMMY CrAg® LFA on CSF, the pooled sensitivity was calculated as 99% (95%CrI 95–100%) and pooled specificity 99% (95%CrI 95–100%).

The estimated sensitivity and specificity of the IMMY CrAg® LFA in each study as well as the pooled estimates from testing on serum and CSF are shown in Figs. 2 and 3.

Methodological quality of included studies

Table 3 summarises the risk of bias and applicability concerns for each study. Overall there were no concerns about the applicability of the included studies. All studies were classified as having some risk of bias however, either with respect to (i) patient selection, (ii) interpretation of the index test, (iii) choice and/or interpretation of the reference standard, or (iv) sample flow and timing. In seven of nine studies, bias concerns were raised in ≥ 2 of the categories; three studies were classified as being at high risk of bias. The primary risk of bias category highlighted was in relation to interpretation of the index test, as in seven studies it was unclear whether the IMMY CrAg® LFA result was interpreted in isolation, without prior

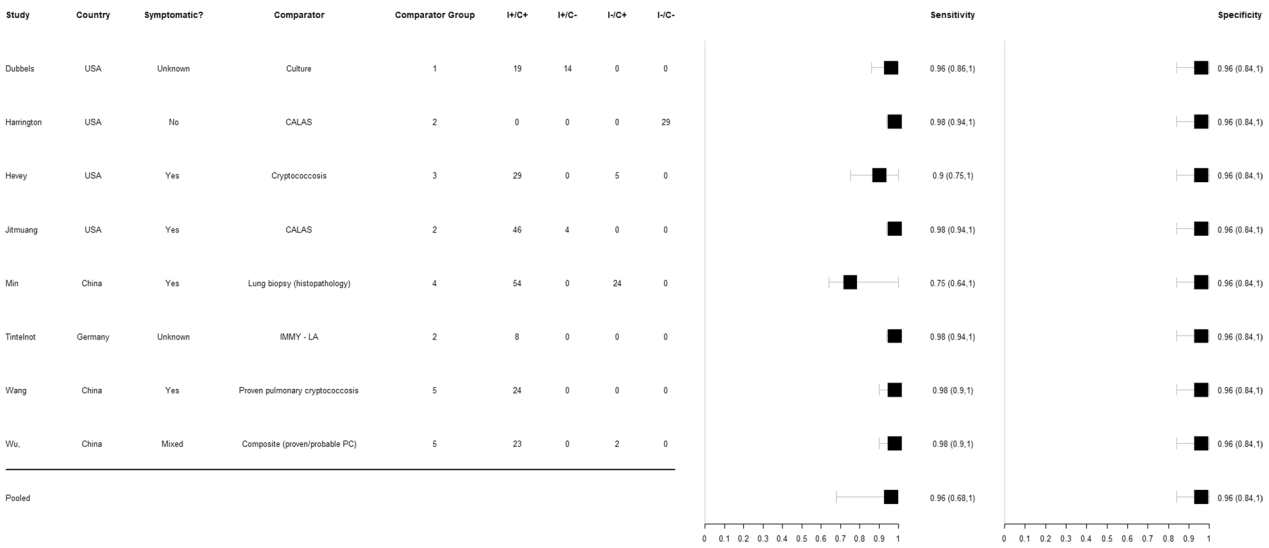


Fig. 2 Forest plot of IMMY CrAg® lateral flow assay sensitivity and specificity on serum

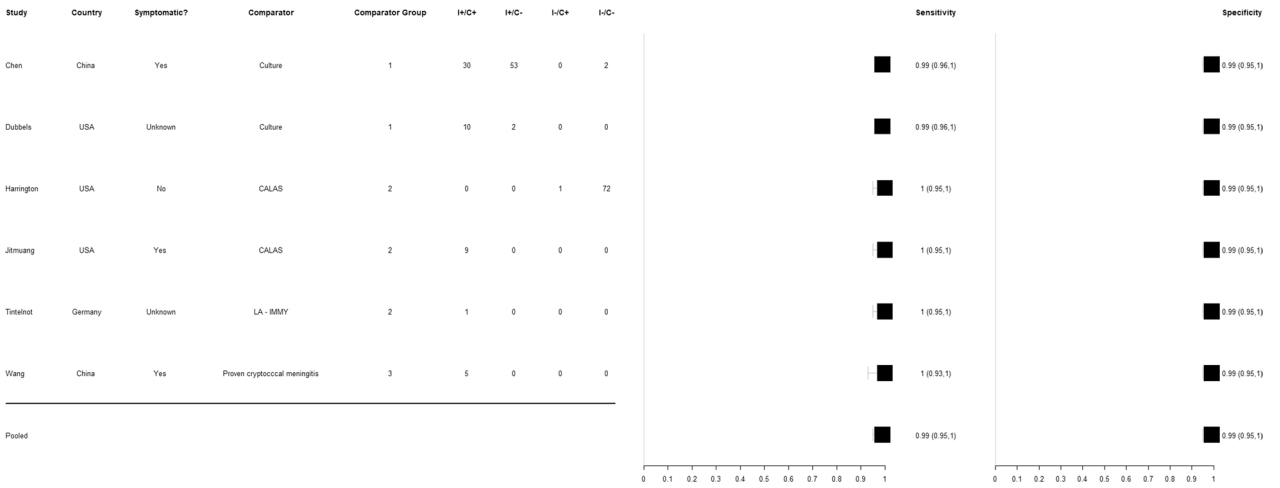


Fig. 3 Forest plot of IMMY CrAg® lateral flow assay sensitivity and specificity on cerebrospinal fluid

knowledge of the results of the comparator test/s. Additionally, four studies were classified as being an unclear risk of bias with respect to patient selection, because it was not reported if patient sampling was random and/or whether the study avoided inappropriate exclusions.

Discussion

In this review, we evaluated the diagnostic performance of the IMMY CrAg® LFA for diagnosis of cryptococcal disease amongst 528 HIV-negative persons from 9 studies. The point estimate for the sensitivity and specificity of the IMMY CrAg® LFA from the pooled values were good in both serum and CSF (>95% for both), in keeping with estimates reported in HIV positive cohorts [12, 13, 15, 16]. This is an important finding because, although the greatest burden of cryptococcal disease occurs in PLWH, this globally endemic fungal pathogen also infects HIV-negative individuals in increasing proportions.

These findings are consistent with the diagnostic accuracy literature from HIV-positive cohorts. In all published studies in PLWH, the IMMY CrAg® LFA has been found to be more sensitive than all other cryptococcal diagnostic tests. In a large multi-site validation study amongst PLWH in Uganda and South Arica, the IMMY CrAg® LFA performed on CSF was more sensitive than CSF culture (99.3% vs 90.0%), and more sensitive and specific than India ink microscopy on CSF (99.3% vs 86.1% and 99.1% vs 97.3% respectively) [12]. A study comparing IMMY CrAg® LFA to Meridian Cryptococcal Antigen Latex Agglutination System (CALAS®) and Meridian enzyme immunoassay (EIA), which tested 1,000 specimens (589 serum and 411 CSF) in parallel demonstrated higher sensitivity of the IMMY CrAg® LFA due to improved sensitivity for serotype C Glucuronoxylomannan (GXM) [13]. Similarly, the IMMY CrAg® LFA has better diagnostic performance than current PCR-based cryptococcosis diagnostics [15]. Amongst 328 adult and 42 paediatric CSF specimens evaluated using a multiplex PCR-based commercial assay (the BioFire® FilmArray® Meningitis/Encephalitis (ME) panel; BioFire Diagnostics, Salt Lake City, Utah, USA), for *Cryptococcus spp*., sensitivity was 82% and specificity was 98%, using CSF CrAg testing as the reference standard [15].

A systematic review and meta-analysis of 11 studies compared CrAg testing, of serum or CSF, to CSF microscopy with India ink staining, and CSF culture for the diagnosis of cryptococcal meningitis in symptomatic PLWH [16]. In all studies fungal culture was the reference standard for confirming cryptococcal meningitis. The review calculated the sensitivity and specificity of both LA and LFA CrAg tests on serum and CSF, using pooled data from multiple studies. For LA on serum (five diagnostic cohorts, 256 participants) the pooled sensitivity estimate was 100% (99.5–100) with pooled specificity estimate 96.7% (93.8–98.9). For LFA on serum (three diagnostic cohorts, 1690 participants) the pooled sensitivity estimate was 97.9% (87.9–100) and pooled specificity estimate was 89.5% (74.3–98.5). LA showed similar sensitivity in serum as LFA ($P=0.08$) and there was no statistically significant difference in specificity ($P=0.14$). For LA on CSF (10 diagnostic cohorts, 1810 participants) the pooled sensitivity was 97.1% (91.9–99.0) and pooled specificity was 99.1% (93.8–99.9). For LFA on CSF (6 diagnostic cohorts, 3099 participants) the pooled sensitivity was 99.5% (97.2–99.9) and pooled specificity was 99.5% (94.2–99.9). There was some evidence that LFA may have better sensitivity in CSF ($P=0.07$) than LA

Table 3 Quality assessment evaluating the risk of bias and applicability of all included studies using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Chen	☺	?	☺	☺	☺	☺	☺
Dubbels	☺	?	☹	?	☺	☺	☺
Harrington	☺	?	?	☺	☺	☺	☺
Hevey	☺	?	☺	☺	☺	☺	☺
Jitmaung	?	?	☺	☹	☺	☺	☺
Min	?	☺	☺	?	☺	☺	☺
Tintelnot	?	?	☺	☺	☺	☺	☺
Wang	?	☺	☺	?	☺	☺	☺
Wu	☺	?	?	☹	☺	☺	☺

Level of risk

☺Low Risk ☹High Risk ? Unclear Risk

QUADAS-2 Scoring System

Domain 1: patient selection

Risk of bias: could the selection of patients have introduced bias? Signalling question 1: was a consecutive or random sample of patients enrolled? Signalling question 2: was a case–control design avoided? Signalling question 3: did the study avoid inappropriate exclusions?

Applicability: are there concerns that the included patients and setting do not match the review question?

Domain 2: index test risk of bias: could the conduct or interpretation of the index test have introduced bias? Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard? Signalling question 2: If a threshold was used, was it pre-specified?

Applicability: are there concerns that the index test, its conduct, or interpretation differ from the review question?

Domain 3: reference standard risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias? Signalling question 1: is the reference standard likely to correctly classify the target condition? Signalling question 2: were the reference standard results interpreted without knowledge of the results of the index test?

Applicability: are there concerns that the target condition as defined by the reference standard does not match the question?

Domain 4: flow and timing risk of bias: could the patient flow have introduced bias? Signalling question 1: was there an appropriate interval between index test and reference standard?

Signalling question 2: did all patients receive the same reference standard?

Signalling question 3: were all patients included in the analysis?

but specificities were comparable ($P=0.54$) [16]. From our analysis the high sensitivity and specificity of IMMY CrAg® LFA in serum and CSF of HIV-uninfected people is in keeping with previously reported values in studies of CrAg testing on PLWH.

There were several limitations to our review. Firstly, due to the lack of data on performance of IMMY CrAg® LFA in HIV-negative people, only nine studies reporting results from a total of 528 participants were suitable for inclusion in the review. Amongst these studies, the diverse patient characteristics, range of comparator tests and cryptococcal disease phenotype made comparison difficult. The majority of studies recruited symptomatic patients or tested samples of patients known to have cryptococcal disease, with only one study screening asymptomatic patients. This limited the statistical analysis as there were very few negative IMMY CrAg® LFA results in the 2×2 tables. For this reason, a fixed effect meta-analysis was used. As a consequence of using a fixed-effect framework we do not suggest that these results are generalizable to other studies not included in this review. A fixed-effect meta-analysis assumes that the sensitivity and specificity is homogenous across studies and so does not account for variability between studies. As a result, our pooled estimates will underestimate the uncertainty by failing to account for this variability. Although we did not account for between-study heterogeneity, we did account for

within-study heterogeneity through the use of a fixed-effect conditional dependence structure between diagnostic tests in a study [35]. The small number of studies and limited data also prevented any further sub-analyses regarding the performance of the IMMY CrAg® LFA between different patient groups or between different cryptococcal species.

Another limitation was that the quality assessment using the QUADAS-2 tool identified unclear or high risk of bias in all studies. A key concern was that reference standard tests were interpreted with prior knowledge of the result of the index test and that populations being tested had already been classified as having cryptococcal infection. The flow and timing of testing was also unclear in a number of studies, with a lack of information regarding exclusions.

The main strengths of this review are that this is the first review looking at CrAg LFA testing of participants without HIV. It is also novel in calculating a value for specificity, where the majority of studies included in the review have focussed on sensitivity estimates only.

Conclusions

This review estimates a high sensitivity and specificity for IMMY CrAg® LFA in HIV-negative populations, as previously described for PLWH. However, our review was limited by a small number of disparate studies reporting IMMY CrAg® LFA testing on HIV-negative persons. Further studies using IMMY CrAg® LFA on both symptomatic patients being evaluated for cryptococcal disease and asymptomatic screening cohorts in HIV-negative populations are required to better predict the diagnostic value of this test. This is important given the increasing proportion of HIV-negative patients with cryptococcal infection.

Abbreviations

CALAS	Cryptococcal Antigen Latex Agglutination System
CSF	Cerebrospinal fluid
CrAg	Cryptococcal antigen
EIA	Enzyme immunoassay
ELISA	Enzyme-linked immunosorbent assay
GXM	Glucuronoxylomannan
HIV	Human immunodeficiency virus
HSROC	Hierarchical Summary Receiver Operating Characteristic
LA	Latex agglutination
LFA	Lateral flow assay
MALDI-TOF	Matrix-assisted laser desorption ionization–time-of-flight mass spectrometry
PCR	Polymerase chain reaction
PLWH	People living with HIV
QUADAS	Quality assessment of diagnostic-accuracy studies
RCT	Randomised controlled trial

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-023-08135-w.

Additional file 1. Search Methodology.

Acknowledgements

Not applicable.

Author contributions

CM, JE, HH, SHK and JNJ conceived the study. CM and JE assessed the eligibility of the studies, extracted the data, and assessed the methodological quality of the included studies. JF developed the search strategy and conducted the literature search. SHK carried out the statistical analysis. JB, RK and OB advised on the statistical analysis. CM, JE and SHK prepared the original draft of the manuscript, with considerable input from HH and JNJ. All authors contributed to the interpretations of results and all authors reviewed, edited and approved the final manuscript.

Funding

SHK is supported by the Medical Research Council London Intercollegiate Doctoral. Training Partnership Studentship (MR/N013638/1). The funders had no role in study design, data collection and analysis, or preparation of the manuscript.

Availability of data and materials

Code used for meta-analysis is publicly available at: https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/CrAg. Data included in analyses can be found in Tables 1 and 2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

Author details

[1]Infectious Diseases Unit, NHS Lanarkshire, University Hospital Monklands, Monkscourt Avenue, Airdrie ML6 0JS, UK. [2]London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. [3]Infectious Diseases Institute, Makerere University, PO Box 22418, Kampala, Uganda. [4]Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana.

Received: 22 December 2022 Accepted: 3 March 2023
Published online: 06 April 2023

References

1. Hagen F, et al. Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet Biol. 2015;78:16–48.
2. Kronstad JW, et al. Expanding fungal pathogenesis: Cryptococcus species break out of the opportunistic box James. Nat Rev Microbiol. 2011;9:193–203.
3. MacDougall L, Fyfe M, Romney M, Starr M, Galanis E. Risk factors for *Cryptococcus gattii* infection, British Columbia, Canada. Emerg Infect Dis. 2011;17:193–9.
4. Pyrgos V, Seitz AE, Steiner CA, Prevots DR, Williamson PR. Epidemiology of cryptococcal meningitis in the US: 1997–2009. PLoS ONE. 2013;8:e56269.

5. Bratton E, et al. Comparison and temporal trends of three groups with Cryptococcosis: HIV-infected, solid organ transplant, and HIV-negative/non-transplant. PLoS ONE. 2012;7:e43582.
6. Silveira FP, et al. Cryptococcosis in liver and kidney transplant recipients receiving anti-thymocyte globulin or alemtuzumab. Transpl Infect Dis. 2007;9:22–7.
7. Vijay S, Ingole N, Wanjare S, Mehta P. Prevalence of cryptococcaemia in HIV seropositive patients in an Indian setting. J Clin Diagnostic Res. 2019;13:DC01–4.
8. Pappas PG. Cryptococcal infections in non-HIV-infected patients. Trans Am Clin Climatol Assoc. 2013;124:61–79.
9. Chen J, et al. *Cryptococcus neoformans* strains and infection in apparently immunocompetent patients. China Emerg Infect Dis. 2008;14:755–62.
10. Tarumoto N, et al. Identification of disseminated cryptococcosis using MALDI-TOF MS and clinical evaluation. Med Mycol J. 2016;57:E41–6.
11. World Health Organisation. The diagnosis, prevention and management of cryptococcal disease in HIV-infected adults, adolescents and children. 2018.
12. Boulware DR, et al. Multisite validation of cryptococcal antigen lateral flow assay and quantification by laser thermal contrast. Emerg Infect Dis. 2014;20:45–53.
13. Hansen J, et al. Large-scale evaluation of the immuno-mycologics lateral flow and enzyme-linked immunoassays for detection of cryptococcal antigen in serum and cerebrospinal fluid. Clin Vaccine Immunol CVI. 2013;20:52–5.
14. Percival A, Thorkildson P, Kozel TR. Monoclonal antibodies specific for immunorecessive epitopes of glucuronoxylomannan, the major capsular polysaccharide of *Cryptococcus neoformans*, reduce serotype bias in an immunoassay for cryptococcal antigen. Clin Vaccine Immunol. 2011;18:1292–6.
15. Bridge S, et al. Evaluation of the biofire filmarray meningitis/encephalitis panel in and adult and pediatric Ugandan population. J Med Mycol. 2021;31: 101170.
16. Temfack JJB, Spijker R, Loyse A, Chiller T, Pappas PG, Perfect J, Sorell TC, Harrison TS, Cohen JF, Lortholary OER. Cryptococcal antigen in serum and cerebrospinal fluid for detecting *Cryptococcal meningitis* in adults living with human immunodeficiency virus: systematic review and meta-analysis of diagnostic test accuracy studies. Clin Infect Dis. 2021;72:1268–78.
17. Hopkins H, et al. Febrile illness evaluation in a broad range of endemicities (FIEBRE): protocol for a multisite prospective observational study of the causes of fever in Africa and Asia. BMJ Open. 2020;10: e035632.
18. von Hagen A, Balbi A, Arbildi C, Bonilla M. International prospective register of systematic reviews Citation Review question Participants/population Intervention (s), exposure (s) Comparator (s)/control Main outcome (s) International prospective register of systematic reviews Additional. Int Prospect Regist Syst Rev. 2018;1–4.
19. Falconer J. Removing duplicates from an EndNote Library. Library & Archives Service Blog. London School of Hygiene & Tropical Medicine (2018). Available at: https://blogs.lshtm.ac.uk/library/2018/12/07/removing-duplicates-from-an-endnote-library/.
20. Whiting PF, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.
21. Harrington K, et al. Evaluation of a cryptococcal antigen lateral flow assay and cryptococcal antigen positivity at a large public hospital in Atlanta, Georgia. Open Forum Infect Dis. 2021;8:123.
22. Murad MH, et al. Fixed-effects and random-effects models. in users' guides to the medical literature: a manual for evidence-based clinical practice. In: Guyatt G, Rennie D, Meade MO, Cook DJ. (eds) 3rd edn. McGraw-Hill Education; 2015.
23. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20:2865–84.
24. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics. 2012;68:1285–93.
25. Jullien S, et al. Diagnostic accuracy of multiplex respiratory pathogen panels for influenza or respiratory syncytial virus infections: systematic review and meta-analysis. BMC Infect Dis. 2022;22:1–11.
26. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual. (2019). Available at: https://mc-stan.org. Accessed: 12th Dec 2021.
27. Chen M, et al. Evaluation of five conventional and molecular approaches for diagnosis of cryptococcal meningitis in non-HIV-infected patients. Mycoses. 2016;59:494–502.
28. Dubbels M, Granger D, Theel ES. Low cryptococcus antigen titers as determined by lateral flow assay should be interpreted cautiously in patients without prior diagnosis of cryptococcal infection. J Clin Microbiol. 2017;55:2472–9.
29. Wang X, et al. Evaluation of low cryptococcal antigen titer as determined by the lateral flow assay in serum and cerebrospinal fluid among HIV-negative patients: a retrospective diagnostic accuracy study. IMA Fungus. 2020;11:6.
30. Tintelnot K, Hagen F, Seibold M, Bernhardt A, Boekhout T. Are cryptococcoses due to *Cryptococcus gattii* underdiagnosed in Europe? Mycoses. 2012;55:277–8.
31. Min J et al. Pulmonary cryptococcosis: comparison of cryptococcal antigen detection and radiography in immunocompetent and immunocompromised patients. BMC Infect Dis. 2020; 20.
32. Jitmuang A, et al. Performance of the cryptococcal antigen lateral flow assay in non-HIV-related cryptococcosis. J Clin Microbiol. 2016;54:460–3.
33. Wu HH, Chen YX, Fang SY. Clinicopathological features of isolated pulmonary cryptococcosis in HIV-negative patients. J Int Med Res. 2020;48:030006052092787.
34. Hevey IA, Rauseo AM, Larson L, Powderly W, Spec AMAG. Performance of the lateral flow assay and the latex agglutination serum cryptococcal antigen test in cryptococcal disease in patients with and without HIV. J Clin Microbiol. 2020;58:21.
35. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57:158–67.
36. De Pauw B et al. Revised definitions of invasive fungal disease from the european organization for research and treatment of cancer/invasive fungal infections cooperative group and the national institute of allergy and infectious diseases mycoses study group (EORTC/MSG) C. Clin Infect Dis. 2008;46:1813–21.

Publisher's Note

Appendix C

Leishmaniasis diagnostic test accuracy meta-analysis

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1702488	**Title**	Miss
First Name(s)	Suzanne		
Surname/Family Name	Keddie		
Thesis Title	Latent class models for diagnostic test accuracy with application to fever aetiology		
Primary Supervisor	John Bradley		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMC Infectious Diseases		
When was the work published?	09/11/2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

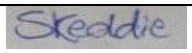
Where is the work intended to be published?	

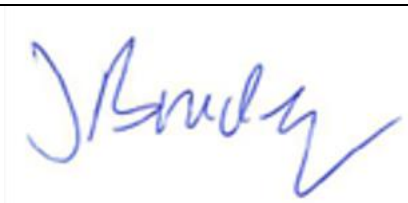
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I helped conceive the study as well as carrying out the statistical analysis, reporting and interpretation of results as well as helping prepare the original draft of the manuscript.

SECTION E

Student Signature	
Date	15/11/2023

Supervisor Signature	
Date	25/03/2024

BMC Infectious Diseases

RESEARCH

Open Access

Accuracy of the direct agglutination test for diagnosis of visceral leishmaniasis: a systematic review and meta-analysis

Tamalee Roberts[1,2*], Suzanne H. Keddie[3], Sayaphet Rattanavong[1], Santiago Rayment Gomez[4], John Bradley[3], Ruth H. Keogh[3], Oliver Bärenbold[3], Jane Falconer[3], Petra F. Mens[5], Heidi Hopkins[3] and Elizabeth A. Ashley[1,2]

Abstract

Background Parasitological investigation of bone marrow, splenic or lymph node aspirations is the gold standard for the diagnosis of visceral leishmaniasis (VL). However, this invasive test requires skilled clinical and laboratory staff and adequate facilities, and sensitivity varies depending on the tissue used. The direct agglutination test (DAT) is a serological test that does not need specialised staff, with just minimal training required. While previous meta-analysis has shown DAT to have high sensitivity and specificity when using parasitology as the reference test for diagnosis, meta-analysis of DAT compared to other diagnostic techniques, such as PCR and ELISA, that are increasingly used in clinical and research settings, has not been done.

Methods We conducted a systematic review to determine the diagnostic performance of DAT compared to all available tests for the laboratory diagnosis of human VL. We searched electronic databases including Medline, Embase, Global Health, Scopus, WoS Science Citation Index, Wiley Cochrane Central Register of Controlled Trials, Africa-Wide Information, LILACS and WHO Global Index. Three independent reviewers screened reports and extracted data from eligible studies. A meta-analysis estimated the diagnostic sensitivity and specificity of DAT.

Results Of 987 titles screened, 358 were selected for full data extraction and 78 were included in the analysis, reporting on 32,822 participants from 19 countries. Studies included were conducted between 1987–2020. Meta-analysis of studies using serum and DAT compared to any other test showed pooled sensitivity of 95% (95%CrI 90–98%) and pooled specificity of 95% (95%CrI 88–98%). Results were similar for freeze-dried DAT and liquid DAT when analysed separately. Sensitivity was lower for HIV-positive patients (90%, CrI 59–98%) and specificity was lower for symptomatic patients (70%, CrI 43–89%). When comparing different geographical regions, the lowest median sensitivity (89%, CrI 67–97%) was in Western Asia (five studies).

Conclusions This systematic review and meta-analysis demonstrates high estimated pooled sensitivity and specificity of DAT for diagnosis of VL, although sensitivity and specificity were lower for different patient groups and geographical locations. This review highlights the lack of standardisation of DAT methods and preparations, and the lack of data from some important geographical locations. Future well-reported studies could provide better evidence to inform test implementation for different patient populations and use cases.

PROSPERO registration CRD42021240830

*Correspondence:
Tamalee Roberts
tamalee.r@tropmedres.ac
Full list of author information is available at the end of the article

Keywords Visceral leishmaniasis, Direct agglutination test, Diagnosis, *Leishmania*, Meta-analysis

Background

Leishmaniasis is a vector-borne disease caused by the protozoan parasite of the genus *Leishmania.* The parasite is transmitted by the bite of female phlebotomine sand flies [1]. There are three main clinical forms of leishmaniasis: cutaneous leishmaniasis (CL) causing ulcerated skin lesions; mucocutaneous leishmaniasis (MCL), which can lead to partial or total destruction of mucous membranes; and visceral leishmaniasis (VL) which is a systemic, potentially lethal disease [2, 3]. There are an estimated 600,000 to 1 million new cases of CL and 50,000 to 90,000 new cases of VL reported annually [4]. As of 2021, there were 99 countries and territories endemic for leishmaniasis with 89% of global VL cases reported from eight countries: Brazil, Ethiopia, India, Kenya, Somalia, South Sudan, Sudan and Yemen [5]. VL has emerged as an opportunistic infection associated with HIV. People living with HIV are more likely to develop VL, and VL is an AIDS-defining condition due to both HIV and *Leishmania* suppressing the immune system which can result in more severe VL disease and higher mortality rates than from either infection in isolation [6].

Parasitological investigation of splenic, lymph node or bone marrow aspirates by microscopic examination for amastigotes remains the gold standard for the diagnosis of VL around the world with sensitivity ranging from 60–99% depending on the sample type [7]. These invasive tests require skilled clinical and laboratory staff and appropriate medical facilities meaning parasitological investigation is often not possible in leishmaniasis-endemic countries. Other diagnostic tests include the direct agglutination test (DAT), enzyme linked immunosorbent assay (ELISA), immunofluorescence antibody test (IFAT), immunochromographic tests, latex agglutination tests, leishmanin skin test (LST) and molecular techniques, including polymerase chain reaction (PCR). All of these tests use different sample types and have varying sensitivity and specificity but are more accessible within leishmaniasis-endemic countries.

DAT is a simple and reliable serological test for the diagnosis of VL which has been on the WHO's list of essential in vitro diagnostics since 2021. DAT is a semi-quantitative test that uses V-shaped well microplates with stained killed promastigotes of *L. donovani* or *L. infantum* mixed with increasing dilutions of patient's serum or blood. DAT detects the presence of antibodies against *Leishmania* parasites in the patient's serum or blood. If antibodies are present these will form an agglutination complex with the promastigotes which can be seen as a blue thin film on the walls of the microplates. The results can be read after 18 h of incubation. A titre-cut off for a specific dilution is used to determine if the sample is positive or negative for *Leishmania* with different cut-off titres used in different settings. Freeze dried antigen DAT (FD-DAT) and liquid antigen DAT (LQ-DAT) are the most common methods used and are based on the methods developed by Harith et al. [8]. LQ-DAT was developed first, however due to batch-to-batch variability as well as temperature sensitivity, FD-DAT was developed which remains stable at higher temperatures and has a higher shelf life with early validation studies showing similar results [9]. However, there are several other types of DAT including fast agglutination screening test (FAST-DAT) which uses a single serum dilution (qualitative result), formaldehyde fixed antigen DAT (FF-DAT) and in-house produced aqueous antigen DAT (AQ-DAT). Promastigotes and FD-DAT kits are produced by the former Royal Tropical Institute (KIT) Amsterdam (now Academic Medical Centre (AMC), Amsterdam), the Netherlands and the Institute of Tropical Medicine in Antwerp (ITMA), Belgium but liquid DAT is also often produced locally in-house with local strains.

Previous meta-analysis of DAT compared to parasitological examination for patients with *L. donovani* and *L. infantum* from studies published from 1986 to 2004 showed a pooled sensitivity and specificity of 94.8% and 85.9% respectively [10], while for studies from 2004 to 2019 the pooled sensitivity and specificity were 96% and 95% respectively [11]. In HIV-positive patients, DAT compared to microscopy showed lower sensitivity using random effects models of 81% and a specificity of 90% [12]. While meta-analysis of DAT accuracy compared to parasitological tests has been reported, meta-analysis of DAT compared to other diagnostic techniques, which are increasing in use in clinical and research settings and in low- and middle- income country (LMIC) settings, has not been done. This makes comparison of different studies and burden estimation difficult. Therefore, the aim of this study was to carry out a systematic review and meta-analysis to assess the diagnostic accuracy of DAT for human VL compared to all available tests up to February 2021.

Methods

Selection criteria

Eligible studies included prospective and retrospective studies on the diagnosis of human VL, independent of study design, that reported results of DAT and at least

one comparator test. Case studies with ≤ 5 people and studies on diseases other than VL were excluded. See Table 1 for full inclusion and exclusion criteria.

Literature search strategy

The review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [13] (Additional file 1) and is registered with the international prospective register of systematic reviews (PROSPERO CRD42021240830).

Search terms were developed and carried out in 10 databases: OvidSP Medline (1946 to 12 February 2021), OvidSP Embase Classic + Embase (1947 to 12 February 2021), OvidSP Global Health, (1910 to week 05 2021), Elsevier Scopus (complete database), Clarivate Analytics Web of Science Science Citation Index (1970-present), Clarivate Analytics Web of Science Social Sciences Citation Index (1970-present), Wiley Cochrane Central Register of Controlled Trials (Issue 2 of 12, February 2021), Ebsco Africa-Wide Information (complete database), WHO LILACS (complete database), WHO Global Index Medicus (complete database). All searches were run on 15 February 2021. There was no restriction on language (see Additional file 2 for search strategy).

All citations identified from the searches were imported into EndNote X9 software. Duplicates were identified and removed using the method described on the London School of Hygiene & Tropical Medicine Library & Archives Service blog [14]. Additional eligible studies were hand-searched from the reference list of relevant manuscripts.

Study selection and full-text review

Two independent reviewers (SR, TR) screened all titles and abstracts, as well as full texts when the abstract did not provide sufficient information, for compliance with the inclusion and exclusion criteria. Results were compared and discrepancies discussed with a third reviewer (EA).

Data extraction

Selected full articles were screened independently and data extracted by three reviewers (SR, TR and SRG). All data was checked by a second reviewer and discrepancies discussed. Variables included bibliographic information, sample type, study design, study location, study population, number of participants, HIV status, DAT test kit details, comparator test details, number of samples tested by each test, patient age group, cross-reaction and quality control information, number positive and negative by each test and 2 × 2 tables (DAT + /Comparator + , DAT-/Comparator + , DAT + /Comparator-, DAT-/Comparator-).

Assessment of study quality

The quality of studies was assessed using the Quality Assessment of Studies of Diagnostic Accuracy Approach-2 (QUADAS-2) [15]. Studies were assessed in duplicate by three assessors (SR, TR and SRG) and results compared.

Categorisation of tests

All studies irrespective of DAT type were included. Where possible, further grouping of DAT was done according to manufacturer (e.g. ITMA, KIT, AMC, In-house) but sub-group analysis was not performed due to

Table 1 Systematic review of the direct agglutination test (DAT) for the diagnosis of visceral leishmaniasis in humans: inclusion and exclusion criteria

Inclusion criteria	Prospective and retrospective studies on diagnosis of visceral leishmaniasis, independent of study design
	Any variation of the DAT technique and a comparator standard used for diagnosing visceral leishmaniasis
	Paired data: The same samples tested with any variation of the DAT method compared to any comparator standard
	Epidemiological and or laboratory studies
Exclusion criteria	Lack of data (studies that do not include, for example, individual participant results, comparator standard, *Leishmania* species/origin, study type, sample type; see also Fig. 1)
	Discrepancies suspected between the studied group and the control group
	Reviews
	Commentaries
	Case studies with ≤ 5 cases
	Duplicate publications
	Patients with others types of leishmaniasis or infectious diseases other than visceral leishmaniasis
	Patients with post kalazar dermal leishmaniasis
	Animal studies

the low number of studies for each manufacturer type. However, sub-group analysis was carried out for FD-DAT and LQ-DAT as they were the DAT types most commonly reported.

Geographical classification of countries and study population

Countries were classified by geographic sub-region following United Nations designations [16]. Study populations were grouped into four categories: neonates (aged ≤ 28 days), infants (1 to < 12 months), children (1 to < 13 years), and adolescents/adults (≥ 13 years). If a study reported participants from each age group, they were grouped as participants of "all ages".

Statistical analyses

Data extracted from each study included the 2 × 2 table comparing results of the index test (DAT) and a comparator test. Where a study presented results of DAT and multiple comparator tests, a 2 × 2 table for each comparator test was extracted. Descriptive analysis was completed for all studies.

To summarise the data from this review and estimate the sensitivity and specificity of DAT, the study implemented an extension to the hierarchical summary receiver operating characteristic model [17] described by Dendukuri, et al. [18]. Using this Bayesian model framework allows estimation of the accuracy of a diagnostic test in the absence of gold-standard comparator tests while taking account of within- and between-study variability (for example, each study is assumed to use a different positivity threshold). Within this approach the assumption of conditional independence between an individual's test results, given their disease status, can be relaxed through the use of random-effects [19].

The outputs of this review are the estimated accuracy (sensitivity and specificity) with credible intervals [CrI]) of DAT within each study, as well as a pooled and predicted estimate of the test's accuracy across all studies included. Results are presented in forest plots and as summary receiver operating characteristic (SROC) curves. Pooled sensitivity and specificity represent the summary of DAT test accuracy across studies included in this review, while predicted sensitivity and specificity allow estimation of the accuracy of DAT in a hypothetical future study. Where there is variability among studies, predicted sensitivity and specificity are less precise than pooled sensitivity and specificity.

A meta-analysis was fit on all data from serum samples only due to it being the most common sample type and to limit the number of variables that may affect the accuracy of DAT and therefore to strengthen the analysis and interpretation of results, irrespective of the specific DAT test and comparator test. We investigate heterogeneity by DAT test type (FD-DAT or LQ-DAT), geographic region, participant status (e.g. symptomatic or HIV-positive) and whether or not the assumption of conditional independence is assumed. Where multiple 2 × 2 tables are available from the same study and represent the same individuals, only one table is included in the meta-analysis to prevent including individuals in the model more than once. In these cases, we chose to include the 2 × 2 table where the comparator test's accuracy (sensitivity and/or specificity) was better established (the one with the smallest difference between the 2.5 and 97.5 percentiles of the proposed prior distribution) as this understanding of the comparator test could be included into the model as informative prior distributions. For example, if there was a 2 × 2 table between DAT and PCR as well as DAT and microscopy, the 2 × 2 table with microscopy was chosen.

All analyses were carried out in R using stan [20]. A full model specification can be found in Additional file 3. All code can be found at: https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/Leishmaniasis.

Results

Search results

A total of 2571 articles were retrieved, 1584 of which were duplicates resulting in 987 titles and abstracts screened. Of these, 358 articles were selected for full data extraction and after full data extraction 78 articles were included that had complete 2 × 2 tables (Fig. 1).

Study description

The included studies reported on a total of 32,822 patients from 19 countries. All patients were from countries endemic for leishmaniasis or had travelled to endemic countries. Ten studies included only adolescents/adults, four included only children, 47 included all age groups and 17 did not report participant ages. The studies included in this review were conducted from 1987 to 2020, inclusive.

Serum was tested in 63/78 (80.8%) studies, whole blood in 8/78 (10.3%), plasma in 5/78 (6.4%) and the sample type was not reported in 2/78 (2.6%). There was a range of DAT titre cut-offs with the lowest 1:100 and the highest ≥ 120,000. HIV-positive patients were included in 8/78 (10.3%) studies. Cross-reaction of DAT for VL with other diseases was noted in 8/78 (10.3%) studies, with cross-reaction noted for patients with cutaneous leishmaniasis (two studies), malaria (two studies), leukaemia (two studies), schistosomiasis (one study), Chagas disease (one study) and connective tissue disorder and lymphoma (one study). Symptomatic patients were included in 22/78 (28.2%) studies, asymptomatic in 9/78 (11.5%), both symptomatic and asymptomatic in 38/78 (48.7%)

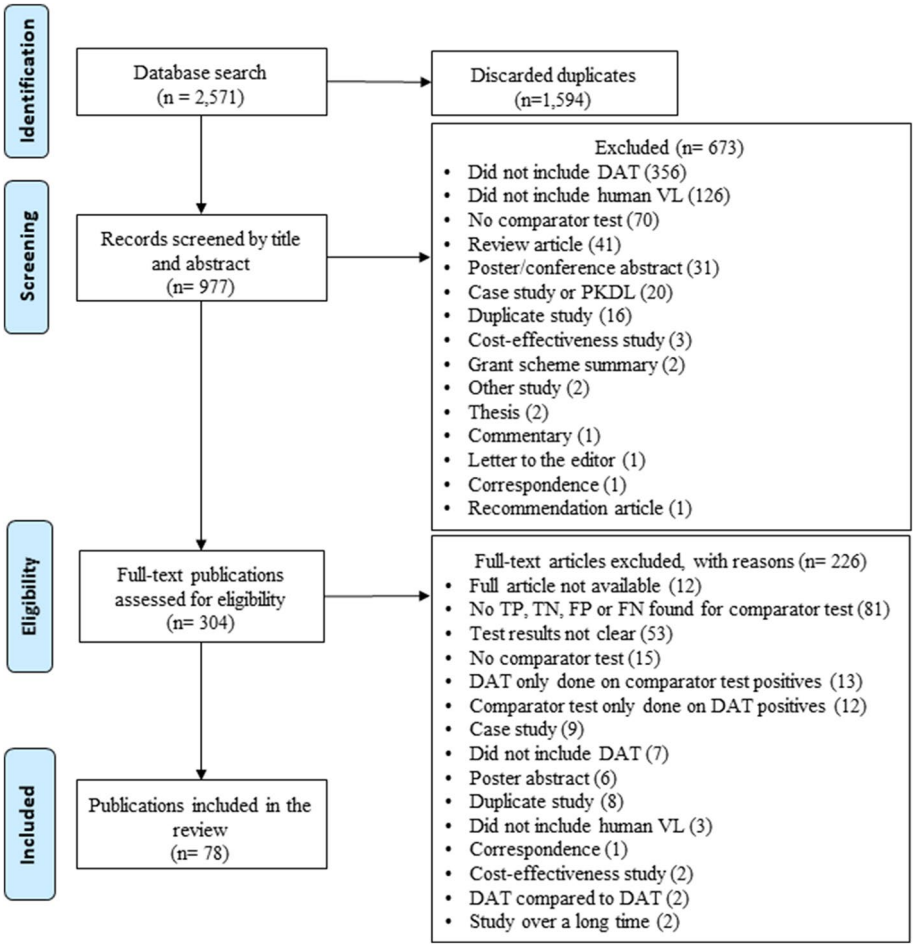


Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram of publications screened in a systematic review of the direct agglutination test for the diagnosis of visceral leishmaniasis. DAT = Direct agglutination test; VL = Visceral leishmaniasis; PKDL = Post-kalazar dermal leishmaniasis; TP = True positive; TN = True negative; FP = False positive; FN = False negative

and it was not reported whether patients were symptomatic or not in 9/78 (11.5%) studies. Laboratory testing quality control was clearly stated in 22/78 (26.8%) studies. There were eight different types of DAT and 21 different comparator tests with the median number of comparator tests in included studies 1 (range 1–4).

Meta-analysis sensitivity and specificity

Sixty-three studies representing 20,364 individuals that used serum samples and any DAT and comparator test were included in a meta-analysis; the results are shown in Fig. 2. The pooled sensitivity across all included studies was 95% (95% CrI 90–98%) and the pooled specificity across all included studies was 95% (95% CrI 88–98%). The estimated median sensitivity and specificity of DAT in included studies ranged from 2 to 100% and from 7 to 100% respectively. The predicted sensitivity and specificity were 96% (95% CrI 8–100%) and 96% (95% CrI 9–100%) respectively. The pooled and predicted estimates are shown as summary ROC curves in Fig. 3.

We investigated heterogeneity in sensitivity and specificity estimates by patient group, geographic region, DAT test type and by relaxing the assumption of conditional independence between diagnostic tests within a study. The pooled estimates of DAT sensitivity and specificity from each analysis are shown in Fig. 4 and the predicted estimates are shown in Fig. 5. The number of studies and individuals included in each analysis are shown in Table 2. The pooled sensitivity estimates of DAT across these different analyses ranged from 89%-97% while the pooled specificity ranged from 70%-98%. Overall, from the studies included in this analysis, estimated DAT accuracy differed only slightly by geographical region. Western Asia (including five studies) had the lowest median sensitivity (89%, CrI 67–97%) and Europe (including two studies) was the region with the most uncertainty in sensitivity and specificity estimates. Pooled sensitivity and

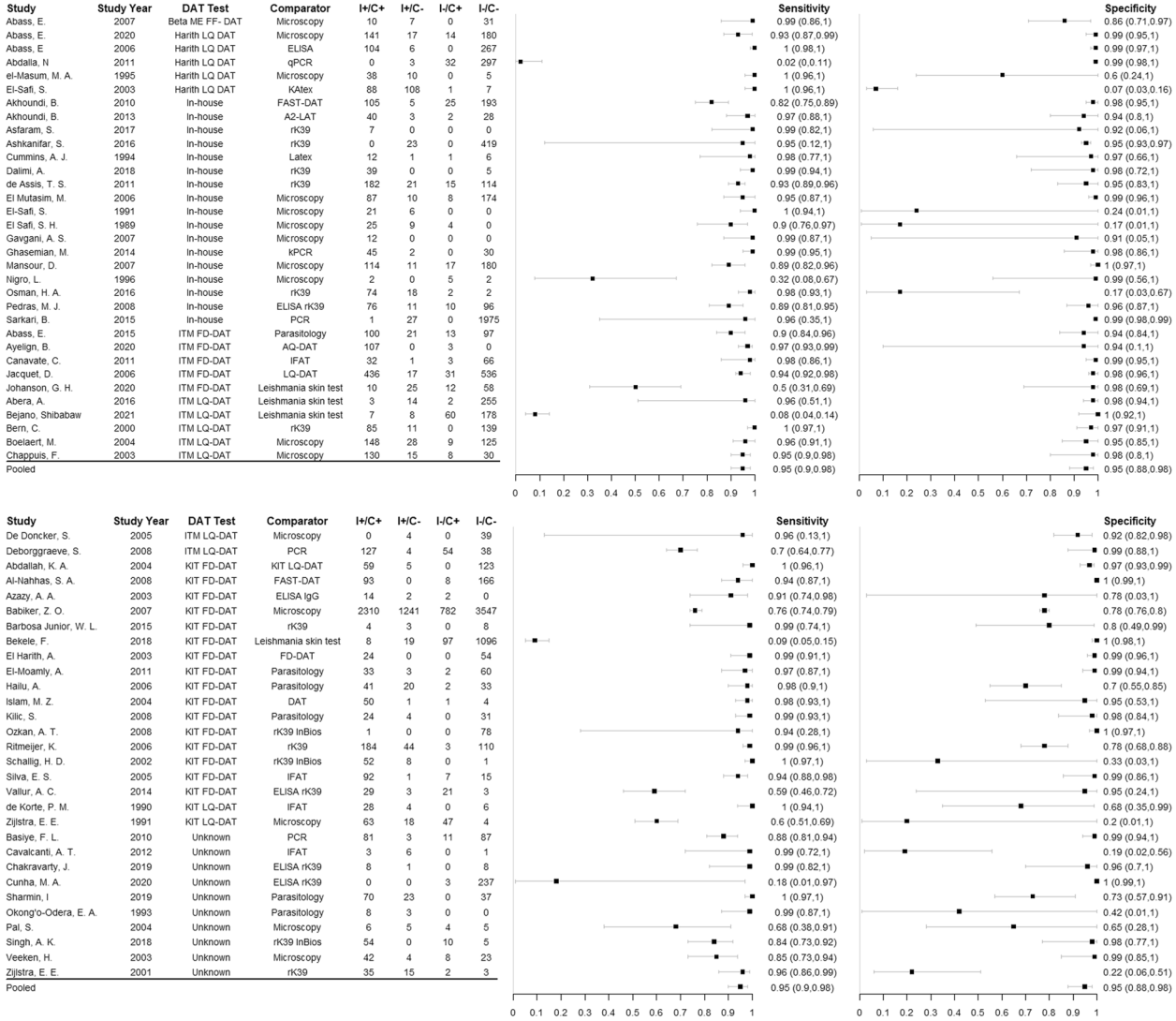


Fig. 2 Forest plot of the sensitivity and specificity of DAT for the 63 studies that included serum sample type

specificity were also similar when FD-DAT and LQ-DAT were analysed separately. Within the different patient groups, sensitivity and specificity estimates varied with a lower sensitivity and specificity with wider uncertainty for the HIV-positive patient group (90%, CrI 59–98% and 91%, CrI 63–99 respectively) and a low specificity for the symptomatic patient group (70%, CrI. 43–89%) compared to analyses that did not differentiate by patient group. The small number of studies that included HIV-positive patients, and the heterogeneity across these, precluded further investigation of accuracy among HIV-infected subgroups. Similarly, the small number of studies that included children, or distinguished between adult and child participants, precluded age-stratified analysis.

The predicted estimates of DAT sensitivity and specificity by sub-group all have wide credible intervals because of heterogenous results from individual studies within a sub-group. The analysis considering only FD-DAT had the narrowest predicted credible intervals at 95% (CrI 38–100%) and 97% (CrI 51–100%) for sensitivity and specificity respectively, representing the DAT type with the least variability between studies.

Assessment of study quality

Additional file 4 summarises the QUADAS-2 risk of bias and applicability concerns for the final included studies. Full information was not available for one study so it was not included in the QUADAS-2 analysis. For patient selection 39/77 (50.6%) studies had a high risk of bias,

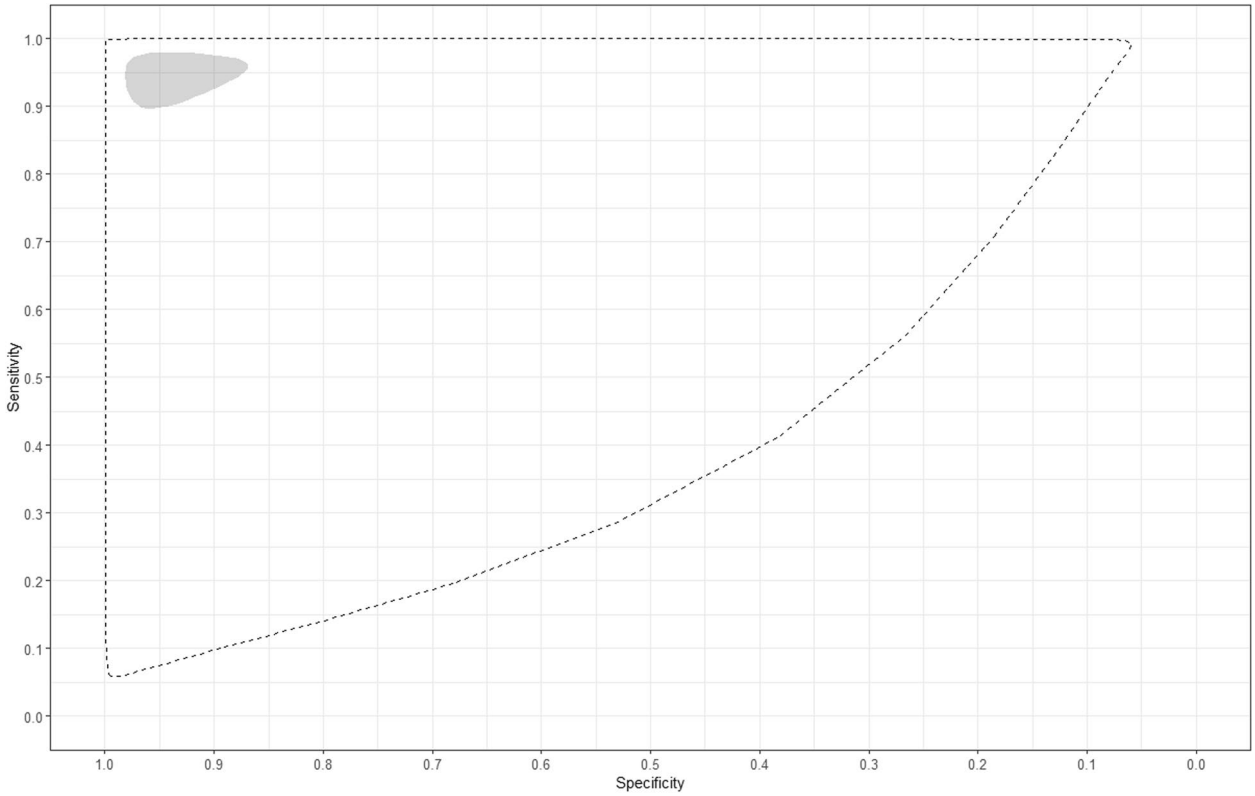


Fig. 3 Summary receiver operating characteristic curves for pooled (shaded) and predicted (dashed) estimates of DAT for the 63 studies including serum sample type

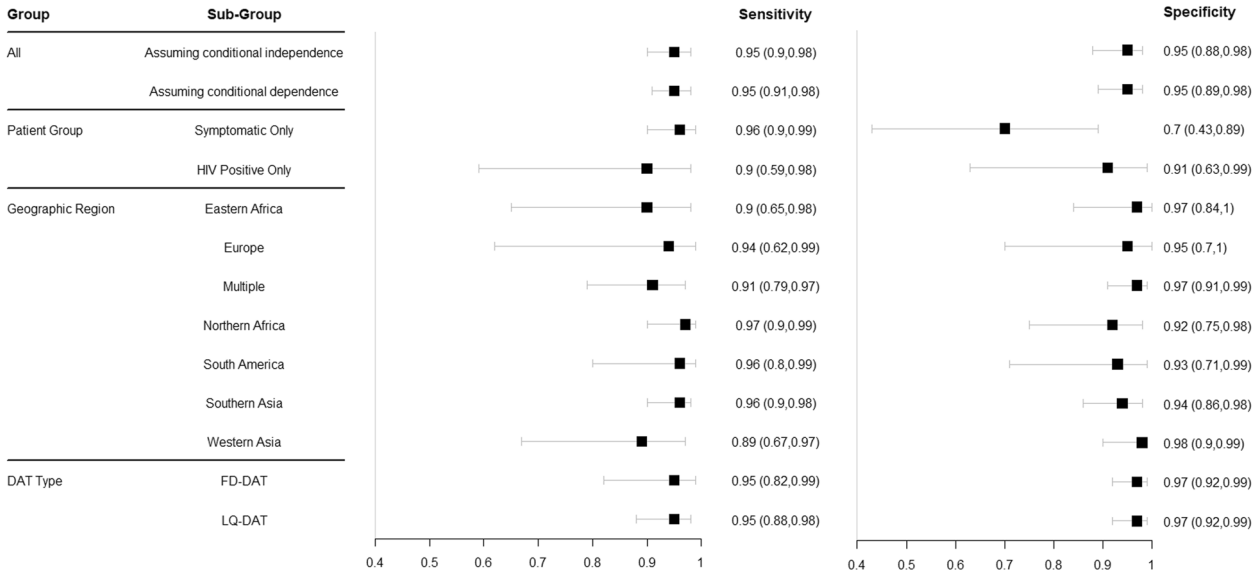


Fig. 4 Forest plot of the pooled estimates of sensitivity and specificity grouped for all the data, studies from symptomatic patients, studies grouped by geographic region, and specific analysis for studies that include FD-DAT or LQ-DAT

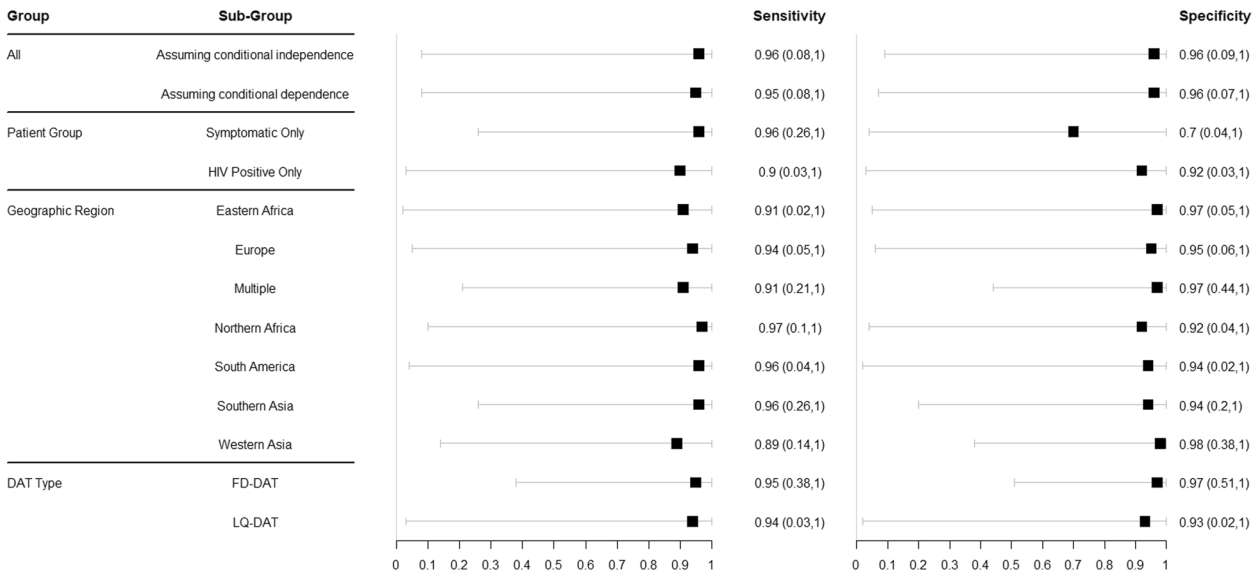


Fig. 5 Forest plot of the predicted estimates of sensitivity and specificity grouped for all the data, studies from symptomatic patients, studies grouped by geographic region, and specific analysis for studies that include FD-DAT or LQ-DAT

7/77 (9.1%) had a high risk of bias for the index test, all had a low risk of bias for the comparator test and 4/77 (5.2%) had a high risk of bias for the flow and timing. All studies had low concern for applicability for patient selection, index test and comparator test.

Discussion

This review evaluated the sensitivity and specificity of DAT, compared with all other available tests, for detecting human VL. We included 63 studies in a meta-analysis and found a high pooled sensitivity for any DAT of 95% (95% CrI 90–98%) and a pooled specificity of 95% (95% CrI 88–98%). This is similar to previous reviews that looked specifically at DAT compared to parasitological examination [10, 11]. We found little variability between geographic regions or DAT test type but sensitivity and specificity did vary when used in only symptomatic or HIV-positive patients.

A number of different DAT types were included in the initial analysis for serum samples with different antigen preparations and methods, and antigens may have been made on site or ordered from a manufacturer. Due to the number of different DAT preparation types, it was not possible to do sub-analysis on all distinct test types. The different DAT preparation types may impact real-world results, however, and future studies may be needed to compare the different preparation types. In this analysis, we found that the predicted estimates of sensitivity and specificity for FD-DAT had narrower credible intervals compared to LQ-DAT when analysed separately, suggesting that FD-DAT would be a more appropriate test.

We estimated the specificity of DAT to be lower when used in studies that enrolled only symptomatic patients, compared with studies that included any patient type (i.e. studies including both symptomatic and asymptomatic patients or those of unknown status). This finding aligns with a previous meta-analysis of DAT compared to microscopy where the lowest specificity was seen in patients clinically suspected to have VL [10]. Cross-reaction of both the FD-DAT and LQ-DAT with other diseases was reported in eight studies. This cross-reaction may have been one cause for the lower specificity of DAT for symptomatic patients. Cross-reaction is also something that laboratories and clinicians need to be aware of especially when testing patients from countries endemic for diseases such as malaria and other parasitic diseases. However, the reporting of patient selection was a concern for bias in 50% of included studies so these results should be treated with caution. Future studies should report clear patient selection criteria so that sub-group analysis can be carried out.

HIV-positive or otherwise immunocompromised patients frequently have low or undetectable anti-leishmanial antibodies meaning there is the potential for false-negative results from serological tests like DAT [12, 21]. Results from this review support this as a sub-analysis of studies with only HIV-positive patients showing a lower sensitivity and specificity with wider credible intervals than when all patients were included. As a result, studies using serological tests for the diagnosis of VL in HIV-positive patients should not rule out infection with a single negative test result.

Table 2 The number of studies and individuals broken down for the meta-analysis for studies using specimen type serum

Model	Number of studies included	Number of Individuals represented
All	63	20,364
Symptomatic only	20	9266
HIV positive only	8	826
Geographic region: Eastern Africa	8	2248
Geographic region: Europe	2	47
Geographic region: Multiple	5	1537
Geographic region: Northern Africa	17	10,628
Geographic region: South America	7	966
Geographic region: Southern Asia	19	4220
Geographic region: Western Asia	5	521
DAT test: LQ	25	6094
DAT test: FD	19	12,273

Geographical region had some impact on pooled estimated sensitivity when FD-DAT and LQ-DAT were combined, with lower sensitivity seen for Western Asia. However, there were only five studies and 521 patients from Western Asia. The geographical variation was also seen in a previous review where sensitivity was higher in South Asia compared to other regions [10]. The sensitivity and specificity for the various DAT types and geographical regions should be taken into consideration by public health decision-makers when implementing diagnostic tests and further country-specific analysis should be done. Unfortunately, in this review there was not enough country-specific data to do focused country analysis and some regions were also lacking data (Europe and Western Asia). For some regions, even when multiple studies were available, data came from only one or two countries; more representative data would help to confirm if there is spatial heterogeneity in test accuracy. The species of promastigotes used in the DAT preparation may also affect test sensitivity and specificity, depending on their match with actual circulating *Leishmania* species in the region of interest.

Only 26% of studies clearly stated whether VL testing was subject to quality control, e.g. by testing samples at a reference laboratory and stating the positive and negative controls for the DAT. While most laboratories probably use controls and carry out quality control, stating this in the methods gives confidence to readers that laboratories' results meet a recognised standard. For many studies, it was difficult to interpret the results due to a lack of clarity in reporting positive and negative results for each test. This review included studies published both before and after the introduction of STROBE (2007), STARD (2015) and MICRO (2019) reporting guidelines; even studies published after introduction did not adhere to the guidelines. Studies reporting laboratory and diagnostic test comparisons should report results to a standardised format, for consistency and comparability.

There are several limitations to this systematic review and meta-analysis. While we tried to be comprehensive and included studies comparing DAT to any other test, this identified reports on a wide range of DAT and comparator tests performed on a variety of sample types in a variety of geographical regions. This diversity may not have been fully captured by the analysis, and as a result estimates of DAT test accuracy may be biased in either direction. Despite this, our results are in line with other published estimates from studies with more restrictive selection criteria [8, 9]. A second limitation is that only a single 2×2 table from each study was included in the meta-analysis and the selection of which 2×2 table to include was somewhat subjective. Finally, different titre cut-offs were used across studies included in this review, but this was not accounted for in analysis due to the large number of different cut-offs used; further analyses that incorporate titre cut-offs would help to improve estimates of DAT accuracy. The different titre cut-offs used in this analysis may impact the sensitivity and specificity with lower cut-offs potentially resulting in false positives and higher titre cut-offs resulting in false negatives.

Strengths of this review include the rigorous and comprehensive approach which included 78 studies representing 32,822 individuals across 19 countries endemic for VL. Another strength is our analysis framework which did not assume any comparator test was perfect and which estimated both pooled and predicted sensitivity and specificity.

Conclusion

Despite variability across studies in terms of geographic location, patient characteristics and comparator tests used, overall this systematic review and meta-analysis demonstrates that DAT performs well compared to other diagnostic methods in most scenarios. However, the test is generally not standardised with many methods and preparations of DAT in use. There is also a lack of data on DAT performance outside of South Asia and Northern Africa, with no data from Southeast Asia. Future studies carried out in a variety of locations with well documented DAT preparations are required to improve estimates of the DAT accuracy, and to better inform implementation for different patient populations and use cases.

Abbreviations

AMC Academic Medical Centre
AQ-DAT Aqueous antigen direct agglutination test

CL	Cutaneous leishmaniasis
CrI	Credible interval
DAT	Direct agglutination test
ELISA	Enzyme linked immunosorbent assay
FAST-DAT	Fast agglutination screening test direct agglutination test
FD-DAT	Freeze dried antigen direct agglutination test
FF-DAT	Formaldehyde fixed antigen direct agglutination test
HIV	Human immunodeficiency virus
IFAT	Immunofluorescence antibody test
ITMA	Institute of Tropical Medicine in Antwerp
KIT	Royal Tropical Institute
LQ-DAT	Liquid antigen direct agglutination test
LST	Leishmanin skin test
MCL	Mucocutaneous leishmaniasis
PCR	Polymerase chain reaction
QUADAS-2	Quality Assessment of Studies of Diagnostic Accuracy Approach- 2
RDT	Rapid diagnostic test
SROC	Summary receiver operating characteristic
VL	Visceral leishmaniasis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-023-08772-1.

Additional file 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.

Additional file 2. What is the accuracy of the direct agglutination test (DAT) for the diagnosis of visceral leishmaniasis in humans? Search methodology.

Additional file 3. *Leishmania* DAT Review Model Specification.

Additional file 4. QUADAS-2 scoring for each study. Studies were scored as high or low risk of bias and applicability.

Acknowledgements
Not applicable.

Authors' contributions
TR, SK, HH, and EA conceived the study. JF developed the search strategy and conducted the literature search. TR, SR and EA assessed the eligibility of the studies. TR, SR, SRG and PM extracted the data and assessed the methodological quality of the included studies. SK carried out the statistical analysis with support from JB, OB and RK. All authors contributed to the interpretation of results. TR and SK prepared the original manuscript draft with considerable input from PM, SRG, SR, EA and HH. All authors edited drafts and read and approved the final manuscript.

Funding
S. Keddie is supported by the Medical Research Council London Intercollegiate Doctoral Training Partnership Studentship (MR/N013638/1). This research was funded in whole, or in part, by the Wellcome Trust [220211]. Funders had no role in study design, analysis or writing for this study.
For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and additional files and available in the github repository (https://github.com/shk313/diagnostic-test-metaanalysis/tree/main/Leishmaniasis).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Lao- Oxford-Mahosot Hospital- Wellcome Trust Research Unit, Mahosot Hospital, Mahosot Road, Vientiane, Lao People's Democratic Republic. [2]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [3]London School of Hygiene & Tropical Medicine, London, UK. [4]Department of Infectious Disease Epidemiology, Faculty of Medicine, London Centre for Neglected Tropical Disease Research, St Marys Campus, Imperial College London, London, UK. [5]Department of Medical Microbiology, Experimental Parasitology Unit, Amsterdam University Medical Centers, Academic Medical Center at the University of Amsterdam, Amsterdam, The Netherlands.

Received: 29 June 2023 Accepted: 30 October 2023
Published online: 09 November 2023

References

1. Dostalova A, Volf P. Leishmania development in sand flies: parasite-vector interactions overview. Parasit Vectors. 2012;5:276.
2. Alvar J, Yactayo S, Bern C. Leishmaniasis and poverty. Trends Parasitol. 2006;22(12):552–7.
3. Bart A, van Thiel PP, de Vries HJ, Hodiamont CJ, Van Gool T. Imported leishmaniasis in the Netherlands from 2005 to 2012: epidemiology, diagnostic techniques and sequence-based species typing from 195 patients. Euro Surveill. 2013;18(30):20544.
4. World Health Organization. Leishmaniasis. Available: https://www.who.int/news-room/fact-sheets/detail/leishmaniasis. [Cited 20 Oct 2022].
5. World Health Organization. The Global Health Observatory. Leishmaniasis. Available: https://www.who.int/data/gho/data/themes/topics/gho-ntd-leishmaniasis. [Cited 22 Feb 2023].
6. World Health Organization. WHO guideline for the treatment of visceral leishmaniasis in HIV co-infected patients in East Africa and South-East Asia.7 June 2022, Guideline. Available: https://www.who.int/publications/i/item/9789240048294. [Cited 22 May 2023].
7. Thakur S, Joshi J, Kaur S. Leishmaniasis diagnosis: an update on the use of parasitological, immunological and molecular methods. J Parasit Dis. 2020;44(2):253–72.
8. Harith AE, Kolk AH, Kager PA, Leeuwenburg J, Muigai R, Kiugu S, et al. A simple and economical direct agglutination test for serodiagnosis and sero-epidemiological studies of visceral leishmaniasis. Trans R Soc Trop Med Hyg. 1986;80(4):583–636.
9. Jacquet D, Boelaert M, Seaman J, Rijal S, Sundar S, Menten J, et al. Comparative evaluation of freeze-dried and liquid antigens in the direct agglutination test for serodiagnosis of visceral leishmaniasis (ITMA-DAT/VL). Trop Med Int Health. 2006;11(12):1777–84.
10. Chappuis F, Rijal S, Soto A, Menten J, Boelaert M. A meta-analysis of the diagnostic performance of the direct agglutination test and rK39 dipstick for visceral leishmaniasis. BMJ. 2006;333(7571):723.
11. Mohebali M, Keshavarz H, Shirmohammad S, Akhoundi B, Borjian A, Hassanpour G, et al. The diagnostic accuracy of direct agglutination test for serodiagnosis of human visceral leishmaniasis: a systematic review with meta-analysis. BMC Infect Dis. 2020;20(1):946.
12. Cota GF, de Sousa MR, Demarqui FN, Rabello A. The diagnostic accuracy of serologic and molecular methods for detecting visceral leishmaniasis in HIV infected patients: meta-analysis. PLoS Negl Trop Dis. 2012;6(5):e1665.
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Reprint–preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther. 2009;89(9):873–80.
14. J. F. Removing duplicates from an EndNote Library. Library & Archives Service Blog 2018: London School of Hygiene & Tropical Medicine. Available from: https://blogs.lshtm.ac.uk/library/2018/12/07/removing-duplicates-from-an-endnote-library/. [Cited 01/03/2021].

15. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med. 2011;155(8):529–36.
16. United Nations. Standard country or area codes for statistical use (M49). Available: https://unstats.un.org/unsd/methodology/m49/overview/. [Cited 04 Feb 2021].
17. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20(19):2865–84.
18. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics. 2012;68(4):1285–93.
19. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001;57(1):158–67.
20. Stan Development Team. Stan Modelling Language Users Guide and Referenece Manual, 2.28. 2019. Available: https://mc-stan.org. [Cited 12 Dec 2021].
21. Lindoso JAL, Moreira CHV, Cunha MA, Queiroz IT. Visceral leishmaniasis and HIV coinfection: current perspectives. HIV AIDS (Auckl). 2018;10:193–201.

Publisher's Note

Appendix D

Dengue diagnostic test accuracy meta-analysis

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	1702488	**Title**	Miss
First Name(s)	Suzanne		
Surname/Family Name	Keddie		
Thesis Title	Latent class models for diagnostic test accuracy with application to fever aetiology		
Primary Supervisor	John Bradley		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

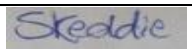
Where is the work intended to be published?	Lancet Infection Diseases

Please list the paper's authors in the intended authorship order:	Kamla Pillay*, Suzanne H. Keddie*, Elizabeth Fitchett, Cassandra Akinde, Oliver Bärenbold, John Bradley, Jane Falconer, Ruth H. Keogh, Zhia Lim, Behrouz Maldonado, Laura Maynard-Smith, Ellen Sugrue, Okuda Taylor, Heidi Hopkins, Audrey Dubot-Peres
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I am joint first author of the paper. I developed, coded and carried out the statistical analysis with support from my supervisors. I also created data visualizations of the results for the manuscript. Along with the other first author, I prepared the original draft of the manuscript.

SECTION E

Student Signature	
Date	25/03/2024

Supervisor Signature	
Date	25/03/2024

Diagnostic accuracy of common reference laboratory tests for acute dengue detection: a systematic review and meta-analysis of NS1 ELISA, IgM ELISA, IgG ELISA, RT-PCR, and viral neutralisation

Authors:**Kamla Pillay***, **Suzanne H. Keddie***, Elizabeth Fitchett, Cassandra Akinde, Oliver Bärenbold, John Bradley, Jane Falconer, Ruth H. Keogh, Zhia Lim, Behrouz Nezafat, Laura Maynard-Smith, Ellen Sugrue, Okuda Taylor, Heidi Hopkins+, Audrey Dubot-Pérès+

Affiliations:

Background

Dengue fever is identified by WHO as among the top ten global health threats. Prompt identification of dengue can guide clinical management and outbreak response, yet laboratory diagnosis is complex, costly, and lacks consensus on performance evaluation. Reliable performance estimates of laboratory reference tests are essential for updating global diagnostic guidance, and for evaluating novel diagnostics including point-of-care tests. 'Gold standard' tests for dengue, such as reverse transcription- polymerase chain reaction (RTPCR), are infrequently used in the highest burden settings, resulting in limited data for traditional diagnostic accuracy meta-analyses.

Methods

Nine literature databases were systematically searched for reports on five common reference tests for dengue infection: IgG, IgM and NS1 enzyme-linked immunosorbent assay (ELISA), RT-PCR and viral neutralisation (VNT). Two-by-two tables were extracted from included articles comparing one of these tests with any comparator. We estimated pooled and predicted values of sensitivity and specificity for each test using Bayesian random-effect meta-analysis, which does not require a 'gold standard' comparator. Risk of bias was assessed using QUADAS-2. This study is registered with Prospero (CRD42022341552).

Findings

Data was extracted from 161 articles, which allowed analysis of three overlapping windows for three tests of interest: 0 to 4, 1-7 and 1-14 days post symptom onset (DPO). Pooled sensitivities of RT-PCR (0 to 4 days), NS1 ELISA (0 to 4 days) and IgM ELISA (1-7 days) were 95% (95% Credible Interval (CrI) 77-99%), 90% (95% CrI 68-98%), 71% (95% CrI 5784%) respectively. The corresponding pooled estimates of specificity were 89% (95% CrI 6098%), 93% (95% CrI 71-99%) and 91% (95% CrI 82-95%).

Interpretation

We reiterate that IgM ELISA has poor diagnostic accuracy early in the symptom course. NS1 ELISA performed with similar diagnostic accuracy to RT-PCR, which has important implications for global public health policy, given its relatively low cost and accessibility.

Panel: Research in context

Evidence before this study

Previously, only traditional diagnostic accuracy meta-analytical methods have been used to estimate diagnostic accuracy of reference laboratory tests for dengue. The lack of a perfect comparator test is a major limitation of these methods. Estimates of diagnostic accuracy for each individual test vary based on which comparator used. In clinical practice, dengue virus (DENV) RT-PCR is seen as a 'gold standard', though performing this assay is expensive and difficult in resource-limited settings, and the short viremic period in dengue infection limits its utility.

Added value of this study.

This systematic review and meta-analysis is novel in terms of methods, size and scope. To our knowledge, this is the largest review of dengue diagnostic accuracy to date. We used Bayesian effect latent class analysis to compare tests of interest with any comparator, enabling us to capture a broad range of real-world data. 11,048 articles were screened and data from 161 full texts were included in the analysis to ascertain the diagnostic accuracy of five reference laboratory tests for dengue. Our results confirm that IgM ELISA should not be used as a single test in the first days of symptom onset, due to poor diagnostic accuracy in this period. Furthermore, we find that the diagnostic accuracy of NS1 ELISA is similar to RTPCR, with

public health implications given ELISA's relatively low cost and ease of access. These estimates can also be used in future diagnostic accuracy studies, for example in the evaluation of novel rapid diagnostic tests (RDTs), which have the potential to further improve dengue diagnosis in high-need settings.

Introduction

Dengue fever is identified by the WHO as one of the top ten threats to global health (1). Dengue incidence has increased in recent years, with 5.2 million reported cases in 2019, a ten-fold increase from 2000 (2). This is likely to underestimate the true numbers given that the majority of cases are self-limiting (3). A probable case of dengue as defined by the WHO is fever plus two additional features of the disease. The WHO further categorises dengue infection into three clinical categories: with or without warning signs, and severe dengue (4). Warning signs, such as abdominal pain and mucosal bleeding, indicate high risk of progression to severe dengue infection (5). Dengue can be categorised as primary, defined as an individual's first DENV infection, or secondary, referring to DENV infection in an individual who has been previously infected by another DENV serotype. Secondary infection brings a higher risk of severe disease, in part due to antibody-dependent enhancement (6).

The diagnostic methods used for suspected DENV infection are wide-ranging and settingdependent. Laboratory confirmation in a probable case of dengue is defined by the WHO as one of: RT-PCR or viral culture positive, IgM or IgG seroconversion in paired sera or fourfold IgG titre increase in paired sera (4). Laboratory results defined as highly suggestive in acute dengue include IgM positivity, IgG positivity with a house index (HI) titre of 1280 or greater, or detection of NS1 antigen by ELISA or rapid test, all of which relate to a single serum sample (4). Use of novel rapid diagnostic tests (RDTs) , which do not require laboratory capacity and which align with the REASSURED criteria, is increasing (7). RDTs allow for timely diagnosis in high-need, low-resource settings. Multiplex RDTs bring the possibility of testing for multiple pathogens simultaneously, of particular use in areas of co-circulation of pathogens with similar clinical presentations. In order to evaluate and assess RDTs, diagnostic accuracy of traditional reference tests must be well defined.

Multiple factors impact the diagnostic accuracy of tests for acute dengue; critical among these is the time point in disease progression when the test is performed, often measured in days post onset of symptoms (DPO). Dengue viraemia is estimated to peak 0 to 4 DPO. While methods that identify the virus, such as viral isolation and RT-PCR, are thought to be most

accurate during this time frame, methods focused on antibody response, such as IgM ELISA, peak later in the illness and remain positive for a longer period, limiting their utility in detecting acute infection (8).

The under-reporting and misdiagnosis due to other causes of febrile illness or co-circulating flaviviruses, such as zika and Yellow Fever viruses, are compelling reasons why dengue control programmes need to use quality-assured diagnostics for surveillance to provide earlier warning of outbreaks, a more accurate estimate of the extent of the outbreak and to monitor the effectiveness of control interventions (9,10). Although there are publications on the evaluation of dengue diagnostics, the lack of consensus on a “gold” standard makes the interpretation of diagnostic accuracy difficult when different reference standards are used. To overcome these limitations, we undertook a large-scale dengue diagnostics systematic review and meta-analysis of the performance of five reference laboratory dengue tests, NS1 ELISA, RT-PCR, IgM ELISA, IgG ELISA and virus neutralisation test (VNT). These tests were chosen given their presence in the WHO and PAHO guidelines, their use in reference laboratory testing for dengue (10) and lack of previous Bayesian diagnostic accuracy meta-analyses. We estimate the sensitivity and specificity of each test using a Bayesian framework that does not assume that the comparator tests are perfect.

Methods

Search strategy and selection criteria

A search strategy was compiled in the OvidSP Medline database by an experienced information specialist (JF). The search strategy included strings of terms, synonyms and controlled vocabulary terms (where available) to reflect two concepts: dengue and diagnostic test of interest (PCR, ELISA or VNT). The search strategy was refined until the results retrieved reflected the scope of the project. The full search strategy including the list of databases searched and the search terms for one representative database are included in supplement 1.

Results were initially limited to articles published from 2000 to the search date; due to the high quantity of extractable data that resulted, at the full-text screening stage the publication date range was further adjusted so that articles ultimately included were from a 10-year period (01 January 2011 to 16 February 2021).

All citations identified by our searches were imported into EndNote X9 software (12) and duplicates were removed. A total of 11,048 unique results were imported into a systematic review software, Cadima, for abstract screening (13). The selection criteria are listed in Table 1.

Table 1: Selection criteria.

Selection criteria	
1	Any study design with n > 5 participants, except for case series which were excluded
2	Article written in English, Spanish, French or Portuguese
3	Article reports primary evidence
4	Article published in a peer-reviewed journal
5	Article reports results of at least one test of interest (NS1, IgG or IgM ELISA, RT-PCR (conventional or real time) or VNT) against any comparator
6	The test of interest was performed on human serum samples
7	The test of interest and comparator were performed on the same samples (for serology paired samples were extracted if the first sample was tested with comparator)
8	Accuracy data reported as sensitivity and specificity with denominators and clear calculation methodology, OR as a 2 x 2 table

All 11,048 abstracts were screened against the selection criteria by at least one of four reviewers (KP, EF, ZL, BM), using the Cadima platform (13). Prior to commencing the abstract screening process, all reviewers undertook a consistency check of 50 abstracts with a kappa value of > 0.80 (defined as ‘excellent’ concordance) between every pair of reviewers. Any disagreements in the consistency check were discussed in order to clarify the cause of disagreement and prevent it from recurring. 10% of all abstracts were screened by two

independent reviewers, in order to highlight any changes in consistency over the course of the abstract screening process. Concordance remained high throughout the screening process. Any abstracts marked as ‘unclear’ were included in the list for full-text screening.

A total of 1715 full-text articles were screened by at least one of seven reviewers (KP, SK, EF, LMS, ES, OT, CA). Consistency checks were performed for every reviewer and compared to outcome based on review by KP and SK. Good concordance was found for every reviewer (kappa value > 0.80). Any full texts marked ‘unclear’ were reviewed by a second reviewer (KP, SK). In total, 570 (33%) texts were reviewed independently by two reviewers.

Data Extraction

Data was extracted in duplicate by two independent reviewers (KP, SK). Discordance between the reviewers was minimal and any disagreement was resolved by discussion. The following data were extracted from included articles into an Excel database: country of data collection, DENV serotype, participant population (including study selection criteria, demographic and clinical details), the test of interest and comparator test used including brand and methodology, duration of sample storage, days post onset (DPO) of symptoms that specimens for testing were collected, and a two-by-two table of the test of interest versus the comparator. If an article included data on multiple comparator tests, two-by-two tables were extracted for each pairwise comparison of a test of interest and a comparator test.

Assessment of study bias

The quality of each study included was assessed using the Quality Assessment of Studies of Diagnostic Accuracy Approach-2 (QUADAS-2), an internationally recognised tool used to assess risk of bias in diagnostic accuracy studies (14). We modified this tool to make it suitable for our review question; the full criteria assessed are detailed in supplement 5. Each study was assessed in duplicate by two assessors (KP and ES).

Statistical analysis

Statistical analysis focused on estimating the diagnostic accuracy of the tests of interest for acute dengue in symptomatic individuals. For the purposes of this review, acute is defined as within two weeks of symptom onset (15). Data on the timing of diagnostic testing days post onset of symptoms were categorised into three groups, hereafter referred to as DPO subgroups: 0 to 4 days, 1 to 7 days, and an overall category of all acute symptomatic cases. The first two groups are mutually exclusive and were chosen to reflect the virologic and immunologic events used as diagnostic test targets and the resultant differences in diagnostic accuracy of tests of interest at different DPO (9,16–18). For NS1 ELISA and RT-PCR the 0 to 4 days subgroup represents the primary analysis model, and for IgM and IgG ELISA the primary analysis model is the 0 to 7 day subgroup. Where data was stated to be from symptomatic individuals but no DPO was given these were assumed to be within the acute window and included in the all acute symptomatic analysis. Data from asymptomatic individuals, data where neither DPO nor symptom status was reported and data from symptomatic individuals with a stated DPO range that exceeded 14 days were excluded from all analyses. Where there were fewer than four studies in a DPO subgroup, meta-analysis was not carried out.

Where articles included data from multiple comparator tests from the same individuals and DPO subgroup, multiple two-by-two tables were extracted. To prevent data from the same individuals being included more than once for the same test of interest, only one of these two-by-two tables was used for analysis. Comparisons with reference tests that have more established performance characteristics were prioritised, because prior understanding of the sensitivity and specificity of reference tests could be incorporated in the model as informative prior distributions (assumptions).

For all statistical analyses, each two-by-two table presented a comparison of results for both a given test of interest and a reference test within a unique cohort of individuals (hereafter referred to as a ‘study’), with some articles contributing two or more studies.

We estimated pooled and predicted sensitivity and specificity, including 95% credible intervals (CrI), for each test of interest. Here, pooled estimates represent the median calculated accuracy across studies included in this analysis while predicted estimates represent the expected accuracy in another hypothetical future study. We used an extension to the Hierarchical Summary Receiver Operating Characteristic Model (19) which relies on Bayesian latent class analysis and is recommended in the Cochrane Handbook for Systematic Reviews

of Diagnostic Test Accuracy (20). This modelling framework takes account of both within- and between-study variability and allows for multiple imperfect comparator tests.

Heterogeneity was explored by looking at the difference in pooled and predicted estimates (21) as well by comparing the primary analysis DPO subgroup to the other subgroups. Initial plans to explore heterogeneity by participant age cohort (adults vs children) and study country were not possible due to small numbers in each subgroup. Sensitivity analyses explored the impact of including only those studies at low risk of bias and with high external validity (supplement 6). All analyses were carried out in R (22) using stan (23). A full model specification can be found in supplement 2.

Results

Description of all eligible texts from systematic review

28,043 articles were identified from the electronic searches, which reduced to 11,048 after removing duplicates and to 2,940 after abstract screening (Figure 1). The date restriction to include a 10-year period of data reduced this further to 2,133. After a full-text review, 193 articles were identified as eligible in the systematic review (98 for RT-PCR, 123 for IgM ELISA, 64 for IgG ELISA, 67 for NS1 ELISA and 12 for viral neutralisation). The included articles contributed 214, 122, 112, 219 and 28 of the two-by-two table comparisons (studies) for IgM ELISA, IgG ELISA, NS1 ELISA, RT-PCR and viral neutralisation respectively.

Figure 1: PRISMA diagram

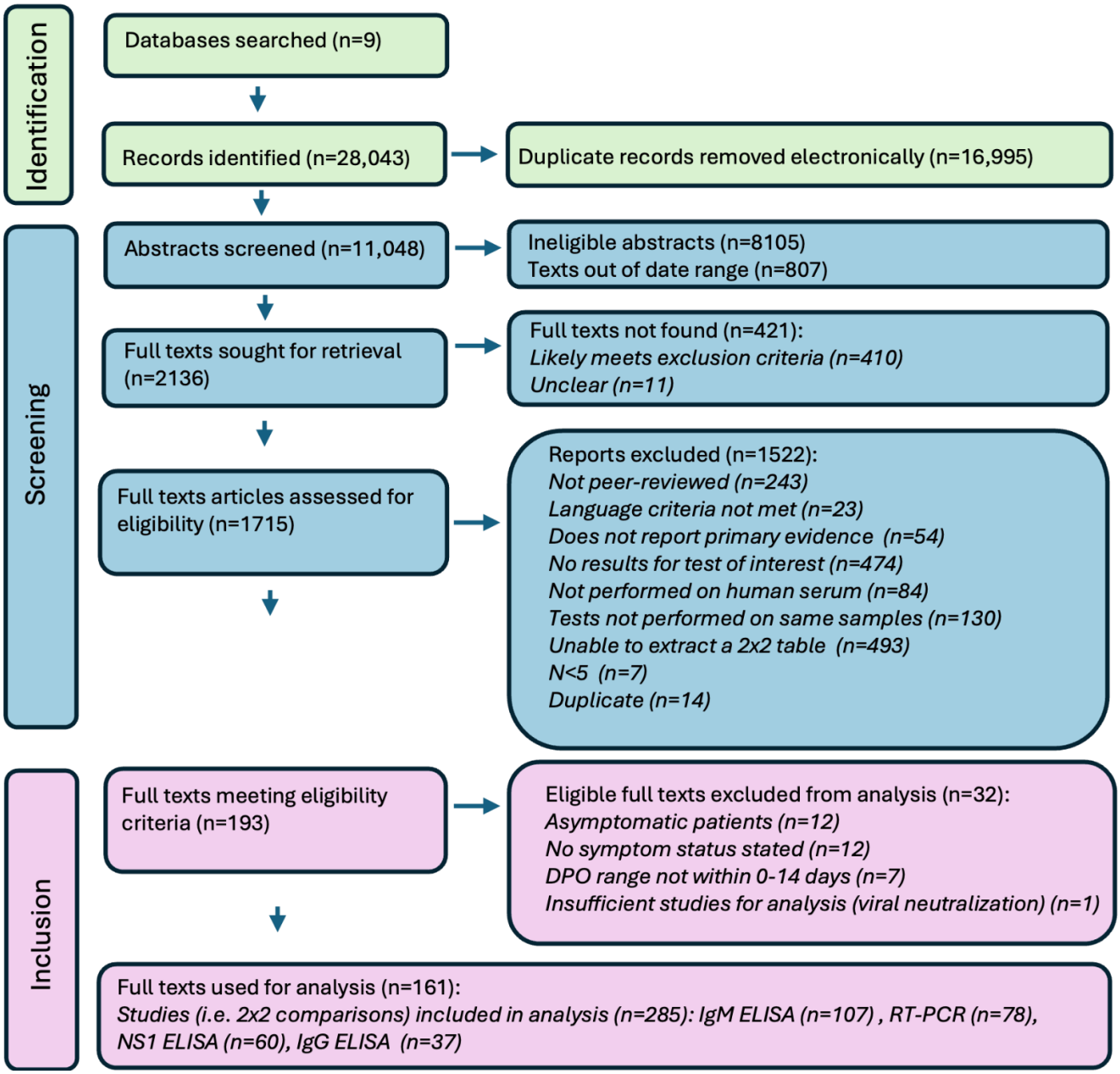


Figure 1: PRISMA diagram of all articles screened.

Description of studies included in the analysis

A full list of studies included in the meta-analyses can be found in supplement 3. IgM ELISA had the largest number of studies to be included in analyses (n=107), representing 67,828 individuals; VNT had the smallest (n=3) representing 278 individuals (Table 2). VNT was the only test with insufficient studies for any meta-analyses and 2 out of 3 studies were from the American region. For all other tests, studies included in analyses reported data from at least

five out of six WHO regions. Overall, just under half (137/285: 48%) of all included studies were from the South-East Asia region and 91% (124/137) of these studies were from India.

Only 17% (48/285) of studies reported data separately for children and adults (13% for NS1 ELISA, 14% for RT-PCR, 67% for VNT, 17% for IgM ELISA and 24% for IgG ELISA). Similarly, only 19% of studies (53/285) reported data separately for inpatients and outpatients (20% for NS1 ELISA, 15% for RT-PCR, 0% VNT, 20% for IgM ELISA and 22% for IgG ELISA (Table 2). NS1 ELISA, RT-PCR and IgM ELISA had greater than four studies in each DPO subgroup. IgG ELISA had less than four studies in the 0 to 4 days DPO subgroup.

Three studies included with IgG ELISA being the test of interest reported a two-by-two table where the definition of a positive test included the comparison of an acute and convalescent sample. No other studies clearly defined a positive IgG ELISA result in terms of sample collection time frames. This resulted in insufficient data to conduct a meta-analysis of paired IgG samples. Our analysis therefore only focused on the diagnostic accuracy of single tests.

Table 2: Characteristics of studies included in analyses by test of interest. In this table, 'n' refers to a study, not an article. WHO regions: African region (AFR), region of the Americas (AMR), European region (EUR), Eastern Mediterranean region (EMR), Southeast Asian region (SEAR), West Pacific region (WPR), Citations for the commercial tests can be found in Supplement [a]The three most common brand or techniques are displayed [b]Inter-quartile range

Test of interest (Number of studies)	**Brand or technique[a] n (%)**	**Study size (median (IQR[b] range))**	**Cohort n (%)**	**Outpatient /Inpatient n (%)**	**Number of studies by WHO region n (%)**	**DPO subgroup (Number of studies; Number of individuals)**
NS1 ELISA (60)	Panbio 22 (37%); Platelia 13 (22%); J-mitra 10 (17%)	193(96-533)	Adults 4 (7%); Children 4 (7%); Both 27 (45%);	Outpatients 7 (15%); Inpatients 5 (8%); Both 6 (10%);	AFR 0 (0%) AMR 9 (15%); EUR 1 (2%); EMR	0 to 4 days (10 ;1814); 1 to 7 days (16; 7949); All acute symptomat

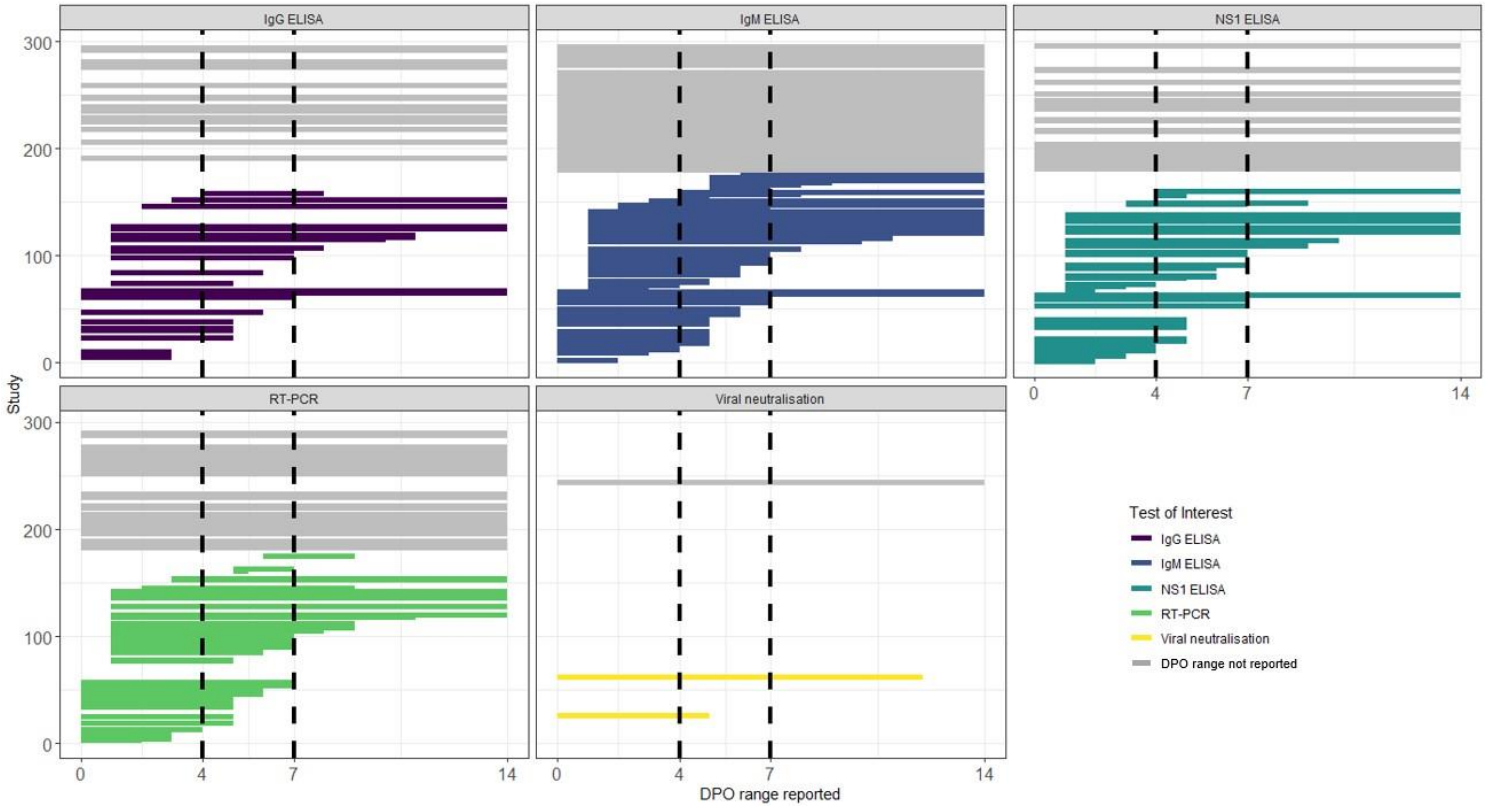
			ot stated 25 (42%)	Not stated 42 (70%)	3 (5%); SEAR 38 (63%); WPR 7 (12%); Multiple 2 (0%)	ic (60;31084)
RT-PCR (78)	Real-time 22 (28%) Conventional 2 (3%) Not stated 54 (69%)	183(98-307)	Adults 4 (5%); Children 7 (9%); Both 36 (46%); Not stated 31 (40%)	Outpatients 5 (6%); Inpatients 7 (9%); Both 13 (17%); Not stated 53 (68%)	AFR 4 (5%) AMR 23 (29%); EUR 1 (1%); EMR 2 (3%); SEAR 29 (37%); WPR 17 (22%) Multiple 2 (0%)	0 to 4 days (4;622); 1 to 7 days (27;7519); All acute symptomatic (78; 21402)
VNT (3)	PRNT 1 (33%) FRNT 1 (33%); ELISPOT-MNT 1 (33%)	41(39-121)	Adults 1 (33%); Children 1 (33%); Both 0 (0%); Not stated 1 (33%)	Outpatients 0 (0%); Inpatients 0 (0%); Both 0 (0%); Not stated 3 (100%)	AFR 0 (0%) AMR 2 (67%); EUR 0 (0%); EMR 0 (0%); SEAR 0 (0%); WPR 0 (0%) Multiple 1 (33%)	0 to 4 days (0; 0); 1 to 7 days (1; 41); All acute symptomatic (3; 278)

IgM ELISA (107)	Panbio 34 (32%) NIV 24 (22%) J-mitra 8 (7%)	164(67-461)	Adults 9 (8%); Children 9 (8%); Both 51 (48%); Not stated 38 (36%)	Outpatients 9 (8%); Inpatients 12 (11%); Both 21 (20%); Not stated 65 (61%)	AFR 5 (5%) AMR 21 (20%); EUR 2 (2%); EMR 4 (4%); SEAR 57 (53%); WPR 14 (13%) Multiple 5 (5%)	0 to 4 days (7;1209); 1 to 7 days (35; 20057); All acute symptomatic (107; 69488)
IgG ELISA (37)	Panbio 14 (38%); Focus 4 (11%); Standard Diagnostics 3 (8%)	158(66-325)	Adults 6 (16%); Children 3 (8%); Both 12 (32%); Not stated 16 (43%)	Outpatients 5 (14%); Inpatients 3 (8%); Both 8 (22%); Not stated 21 (57%)	AFR 2 (5%) AMR 7 (19%); EUR 1 (3%); EMR 1 (3%); SEAR 13 (35%); WPR 10 (27%) Multiple 3 (8%)	0 to 4 days (2; 1156); 1 to 7 days (12; 8642); All acute symptomatic (37;15251)

All (285)	N/A	176 (80-376)	Adults 24 (8%); Children 24 (8%); Both 127 (45%); Not stated 110 (39%)	Outpatients 26 (9%); Inpatients 27 (9%); Both 48 (17%); Not stated 184 (65%)	AFR 11 (4%) AMR 62 (22%); EUR 5 (2%); EMR 10 (4%); SEAR 137 (48%);	0 to 4 days (23; 4801); 1 to 7 days (91;44,208); All acute symptomatic (285;137,4
					WPR 48 (17%) Multiple 11 (4%)	87)

Extracted data showed high variability regarding the timing of dengue diagnostic testing post onset of fever (Figure 2). 41% (116/285) of included studies did not report the timeframe of testing DPO other than stating that testing was carried out on symptomatic individuals for acute dengue. This proportion was similar across tests of interest with 37%, 47%, 33%, 39% and 41% for NS1 ELISA, RT-PCR, viral, IgM ELISA and IgG ELISA respectively.

Figure 2. Extracted data from studies included in analyses by reported days-post-onset (DPO) of fever diagnostic testing and by test of interest. DPO subgroups are: 0 to 4 days, 1 to 7 days and all acute symptomatic, which includes studies that did not report a DPO range but did state that testing occurred on symptomatic individuals for acute dengue (shown as as grey bars spanning 0 to 14 days DPO). Vertical dashed lines indicate DPO subgroup upper limits.



Results from meta-analyses

The estimated sensitivity and specificity from our analysis for each test of interest is shown in Figure 3. Subgroup analyses of results reported at 0 to 4 days DPO compared with 1 to 7 days DPO and during the 'all acute symptomatic' group showed that both RT-PCR and NS1 ELISA were less sensitive during longer or unspecified time periods after symptom onset; yet the overall point estimates for sensitivity and specificity of RT-PCR were above 85% and 89%, respectively. Similarly, NS1 ELISA point estimates for sensitivity remained above 85% irrespective of DPO subgroup, with specificity above 90%.

Subgroup analyses of results reported at 1 to 7 days DPO compared with 0 to 4 days DPO demonstrated a marked difference in IgM ELISA sensitivity – at 71% (95% CrI 57-84%) and 17% (95% CrI 3-51%) respectively (Figure 3). Comparing the IgM ELISA primary analysis model to the model containing all acute symptomatic data, sensitivity and specificity estimates were lower at 62% (95% CrI 45-75%) and 85% (95% CrI 76-91%) respectively. For the IgG ELISA comparison of the 1 to 7 DPO with the 'all acute symptomatic' subgroup, the point estimate for sensitivity was 10 percentage points lower, at 74% (95% CrI 44-94%) and 64% (95% CrI 33-83%) respectively. Similarly, the specificity point estimate was 93% (95% CrI 77-

98%) at 1 to 7 DPO compared with 86% (95% CrI 62-94%) in the 'all acute symptomatic' group.

Individual study estimates for both sensitivity and specificity for each test of interest in each DPO subgroup are presented in supplement 4.

Overall, heterogeneity was seen across all meta-analyses shown by wider credible intervals for the predicted estimates when compared to the pooled estimates. For all tests of interest, the model with the narrowest credible interval for the predicted estimates was the primary analysis model and the widest credible interval from the 'all acute symptomatic' model, representing the widest DPO range.

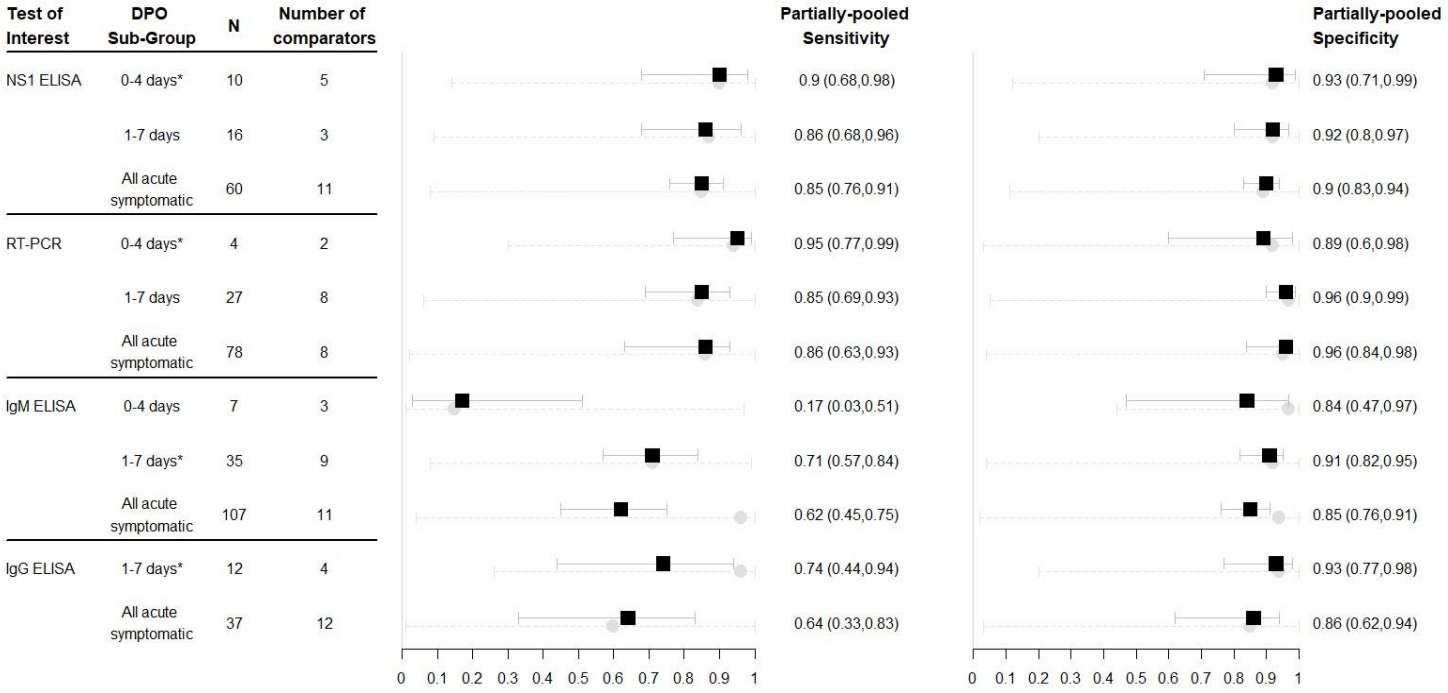


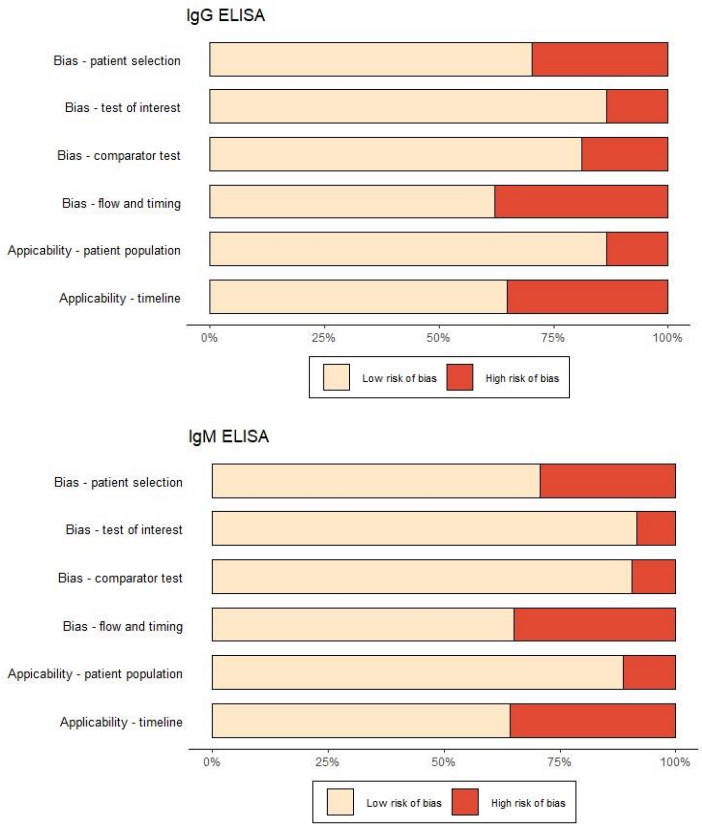
Figure 3: Summary forest plot of pooled (black - solid line) and predicted (grey dashed-line) estimates of sensitivity and specificity from separate meta-analyses for each test of interest. Stratified by DPO subgroup. *Indicates primary analysis subgroup for each test of interest.

Methodological quality of included studies

Figure 4 summarises the risk of bias and applicability concerns for studies from each test of interest with modified criteria of assessment (domains). 44% (n=124) of all included studies were judged overall as having a low risk of bias and low concern regarding applicability. This

varied by test of interest with 43% (n=26), 37% (n=29), 49% (n=52) and 43% (n=16) for NS1 ELISA, RT-PCR, IgM ELISA and IgG ELISA included studies respectively. Of all domains of assessment, the risk of bias from the index test and comparator test were deemed the lowest risk. Over 25% of included studies had a high risk of bias from the flow and timing domain, primarily due to studies that performed the test in the acute period with no specific DPO range stated.

The majority of included studies demonstrated low concerns of applicability for the timeline used. A meta-analysis for the 'all acute symptomatic' subgroup that includes only those studies with a low risk of bias and low risk of applicability concerns is presented in the supplementary material (Figure S6). The estimated sensitivity and specificity of each test were not dissimilar from those results presented in Figure 3, including all studies.



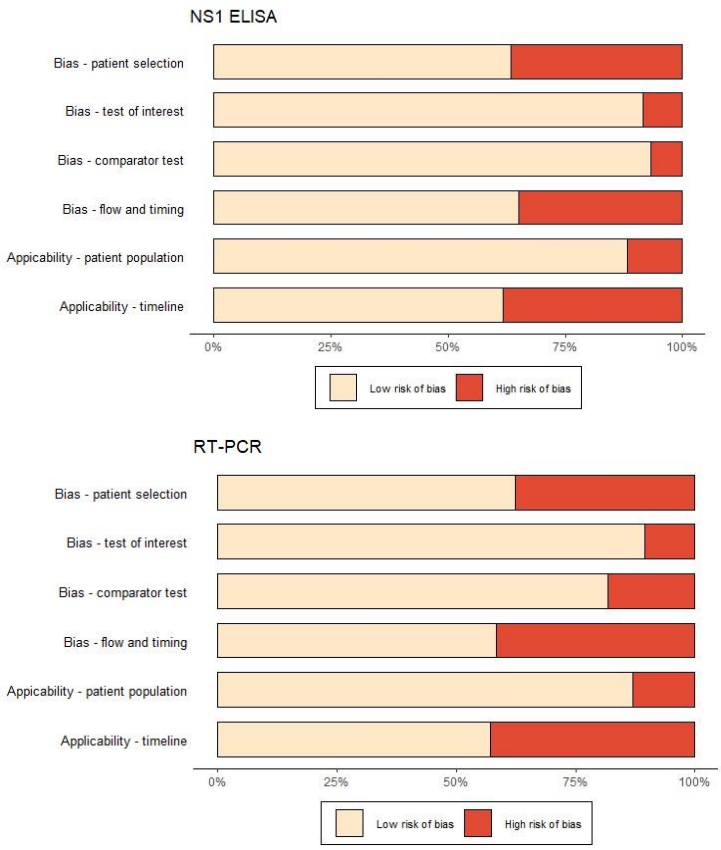


***Figure 4**: Summary of the risk of bias and applicability concerns across six domains in the accuracy of dengue diagnostics for the detection of acute dengue.*

Discussion

We report, to our knowledge, the most comprehensive review of dengue diagnostic accuracy to date. The estimated sensitivity and specificity of NS1 ELISA from our primary analysis (0 to 4 DPO) was 90% (95% CrI 68-98%) and 93% (95% CrI 71-99%) respectively, which is more sensitive and less specific than has previously been reported (24). Generally, previous analyses of NS1 ELISA accuracy have used RT-PCR as a 'gold-standard' comparator test (25,26), making the assumption that RT-PCR is 100% sensitive and 100% specific. If this assumption is incorrect, estimates of NS1 ELISA diagnostic accuracy will have been underestimated in these analyses (assuming conditional independence). One study compared NS1 ELISA and RT-PCR using a third diagnostic method as a 'gold standard' comparator and found that the two tests had similar diagnostic accuracy, reflecting our findings. In comparison to the 0 to 4 DPO subgroup, where RT-PCR has a better sensitivity than NS1 ELISA, In the 1 to 7 DPO subgroup, NS1 ELISA was found to have a similar sensitivity (86%, 95% CrI 6896%)

to RT-PCR (85%, 95% CrI 68-93%). This reflects the greater longevity of NS1 antigen compared to DENV RNA. These results call for increased recognition of the utility of NS1 ELISA as a more accessible alternative to RT-PCR, particularly given poor patient and clinician reporting of illness duration, which has a substantial impact on RT-PCR diagnostic test accuracy.

IgM ELISA was found to be 17% sensitive and 84% specific in the 0 to 4 DPO range, reflecting the recognised lag in DENV IgM antibody titres in response to infection (27,28). These results reiterate that IgM ELISA as a single test should not be used to confirm dengue infection in this timeframe. The current WHO outbreak recommendation states that IgM ELISA positivity in a single acute sample is highly suspicious of dengue infection rather than confirmatory (4). Our results suggest that IgM ELISA in a single acute sample must be interpreted with DPO in mind given the increase in sensitivity after 4 DPO. The pooled sensitivity and specificity of IgM ELISA in the 1 to 7 DPO subgroup was 71% (95% CrI 57 - 85%) and 91% (95% CrI 82 - 95%) respectively, reflecting increasing antibody titres over this timeframe. These results are likely to underestimate the actual diagnostic accuracy of IgM ELISA in the 5 to 7 DPO, given the possible inclusion of samples from less than 5 DPO. One previously reported estimate of IgM ELISA sensitivity and specificity was 38.1% and 100%, respectively. Notably, although these estimates were not stratified by DPO, the majority of patients included in the analysis (44/86) were in the 1 to 3 DPO range (25), which may explain the poor sensitivity reported.

Additionally, IgM titres are recognised to increase less in response to secondary dengue infection when compared to primary infection, further limiting the utility of this test particularly among patients living in dengue endemic regions and/or with a history of previous infection (27). While in practice IgM ELISA is often used as a single test in the acute phase, its accuracy may be increased by using paired samples, interpretation as a ratio with IgG ELISA or, using a single measurement in combination with other diagnostic methods as part of a diagnostic algorithm (9,28). Further meta-analyses to ascertain diagnostic accuracy of such algorithms are needed.

IgG ELISA is primarily used to distinguish primary from secondary dengue, and/or in paired samples for acute dengue. While this review aimed to assess the diagnostic accuracy of acute and convalescent paired samples, too few articles reported using paired samples to carry out a meta-analysis. The remaining included studies used single IgG measurement in the acute phase, which have limited utility in clinical practice, these results should therefore be

interpreted with caution and we do not recommend single IgG ELISA to be used in acute dengue diagnosis.

Strengths of this review include the comprehensive and robust search strategy, encapsulating five different dengue diagnostics, across eleven different bibliographic databases and incorporating publications in four different languages that included independent duplicate review and adherence to QUADAS-2 and PRISMA guidelines. This inclusive search strategy identified a large number of studies that collected data in 43 countries with a wide range of comparator tests. To ensure our results were not biased by assumptions regarding the comparator tests accuracy, a Bayesian analysis approach was used that did not assume any comparator was perfect. While this approach has previously been used for rapid tests for dengue (29), to our knowledge it has not previously been applied to a dataset this large for our tests of interest.

The studies included in the meta-analysis brought multiple limitations. Firstly, dengue diagnostics have been shown to have optimal performance at certain DPOs, depending on the diagnostic target (9). However, included studies reported overlapping DPO ranges for acute dengue from 1 DPO to 36 days DPO for all tests of interest (Figure 2). This limited our ability to stratify results into discrete DPO windows (i.e. 0 to 4, 5 to 7, 8 to 14 DPO), which would have strengthened interpretability of the results. This inconsistent reporting highlights a lack of consensus on timeframes for acute dengue diagnosis, which inhibits the ability to perform meta-analyses. Second, there was substantial heterogeneity of reporting of important variables that were extracted but could not be used for subgroup analyses due to insufficient data. Such variables include; DENV serotype, patient population (including clinical suspicion of dengue infection, which would affect the pre-test probability), primary or secondary infection, brand of test of interest and geographical location of study. DPO subgroup analysis, though adding depth to our conclusions, resulted in data paucity for some included windows, notably DPO 0 to 4 RT-PCR, which was based on only 4 included studies, and needs to be interpreted with caution. Third, in order to maximise the data available for analysis, we included any clinical study that met the inclusion criteria, rather than limiting inclusion to diagnostic accuracy studies. As a consequence, many studies poorly described various aspects of the test methodology and the population group studied, highlighted by the results of QUADAS-2, with potential impacts on the overall estimates of both sensitivity and specificity. Our meta-analytical approach also has limitations. While our analysis framework improves upon previous analyses by accounting for imperfect comparator tests, there are other model assumptions that have to be made that can still lead to biased estimates of sensitivity and specificity.

Notably, the choice of prior distributions for the comparator tests' diagnostic accuracy and the assumption of a two-state latent class model (30).

Our findings have the potential to contribute to policy development, with particular relevance to the utility of NS1 ELISA. Furthermore, our estimates can provide a basis for future work in dengue diagnostic accuracy assessment, for example in the evaluation of novel RDTs and diagnostic algorithms, using both Bayesian and traditional approaches.

Contributors

HH, EF and KP conceived the study. KP, SK, EF, LMS, ES, OT and CA assessed the eligibility of the studies, KP, SK and ES extracted the data, and KP and ES assessed the methodological quality of the included studies. JF developed the search strategy and conducted the literature search. SK carried out the statistical analysis. JB, RK and OB advised on the statistical analysis. KP and SK prepared the original draft of the manuscript, with considerable input from EF, HH, and AD. All authors reviewed, edited and approved the final manuscript.

Acknowledgements

We wish to acknowledge the input of Professor Rosanna Peeling for her insightful contribution to the manuscript.

Declaration of interests

We declare no competing interests

Data Sharing

All model code can be found at: https://github.com/shk313/diagnostic-testmetaanalysis/tree/main/Dengue

Funding

SK is supported by the Medical Research Council London Intercollegiate Doctoral Training Partnership Studentship (MR/N013638/1). The funders had no role in study design, data collection and analysis, or preparation of the manuscript

Bibliography

1. Akbar R. Ten threats to global health in 2019 [Internet]. WHO; 2019. Available from: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019
2. WHO. Disease Outbreak News; Dengue – Global situation [Internet]. 2023. Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498
3. Zeng Z, Zhan J, Chen L, Chen H, Cheng S. Global, regional, and national dengue burden from 1990 to 2017: A systematic analysis based on the global burden of disease study 2017. eClinicalMedicine [Internet]. 2021 Feb 1 [cited 2023 Dec 15];32. Available from: https://doi.org/10.1016/j.eclinm.2020.100712
4. WHO. Dengue Outbreak Toolbox [Internet]. WHO; 2019. Available from: https://www.who.int/docs/default-source/outbreak-toolkit/updates-documents_july5/dengue-outbreak-data-collection-toolbox---inis-3-july-1.pdf?sfvrsn=ec3ffcf3_2#:~:text=WHO%20surveillance%20case%20definition,leukopenia%2C%20or%20any%20warning%20sign.
5. Guzman MG, Gubler DJ, Izquierdo A, Martinez E, Halstead SB. Dengue infection. Nat Rev Dis Primers. 2016 Aug 18;2:16055.
6. Halstead SB. Dengue Antibody-Dependent Enhancement: Knowns and Unknowns. Microbiol Spectr. 2014 Dec;2(6).
7. Land KJ, Boeras DI, Chen XS, Ramsay AR, Peeling RW. REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. Nat Microbiol. 2019 Jan;4(1):46–54.
8. WHO. Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition. 2009.
9. Raafat N, Blacksell SD, Maude RJ. A review of dengue diagnostics and implications for surveillance and control. Trans R Soc Trop Med Hyg. 2019 Nov 1;113(11):653–60.
10. Senanayake A M Kularatne. Dengue fever. BMJ. 2015 Sep 15;351:h4661.
11. de Vet H, Eisinga A, Riphagen I, Aertgeerts B, Pewsner D. Chapter 7: Searching for Studies in Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 0.4 [updated September 2008]. (The Cochrane Collaboration).
12. The EndNote Team. EndNote. Philadelphia, PA;
13. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environmental Evidence. 2018 Feb 1;7(1):8.
14. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529–36.
15. Cogan J. WHO: Dengue and Severe Dengue Fact Sheet [Internet]. WHO; 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue
16. Duyen HTL, Ngoc TV, Ha DT, Hang VTT, Kieu NTT, Young PR, et al. Kinetics of plasma viremia and soluble nonstructural protein 1 concentrations in dengue: differential effects according to serotype and immune status. J Infect Dis. 2011 May 1;203(9):1292–300.
17. Erra EO, Korhonen EM, Voutilainen L, Huhtamo E, Vapalahti O, Kantele A. Dengue in travelers: kinetics of viremia and NS1 antigenemia and their associations with clinical parameters. PLoS One. 2013;8(6):e65900.

18. Tricou V, Minh NN, Farrar J, Tran HT, Simmons CP. Kinetics of viremia and NS1 antigenemia are shaped by immune status and virus serotype in adults with dengue. PLoS Negl Trop Dis. 2011 Sep;5(9):e1309.
19. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics. 2012 Dec;68(4):1285–93.
20. Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C,. Chapter 10: Undertaking meta-analysis. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 [updated July 2023]. https://training.cochrane.org/handbookdiagnostic-test-accuracy/current; (The Cochrane Collaboration).
21. Takwoingi Y, Dendukuri N, Schiller I, Rücker G, Jones HE, Partlett C,. Chapter 11: Presenting findings in Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 [updated July 2023]. https://training.cochrane.org/handbookdiagnostic-test-accuracy/current; (The Cochrane Collaboration).
22. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2022. Available from: https://www.R-project.org/.
23. Stan Development Team. Stan modeling language user's guide and reference manual [Internet]. 2017. Available from: http://mc-stan.org/users/documentation/
24. Costa VGD, Marques-Silva AC, Moreli ML. A Meta-Analysis of the Diagnostic Accuracy of Two Commercial NS1 Antigen ELISA Tests for Early Dengue Virus Detection. PLoS ONE. 2014;9(4):e94655.
25. Luvira V, Thawornkuno C, Lawpoolsri S, Thippornchai N, Duangdee C, Ngamprasertchai T, et al. Diagnostic Performance of Dengue NS1 and Antibodies by Serum Concentration Technique. Tropical Medicine and Infectious Disease. 2023;8(2):117.
26. Iqbal G, Javed H, Raza FA, Gohar UF, Fatima W, Khurshid M. Diagnosis of Acute Dengue Virus Infection Using Enzyme-Linked Immunosorbent Assay and Real-Time PCR. Canadian Journal of Infectious Diseases and Medical Microbiology. 2023;2023:1–7.
27. Chanama S, Anantapreecha S, A-nuegoonpipat A, Sa-gnasang A, Kurane I, Sawanpanyalert P. Analysis of specific IgM responses in secondary dengue virus infections: levels and positive rates in comparison with primary infections. J Clin Virol. 2004 Nov;31(3):185–9.
28. Hunsperger EA, Muñoz-Jordán J, Beltran M, Colón C, Carrión J, Vazquez J, et al. Performance of Dengue Diagnostic Tests in a Single-Specimen Diagnostic Algorithm. J Infect Dis. 2016 Sep 15;214(6):836–44.
29. Pan-ngum W, Blacksell SD, Lubell Y, Pukrittayakamee S, Bailey MS, de Silva HJ, et al. Estimating the true accuracy of diagnostic tests for dengue infection using bayesian latent class models. PLoS One. 2013;8(1):e50765.
30. Schofield MR, Maze MJ, Crump JA, Rubach MP, Galloway R, Sharples KJ. On the robustness of latent class models for diagnostic testing with no gold standard. Stat Med. 2021 Sep 30;40(22):4751–63.

Appendix E

Example Stan code for a single-pathogen model

```
    // Example Stan program for the malaria single-
       pathogen model

data {
  int<lower=0> N; // Number of individuals (cases
     and controls)
  int<lower=0> Nstrata;  // Number of strata to
     consider for regression
  int<lower=0> N_Cases; // Number of cases only
  int<lower=0> J; // Number of possible causes
  int<lower=0> M; // Number of diangostic tests
  vector[N] I1; // Indicator for whether an
     individual has a result for diagnostic test 1
  vector[N] I2; // Indicator for whether an
     individual has a result for diagnostic test 2
  int t1[N]; // Observed diagnostic test results for
      test 1
  int t2[N]; // Observed diagnostic test results for
      test 2
  int Strata[N]; // Indicator for wich strata does
     each individual belong to
}

parameters {
  // Specificty is allowed to vary by strata and is
     between 0 and 1
  real<lower=0,upper=1.0> Sp1[Nstrata];
```

```
  real<lower=0,upper=1.0> Sp2[Nstrata];

   // Sensitivity is assumed to be greater than 1-Sp
      , does not vary by site and is between 0 and 1
  real<lower=inv_logit(logit(1-min(Sp1))),upper=1.0>
      Se1;
  real<lower=inv_logit(logit(1-min(Sp2))),upper=1.0>
      Se2;

 // Aetiology is a simplex
  simplex[J] aeti[Nstrata];// simplex of length
     number of causes

  // Parameters for random effect
  vector[N] REp; // e.g. infection intensity
  real<lower=0> b;

}

transformed parameters {
  /// Probability of a positive test for each test
     and each individual
  vector<lower=0,upper=1>[N] prob[2,M];

  for(n in 1:N){
    prob[1,1,n] = 1-Sp1[Strata[n]];
    // equation from Dendukuri et al. []
    prob[2,1,n] = inv_logit(logit(Se1) + REp[n]*b);
    prob[1,2,n] = 1-Sp2[Strata[n]];
    prob[2,2,n] = inv_logit(logit(Se2) + REp[n]*b);
  }
}

model {
  real lp[J]; // 2d vector with each entry
     corresponding
  real lp2;// controls
  // to the un-normalised posterior prob for each
     cause
```

```
// Priors
// Aetiology prior (pi)
aeti ~ dirichlet(rep_vector(1.0,J));

//Sensitivity priors
// RDT 65.2-89.0% with 95% probability
Se1~beta(34.68921,9.609493);
// Microscopy
Se2~beta(1,1);

//Specificity priors
// RDT estimated from controls so 0-100%
Sp1~beta(1,1);
// Microscopy estimated from controls in Laos,
   Mozambique and Zimbabwe
// Informative prior assuming specificity between
   95 and 100% in Malawi
Sp2[1]~beta(1,1);
Sp2[2]~beta(50,0.5);
Sp2[3]~beta(50,0.5);
Sp2[4]~beta(1,1);
Sp2[5]~beta(1,1);
Sp2[6]~beta(1,1);
Sp2[7]~beta(1,1);
Sp2[8]~beta(1,1);

// Random-effect to account for conditional
   dependence
REp~normal(0,1);
b~gamma(1,1);

for(n in 1:N_Cases){
   for(j in 1:2){
     lp[j] = log(aeti[Strata[n], j]);
   }
     lp[1] = lp[1] + binomial_lpmf(t1[n] | 1, prob
        [2,1,n])*I1[n] + binomial_lpmf(t2[n] | 1,
        prob[2,2,n])*I2[n];
     lp[2] = lp[2] + binomial_lpmf(t1[n] | 1, prob
        [1,1,n])*I1[n] + binomial_lpmf(t2[n] | 1,
```

```
            prob[1,2,n])*I2[n];

    target += log_sum_exp(lp);
  }
    // Controls (only need to consider the false
       positivity rate stored in prob[1,,])
    for(n in (N_Cases+1):N){

        lp2 = binomial_lpmf(t1[n]|1, prob[1,1,n])*I1
           [n] + binomial_lpmf(t2[n]|1, prob[1,2,n])
           *I2[n];

        target += (lp2);
}
}
```

Appendix F

Extended sensitivity analyses from single-pathogen models

Figure F.1: A comparison of the estimated percentage of fever cases due to respiratory pathogens in inpatients using a more informative or diffuse prior for diagnostic test sensitivity by site and age group.

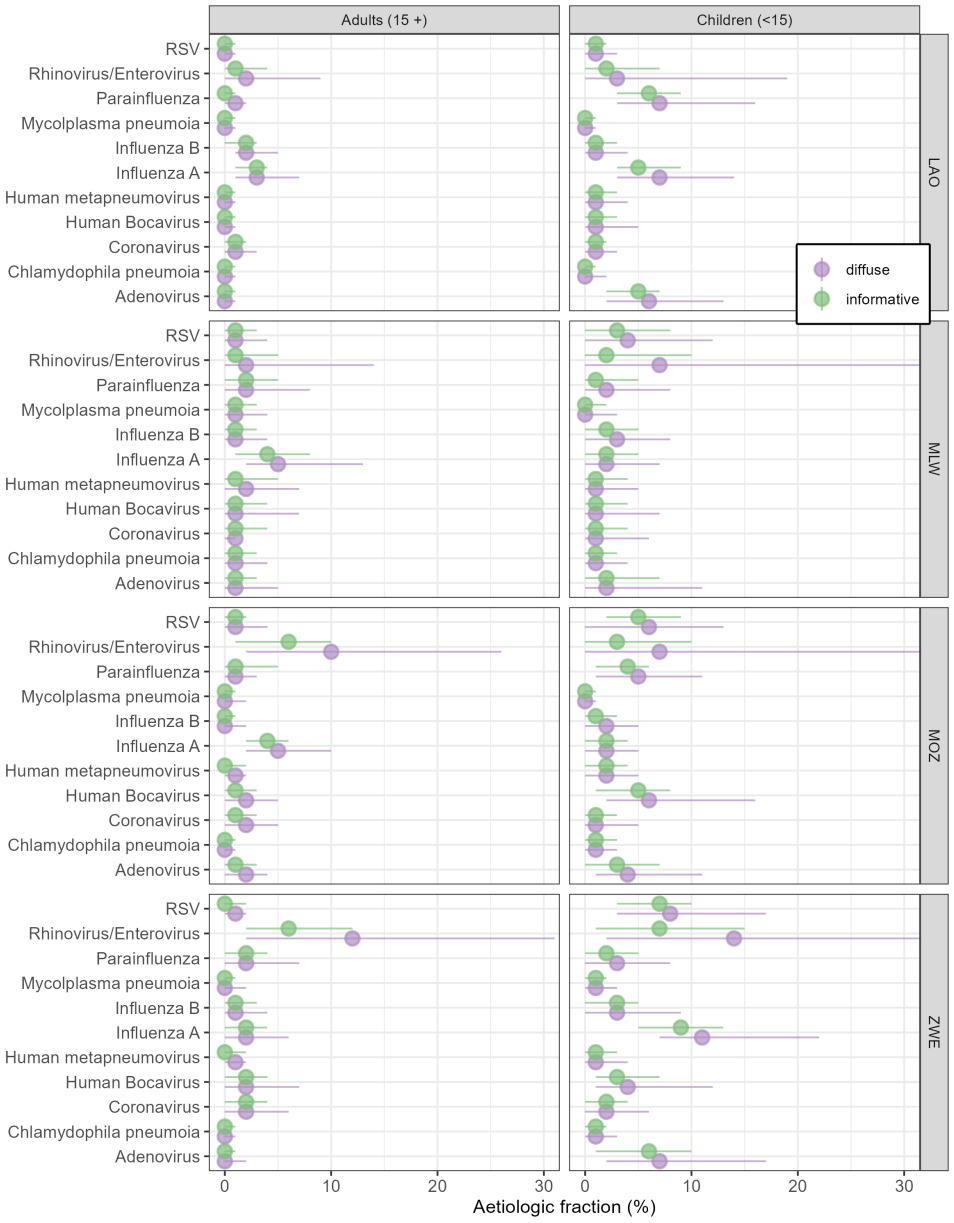
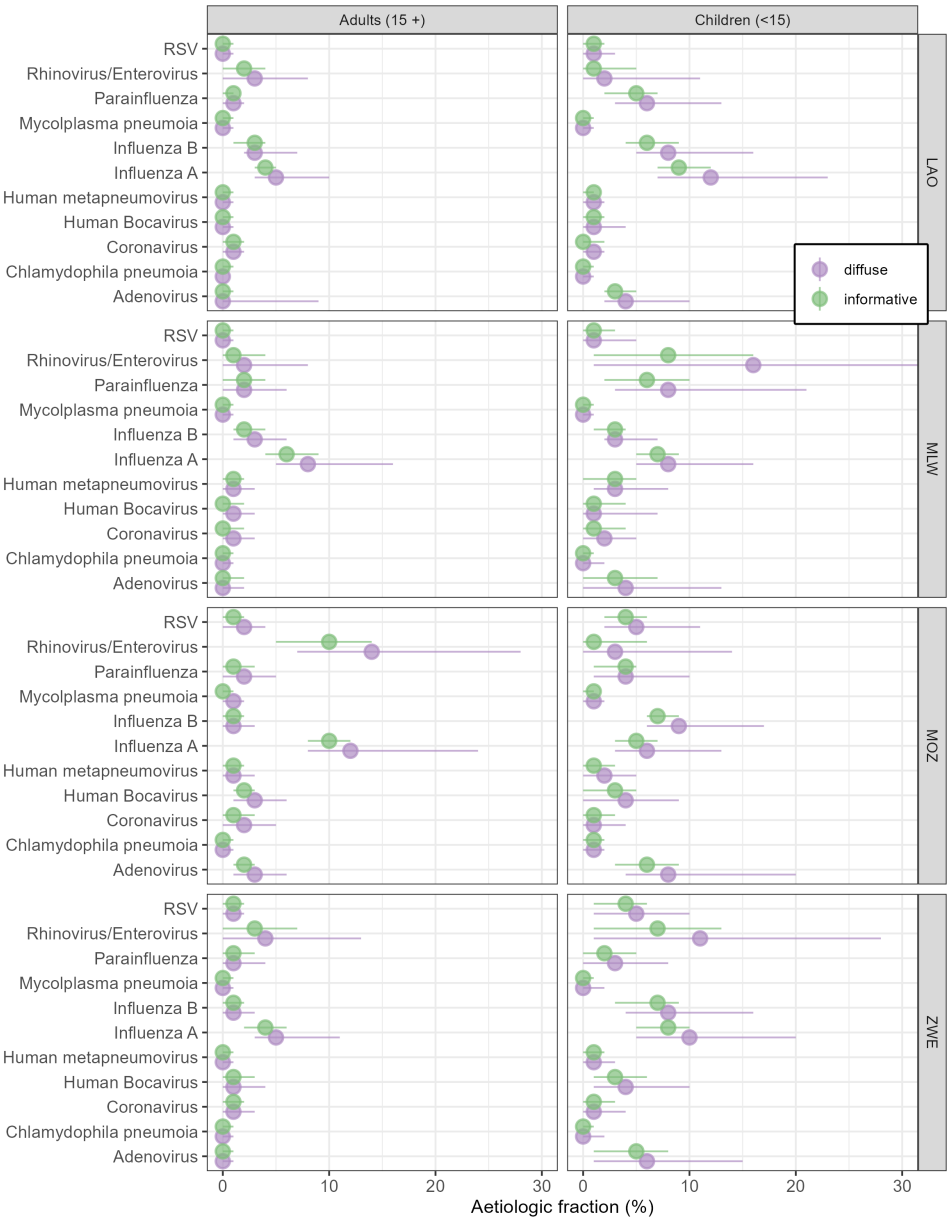


Figure F.2: A comparison of the estimated percentage of fever cases due to respiratory pathogens in all patients using a more informative or diffuse prior for diagnostic test sensitivity by site and age group.



Appendix G

Multi-pathogen model complete results tables

Table G.1: Estimated cause-specific case fractions (median and 95% credible interval) for Inpatients by site and age group

Pathogen	Laos		Malawi		Mozambique		Zimbabwe	
	under 15	15 & over	under 15	15 & over	under 15	15 & over	under 15	15 & over
Malaria	0 (0-1)	0 (0-1)	42 (34-51)	20 (13-27)	46 (42-51)	11 (8-14)	2 (1-4)	3 (2-5)
Influenza A	5 (3-8)	2 (1-4)	2 (0-5)	3 (1-7)	3 (1-5)	4 (2-6)	8 (5-12)	1 (0-3)
Influenza B	1 (0-3)	2 (1-3)	2 (0-5)	0 (0-2)	0 (0-2)	0 (0-1)	2 (0-5)	0 (0-2)
RSV	1 (0-2)	0 (0-1)	4 (1-8)	0 (0-2)	2 (1-5)	0 (0-1)	6 (3-10)	0 (0-1)
Adenovirus	4 (2-7)	0 (0-1)	2 (0-6)	1 (0-3)	1 (0-3)	1 (0-2)	2 (0-6)	0 (0-1)
Rhinovirus/Enterovirus	1 (0-5)	1 (0-4)	3 (0-10)	1 (0-4)	5 (1-9)	5 (2-9)	2 (0-8)	3 (0-8)
Human Bocavirus	0 (0-2)	0 (0-1)	1 (0-3)	1 (0-4)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)
Parainfluenza	5 (3-8)	0 (0-1)	2 (0-6)	1 (0-4)	1 (0-2)	1 (0-2)	1 (0-3)	1 (0-3)
Coronavirus	0 (0-2)	0 (0-2)	1 (0-3)	1 (0-3)	1 (0-2)	2 (0-3)	2 (0-4)	1 (0-3)
Human Metapneumovirus	1 (0-3)	0 (0-1)	1 (0-3)	1 (0-4)	0 (0-2)	0 (0-1)	1 (0-2)	0 (0-2)
Chlamydophila pneumonia	0 (0-2)	0 (0-1)	0 (0-3)	0 (0-2)	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)
Mycoplasma pneumonia	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)
Brucella	1 (0-3)	0 (0-1)	0 (0-2)	1 (0-3)	0 (0-1)	0 (0-1)	1 (0-3)	0 (0-1)
Dengue	9 (5-13)	15 (11-19)	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)
Zika	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Chikungunya	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Leptospirosis	0 (0-1)	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)
Burkholderia	0 (0-3)	2 (0-7)	1 (0-3)	1 (0-5)	0 (0-2)	1 (0-3)	1 (0-3)	0 (0-2)
Typhoidal *Salmonella*	2 (0-5)	2 (0-5)	2 (0-5)	4 (0-12)	1 (0-3)	1 (0-5)	4 (0-12)	3 (0-8)
Non-typhoidal *Salmonella*	0 (0-2)	0 (0-1)	2 (0-5)	1 (0-4)	0 (0-2)	1 (0-5)	0 (0-2)	0 (0-2)
Staphylococcus aureus	1 (0-4)	0 (0-2)	1 (0-5)	1 (0-4)	6 (3-13)	2 (0-8)	1 (0-5)	1 (0-3)
Escherichia coli	0 (0-2)	1 (0-5)	0 (0-3)	2 (0-8)	1 (0-3)	5 (0-13)	0 (0-3)	1 (0-5)
Other Enterobacterales	0 (0-2)	1 (0-3)	0 (0-3)	1 (0-4)	1 (0-3)	1 (0-4)	0 (0-3)	1 (0-3)
Klebsiella pneumoniae	0 (0-2)	1 (0-2)	0 (0-3)	1 (0-6)	1 (0-3)	2 (0-7)	1 (0-4)	1 (0-3)
Streptococcus pneumoniae	1 (0-3)	0 (0-2)	1 (0-4)	1 (0-5)	6 (2-11)	1 (0-3)	1 (0-3)	1 (0-3)
NOS	60 (52-68)	67 (60-73)	24 (12-37)	49 (36-61)	18 (8-27)	57 (44-66)	57 (46-66)	76 (68-83)

Note. NOS, not specified. RSV includes both RSV A or RSV B; Parainfluenza includes parainfluenza viruses 1:4; Coronavirus includes 229e, hku1, nl63, oc43.

Table G.2: Estimated cause-specific case fractions (median and 95% credible interval) for outpatients by site and age group

Pathogen	Laos		Malawi		Mozambique		Zimbabwe	
	under 15	15 & over	under 15	15 & over	under 15	15 & over	under 15	15 & over
Malaria	0 (0-1)	0 (0-1)	42 (34-50)	31 (25-39)	5 (4-7)	9 (6-12)	0 (0-2)	1 (0-2)
Influenza A	13 (10-18)	5 (3-7)	8 (6-11)	7 (5-10)	8 (6-11)	14 (11-18)	6 (4-9)	7 (4-9)
Influenza B	13 (9-17)	4 (2-6)	4 (2-5)	3 (2-5)	12 (9-15)	2 (1-4)	9 (6-12)	1 (0-2)
RSV	2 (0-4)	0 (0-1)	1 (0-3)	0 (0-1)	4 (2-6)	1 (0-3)	1 (0-3)	1 (0-2)
Adenovirus	1 (0-4)	0 (0-1)	4 (1-7)	0 (0-1)	8 (6-12)	1 (0-3)	3 (0-5)	0 (0-1)
Rhinovirus/enterovirus	6 (1-10)	4 (1-7)	13 (7-20)	2 (0-5)	6 (1-10)	5 (1-10)	4 (0-9)	1 (0-3)
Human Bocavirus	0 (0-2)	0 (0-1)	0 (0-2)	0 (0-2)	1 (0-2)	1 (0-3)	1 (0-4)	0 (0-1)
Parainfluenza	4 (2-7)	0 (0-1)	9 (6-14)	3 (1-5)	4 (3-6)	1 (0-2)	1 (0-3)	0 (0-1)
Coronavirus	1 (0-2)	1 (0-2)	2 (0-4)	1 (0-2)	2 (0-4)	1 (0-3)	1 (0-2)	0 (0-1)
Human Metapneumovirus	0 (0-1)	0 (0-1)	2 (1-5)	1 (0-2)	2 (1-4)	1 (0-3)	1 (0-2)	0 (0-1)
Chlamydophila pneumonia	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Mycoplasma pneumonia	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-1)	0 (0-1)
Brucella	0 (0-1)	0 (0-3)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)
Dengue	9 (6-14)	22 (18-26)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Zika	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Chikungunya	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Leptospirosis	1 (0-2)	1 (0-2)	0 (0-1)	1 (0-2)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)
Burkholderia	1 (0-4)	3 (0-7)	0 (0-2)	0 (0-2)	0 (0-2)	0 (0-2)	0 (0-2)	0 (0-2)
Typhoidal *Salmonella*	1 (0-4)	1 (0-3)	0 (0-2)	2 (0-5)	0 (0-2)	0 (0-2)	26 (16-38)	58 (38-74)
Non-typhoidal *Salmonella*	0 (0-2)	0 (0-1)	1 (0-3)	1 (0-2)	0 (0-1)	1 (0-3)	0 (0-2)	0 (0-2)
Staphylococcus aureus	1 (0-4)	1 (0-3)	0 (0-2)	0 (0-2)	1 (0-3)	1 (0-2)	0 (0-2)	0 (0-2)
Escherichia coli	0 (0-3)	1 (0-3)	0 (0-2)	1 (0-3)	1 (0-3)	5 (1-12)	1 (0-2)	3 (0-7)
Other Enterobacterales	0 (0-3)	1 (0-2)	0 (0-2)	0 (0-1)	0 (0-2)	1 (0-2)	1 (0-3)	0 (0-2)
Klebsiella pneumoniae	0 (0-3)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-2)	0 (0-2)	1 (0-3)	0 (0-2)
Streptococcus pneumoniae	1 (0-4)	0 (0-2)	0 (0-2)	0 (0-2)	0 (0-2)	2 (0-5)	1 (0-3)	1 (0-3)
NOS	37 (28-46)	52 (45-58)	3 (0-12)	41 (32-50)	40 (33-47)	49 (39-57)	36 (22-49)	16 (1-37)

Note. NOS, not specified. RSV includes both RSV A or RSV B; Parainfluenza includes parainfluenza viruses 1:4; Coronavirus includes 229e, hku1, nl63, oc43

Table G.3: Estimated cause-specific case fractions (median and 95% credible interval) for all patients by site and age group

Pathogen	Laos		Malawi		Mozambique		Zimbabwe	
	under 15	15 & over	under 15	15 & over	under 15	15 & over	under 15	15 & over
Malaria	0 (0-1)	0 (0-0)	41 (35-46)	26 (21-30)	25 (22-28)	10 (7-12)	1 (0-2)	2 (1-3)
Influenza A	10 (7-12)	4 (3-5)	7 (5-9)	6 (4-8)	6 (5-8)	10 (8-12)	7 (5-10)	4 (2-6)
Influenza B	7 (5-9)	3 (2-4)	3 (2-5)	2 (1-4)	7 (5-8)	1 (0-2)	6 (4-9)	0 (0-1)
RSV	1 (0-3)	0 (0-1)	2 (1-4)	0 (0-1)	3 (1-4)	1 (0-2)	3 (1-5)	1 (0-1)
Adenovirus	3 (1-5)	0 (0-0)	4 (1-7)	0 (0-1)	5 (3-7)	1 (0-2)	1 (0-4)	0 (0-1)
Rhinovirus/enterovirus	3 (0-7)	3 (1-5)	12 (6-17)	1 (0-4)	6 (2-9)	5 (1-8)	3 (0-7)	1 (0-4)
Human Bocavirus	0 (0-1)	0 (0-0)	0(0-2)	0 (0-1)	0 (0-2)	1 (0-2)	1 (0-2)	0 (0-1)
Parainfluenza	5 (3-7)	0 (0-1)	8 (5-11)	2 (1-4)	2 (1-3)	0 (0-1)	0 (0-2)	0 (0-2)
Coronavirus	0 (0-2)	1 (0-1)	1 (0-3)	1 (0-2)	1 (0-3)	1 (0-3)	1 (0-3)	0 (0-2)
Human Metapneumovirus	1 (0-2)	0 (0-1)	1 (0-3)	1 (0-2)	1 (0-2)	1 (0-2)	0 (0-2)	0 (0-1)
Chlamydophila pneumonia	0 (0-1)	0 (0-0)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-0)	0 (0-1)	0 (0-0)
Mycoplasma pneumonia	0 (0-1)	0 (0-0)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Brucella	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Dengue	10 (7-13)	19 (16-22)	0 (0-0)	0 (0-1)	0 (0-0)	0 (0-0)	0 (0-1)	0 (0-0)
Zika	0 (0-1)	0 (0-0)	0 (0-0)	0 (0-1)	0 (0-0)	0 (0-0)	0 (0-1)	0 (0-0)
Chikungunya	0 (0-1)	0 (0-1)	0 (0-0)	0 (0-1)	0 (0-0)	0 (0-0)	0 (0-1)	0 (0-0)
Leptospirosis	0 (0-1)	1 (0-1)	0 (0-1)	1 (0-1)	0 (0-0)	0 (0-1)	0 (0-1)	0 (0-1)
Burkholderia	0 (0-2)	3 (0-6)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-1)
Typhoidal *Salmonella*	1 (0-4)	2 (0-3)	1 (0-2)	3 (0-6)	0 (0-1)	1 (0-2)	20 (11-30)	37 (19-55)
Non-typhoidal *Salmonella*	0 (0-1)	0 (0-1)	1 (0-3)	0 (0-2)	0 (0-1)	1 (0-3)	0 (0-1)	0 (0-1)
Staphylococcus aureus	1 (0-3)	1 (0-2)	1 (0-2)	0 (0-1)	3 (1-7)	1 (0-4)	0 (0-2)	0 (0-2)
Escherichia coli	0 (0-1)	1 (0-3)	0 (0-1)	1 (0-3)	1 (0-2)	6 (2-12)	0 (0-2)	2 (0-6)
Other Enterobacterales	0 (0-1)	0 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)	1 (0-2)	0 (0-2)	0 (0-1)
Klebsiella pneumoniae	0 (0-1)	0 (0-1)	0 (0-1)	0 (0-2)	0 (0-1)	1 (0-3)	1 (0-2)	0 (0-1)
Streptococcus pneumoniae	1 (0-2)	0 (0-1)	0 (0-2)	0 (0-1)	2 (1-5)	1 (0-3)	0 (0-2)	1 (0-2)
NOS	53 (46-59)	61 (55-66)	13 (6-21)	52 (45-58)	33 (27-39)	56 (48-63)	48 (35-57)	45 (24-61)

Note. NOS, not specified. RSV includes both RSV A or RSV B; Parainfluenza includes parainfluenza viruses 1:4; Coronavirus includes 229e, hku1, nl63, oc43

Table G.4: Estimated cause-specific case fractions (median and 95% credible interval) for HIV positive cases in Mozambique by In/out-patient status and age group

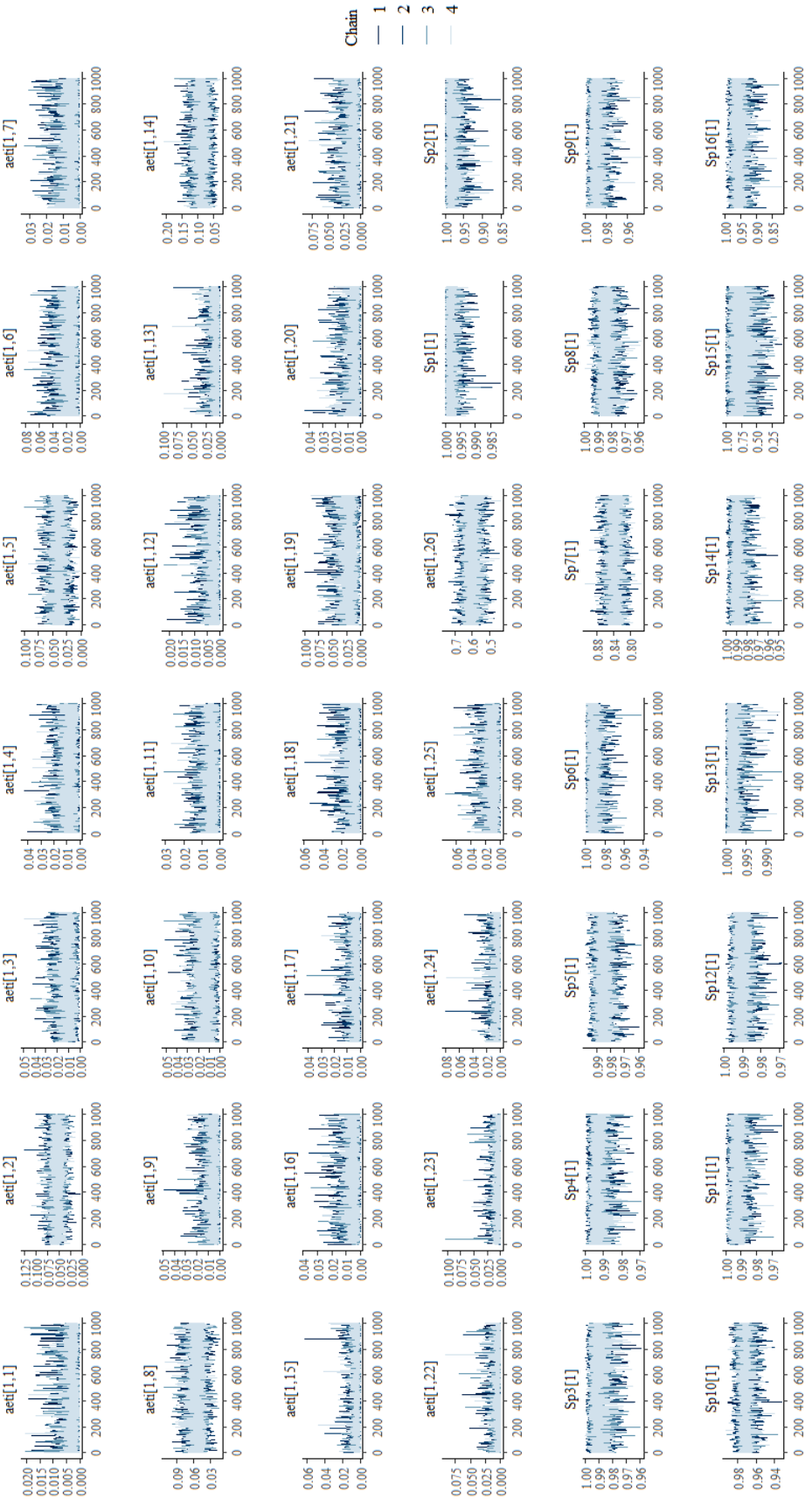
Pathogen	Inpatients		Outpatients	
	under 15	15 & over	under 15	15 & over
Malaria	19 (11-29)	6 (3-10)	5 (1-15)	8 (5-12)
Influenza A	2 (0-7)	4 (1-7)	2 (1-8)	10 (5-15)
Influenza B	2 (0-7)	0 (0-1)	11 (4-22)	1 (0-3)
RSV	4 (0-10)	1 (0-3)	2 (0-8)	1 (0-2)
Adenovirus	3 (0-10)	1 (0-3)	2 (0-8)	2 (1-5)
Rhinovirus/enterovirus	4 (0-15)	7 (2-12)	3 (0-14)	11 (5-17)
Human Bocavirus	2 (0-8)	2 (0-4)	2 (0-9)	1 (0-4)
Parainfluenza	2 (0-7)	1 (0-2)	2 (0-8)	2 (0-4)
Coronavirus	2 (0-7)	1 (0-3)	3 (0-11)	2 (0-5)
Human Metapneumovirus	2 (0-7)	0 (0-2)	2 (0-8)	1 (0-3)
Chlamydophila pneumonia	1 (0-5)	0 (0-1)	2 (0-8)	0 (0-2)
Mycoplasma pneumonia	1 (0-6)	0 (0-1)	2 (0-8)	1 (0-3)
Brucella	2 (0-9)	0 (0-2)	2 (0-10)	0 (0-2)
Dengue	0 (0-1)	0 (0-1)	2 (0-8)	0 (0-2)
Zika	1 (0-6)	0 (0-1)	2 (0-8)	0 (0-1)
Chikungunya	1 (0-6)	0 (0-2)	2 (0-8)	0 (0-2)
Leptospirosis	1 (0-5)	0 (0-1)	2 (0-8)	0 (0-1)
Burkholderia	2 (0-11)	1 (0-4)	4 (0-18)	1 (0-4)
Typhoidal *Salmonella*	2 (0-11)	1 (0-4)	4 (0-18)	1 (0-4)
Non-typhoidal *Salmonella*	3 (0-12)	2 (0-7)	3 (0-15)	2 (0-5)
Staphylococcus aureus	4 (0-16)	1 (0-5)	3 (0-14)	1 (0-4)
Escherichia coli	1 (0-5)	4 (0-12)	3 (0-14)	2 (0-7)
Other Enterobacterales	2 (0-9)	1 (0-4)	2 (0-13)	1 (0-3)
Klebsiella pneumonia	2 (0-9)	1 (0-5)	3 (0-14)	1 (0-3)
Streptococcus pneumonia	7 (0-22)	1 (0-5)	4 (0-17)	2 (0-7)
NOS	9 (13-27)	58 (46-69)	10 (1-28)	45 (35-55)

Note. NOS, not specified. RSV includes both RSV A or RSV B; Parainfluenza includes parainfluenza viruses 1:4; Coronavirus includes 229e, hku1, nl63, oc43

Appendix H

Multi-pathogen model convergence diagnostics

Figure H.1: Trace plots for the multi-pathogen inpatient model for all parameters related to children in Laos




Sp17[1]
Sp18[1]
Sp19[1]
Sp20[1]
Sp21[1]
Sp22[1]
Sp23[1]
Sp24[1]
Sp25[1]
Sp26[1]
Sp27[1]
Sp28[1]
Sp29[1]
Sp30[1]
Sp31[1]
Sp32[1]
Sp33[1]
Sp34[1]
Sp35[1]
Sp36[1]
Sp37[1]
Se1
Se2
Se3
Se4
Se5
Se6
Se7
Se8
Se9
Se10
Se11
Se12
Se13
Se14
Se15
Se16
Se17
Se18
Se19
Se20
Se21
Chain
1
2
3
4

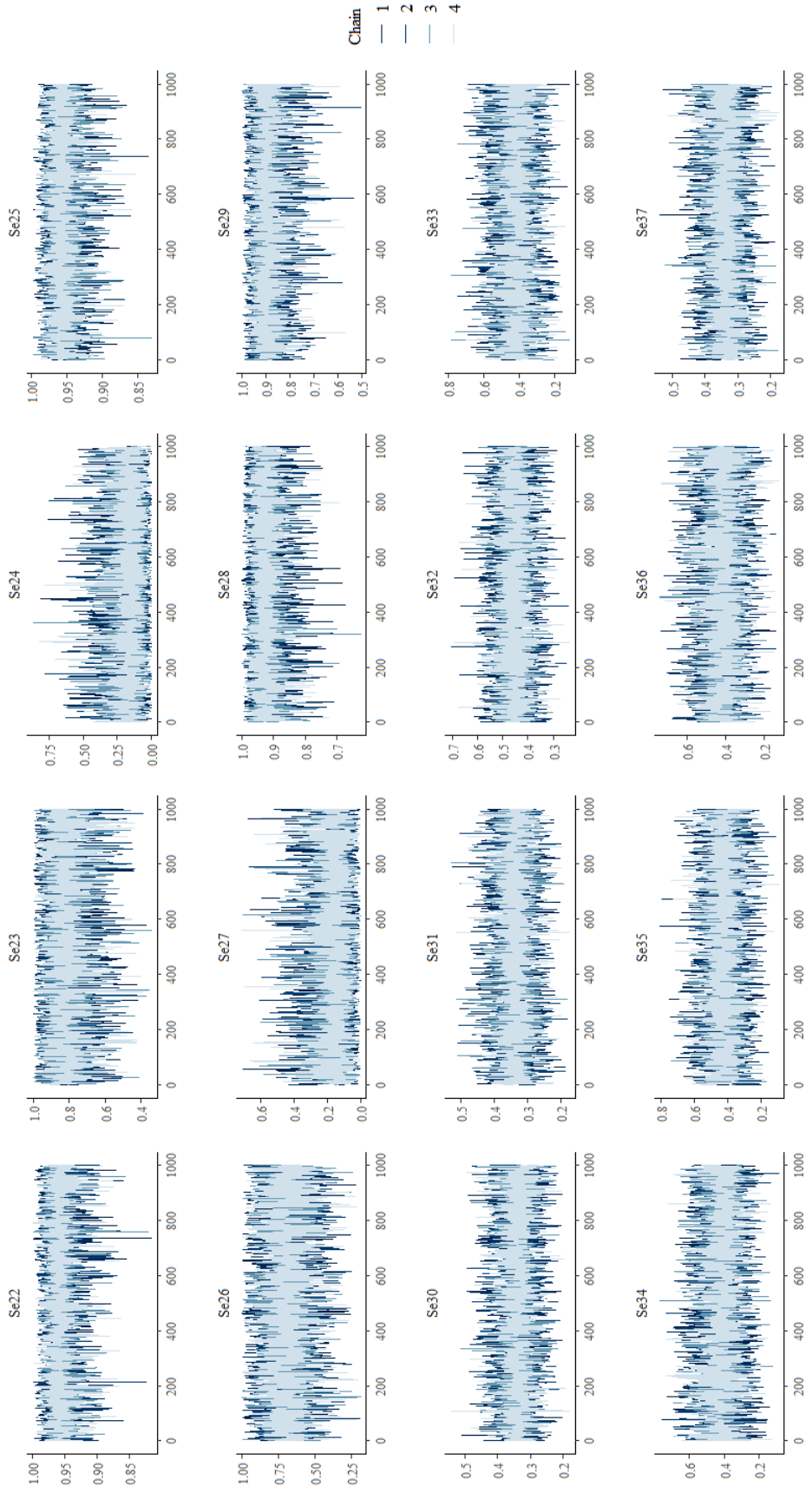
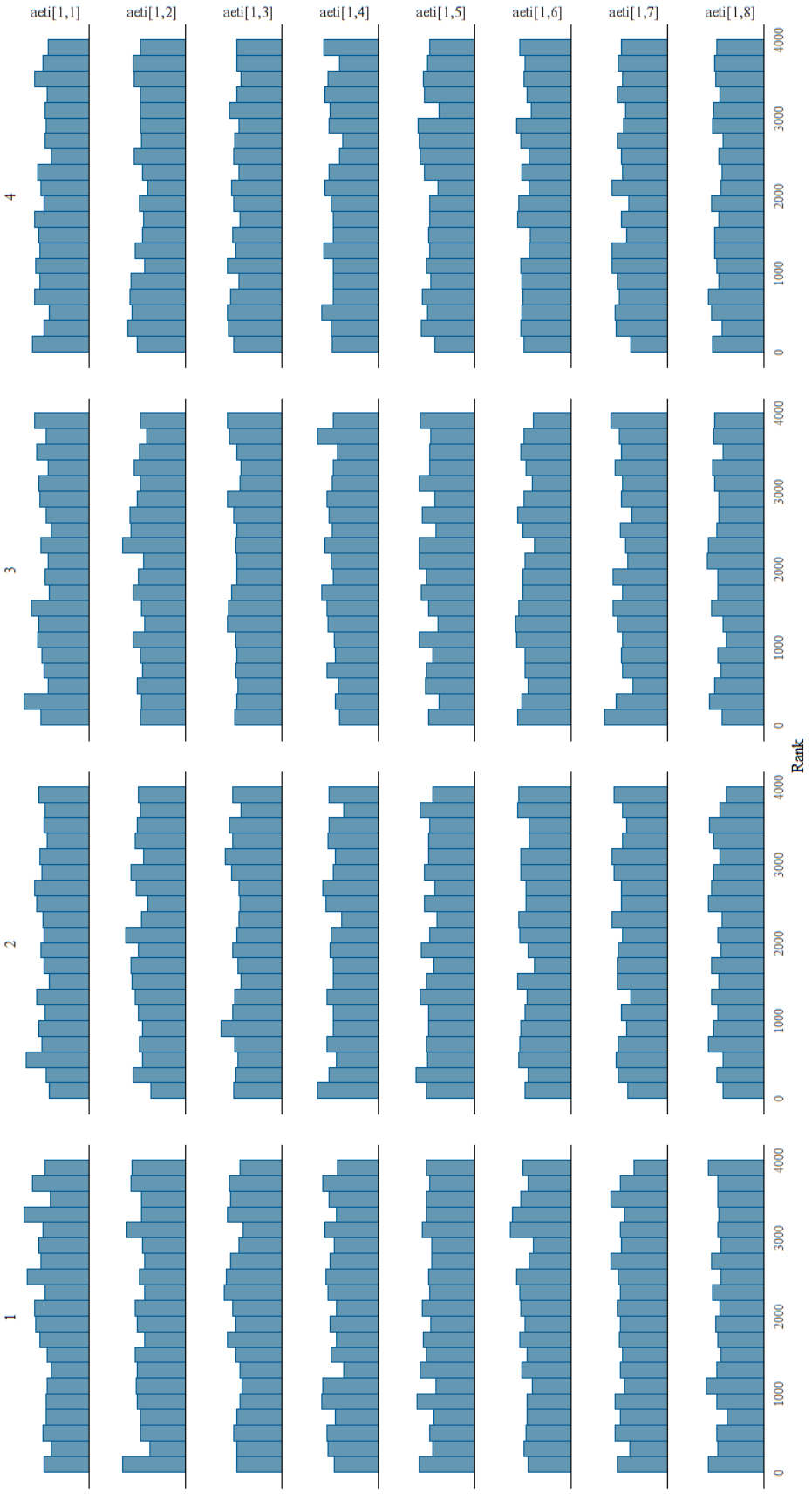
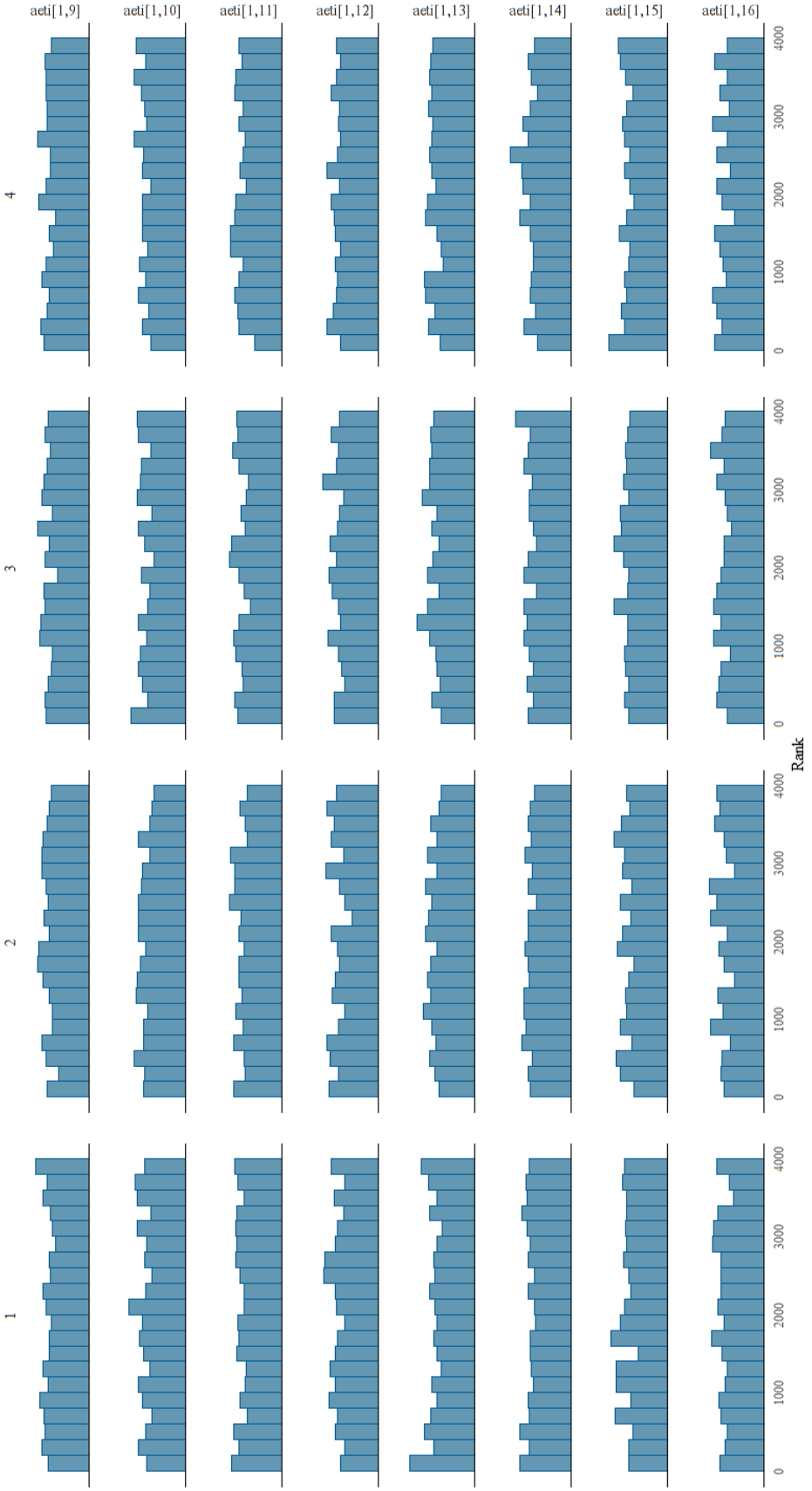
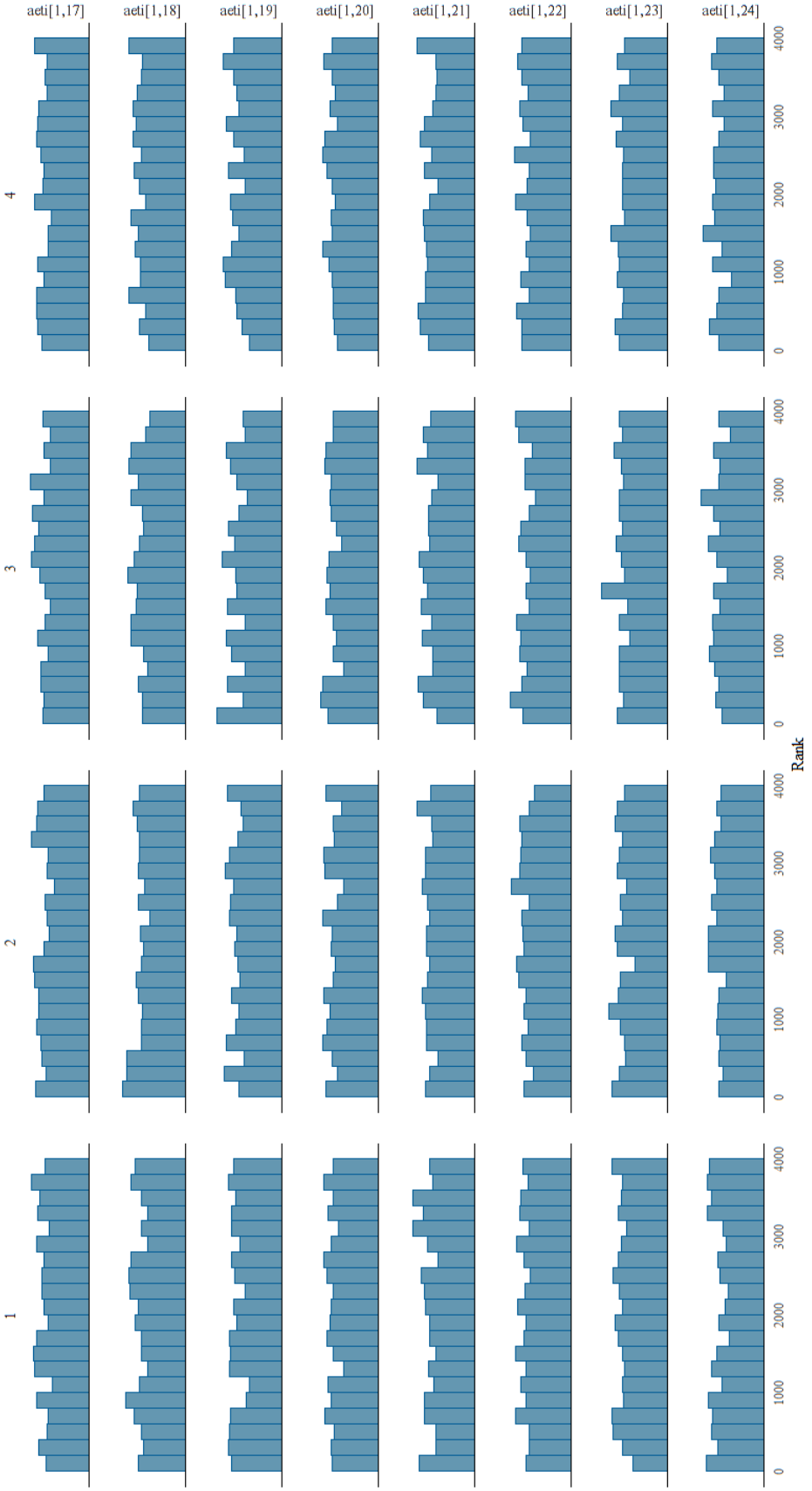
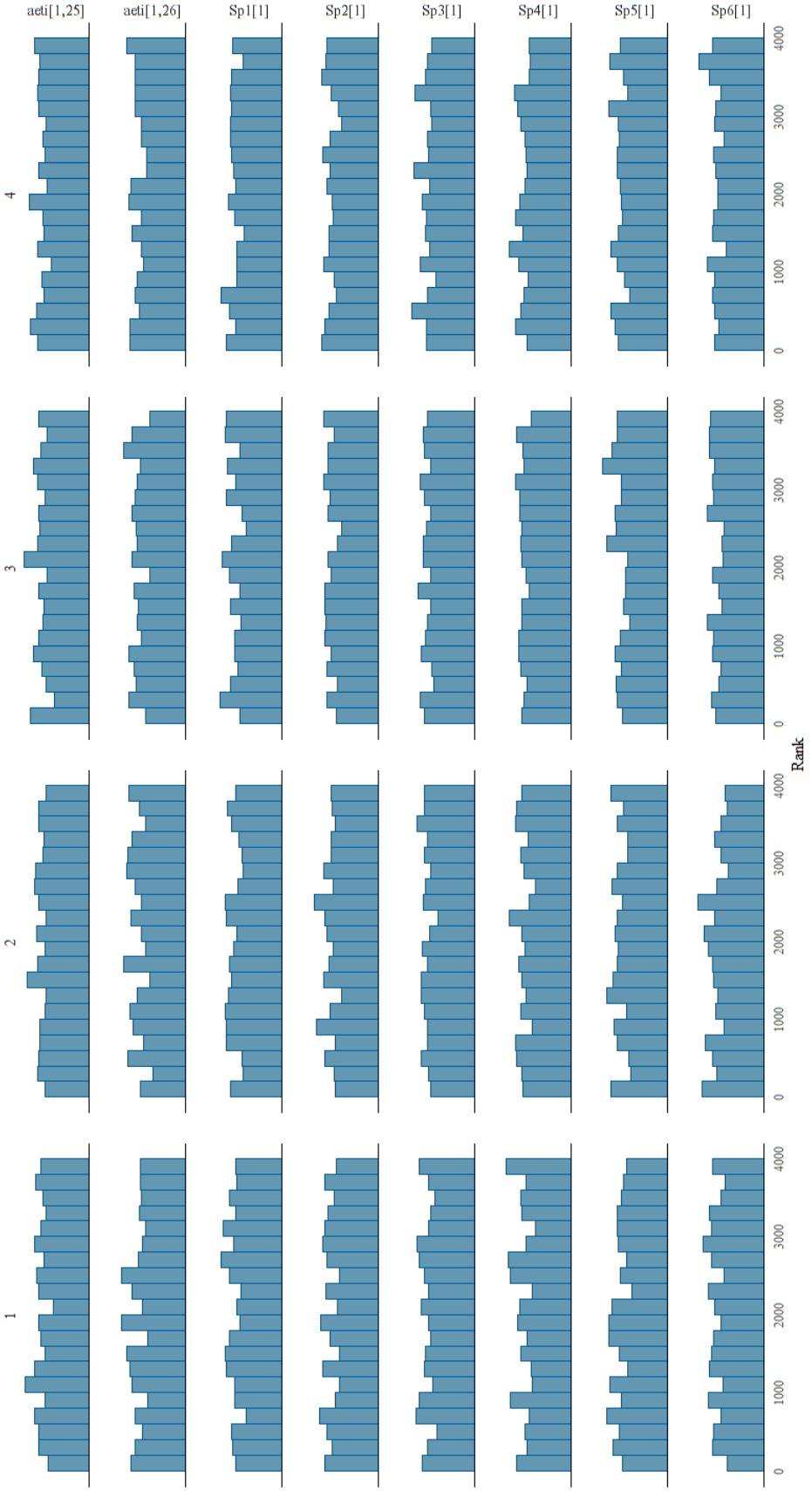

Chain
1
2
3
4
Se22
Se23
Se24
Se25
Se26
Se27
Se28
Se29
Se30
Se31
Se32
Se33
Se34
Se35
Se36
Se37

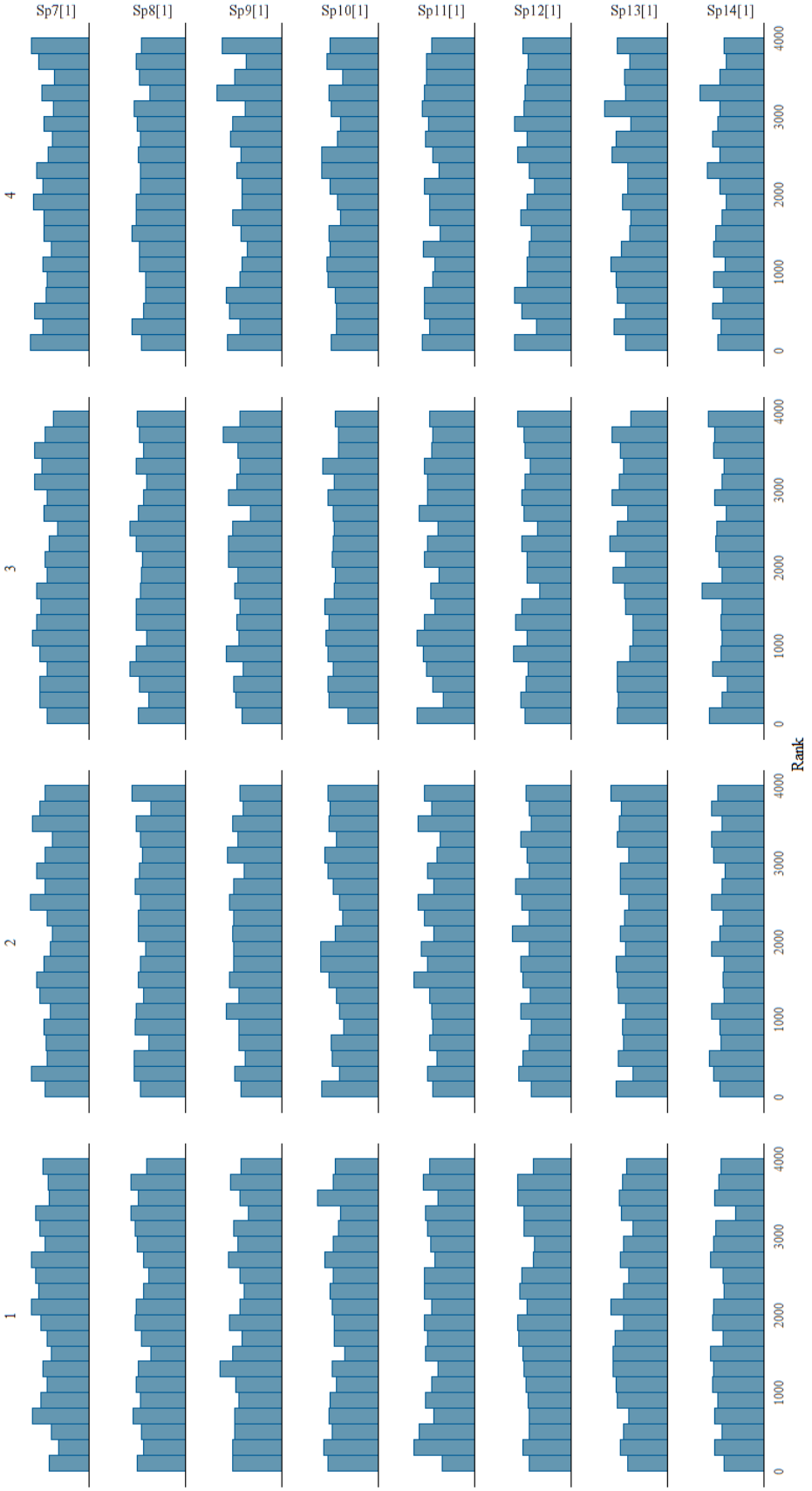
Figure H.2: Rank plots for the multi-pathogen inpatient model for all parameters related to children in Laos

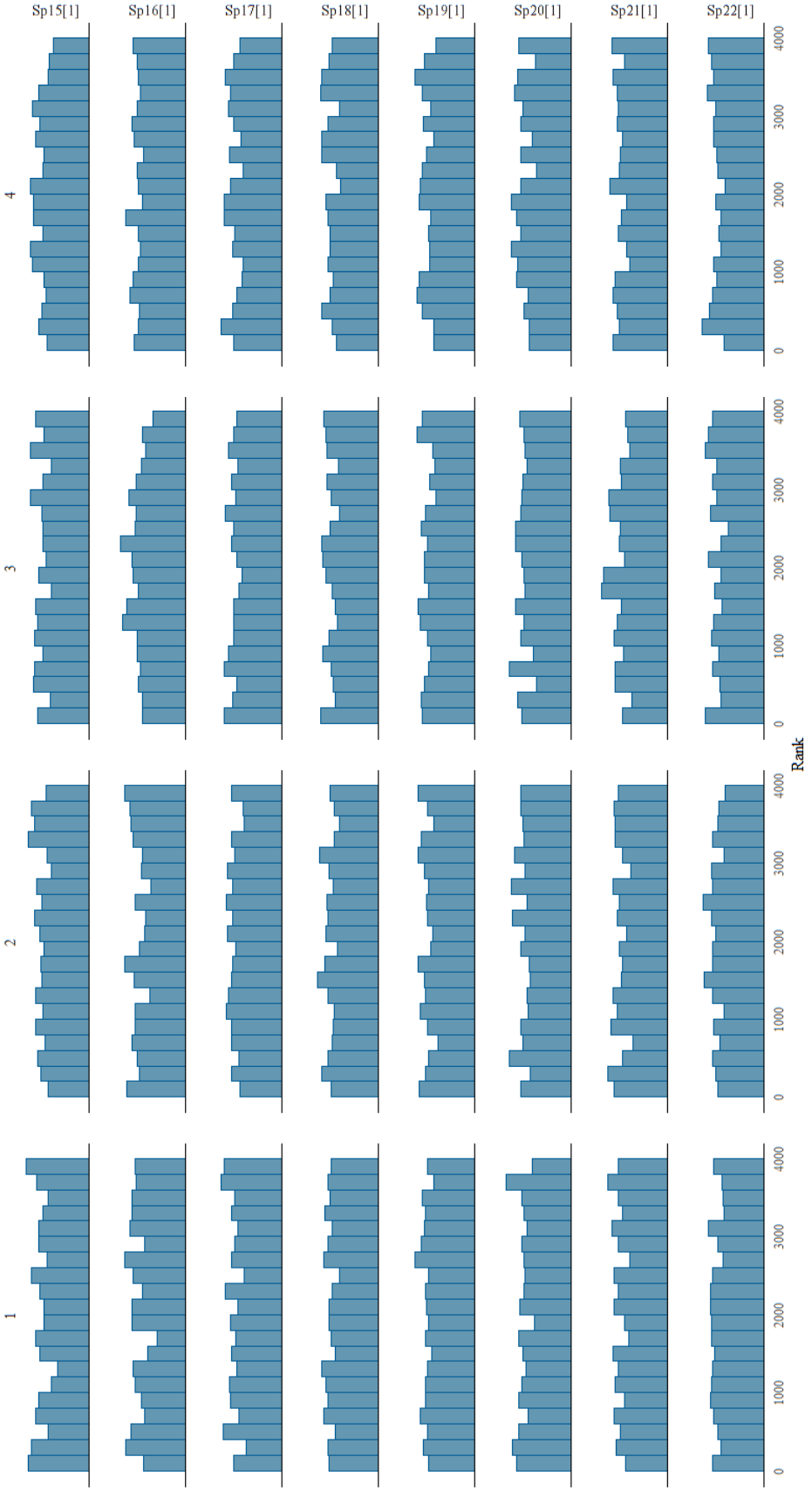


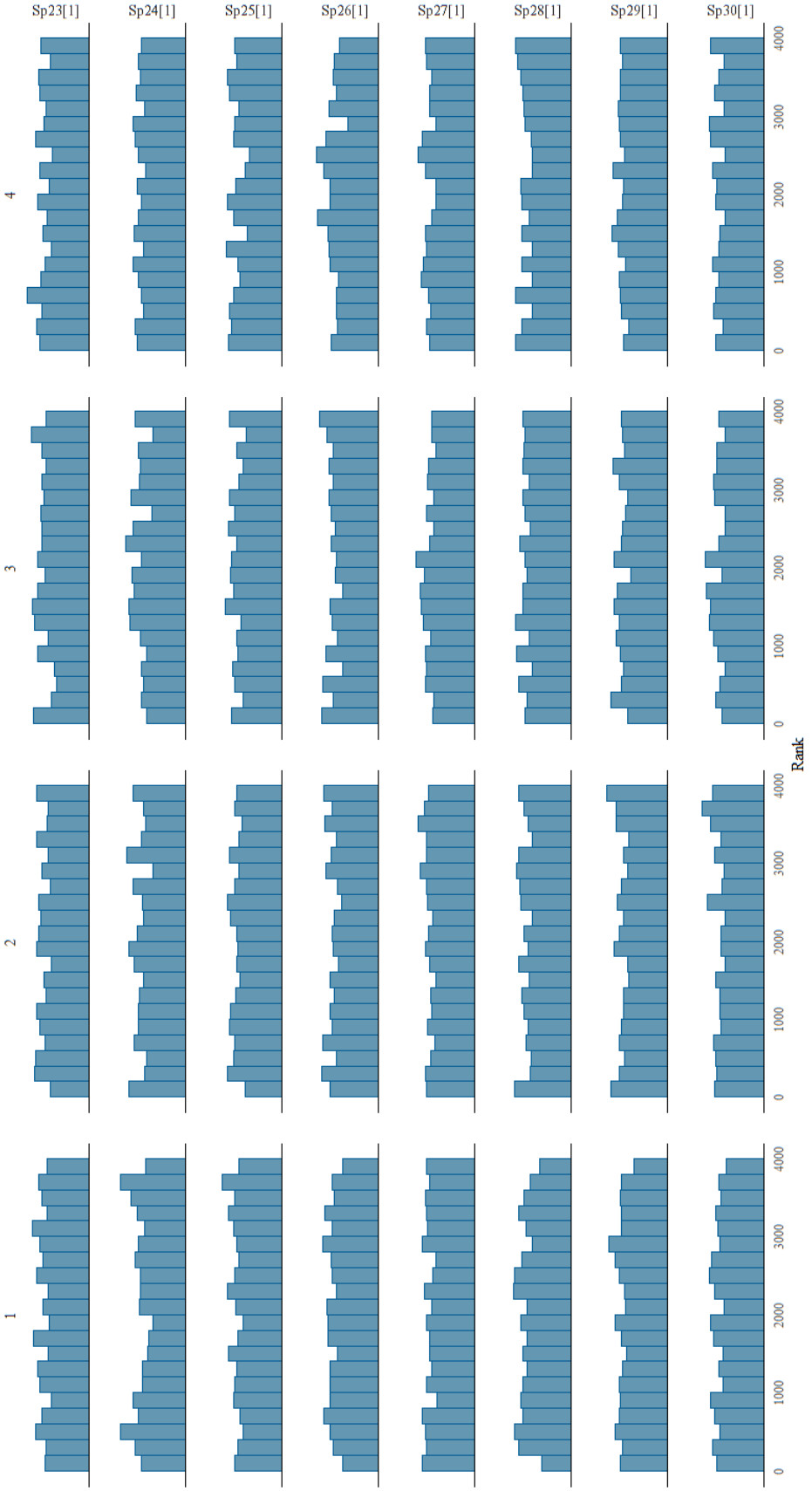

aeti[1,9]
aeti[1,10]
aeti[1,11]
aeti[1,12]
aeti[1,13]
aeti[1,14]
aeti[1,15]
aeti[1,16]
4
3
2
1
0
1000
2000
3000
4000
Rank

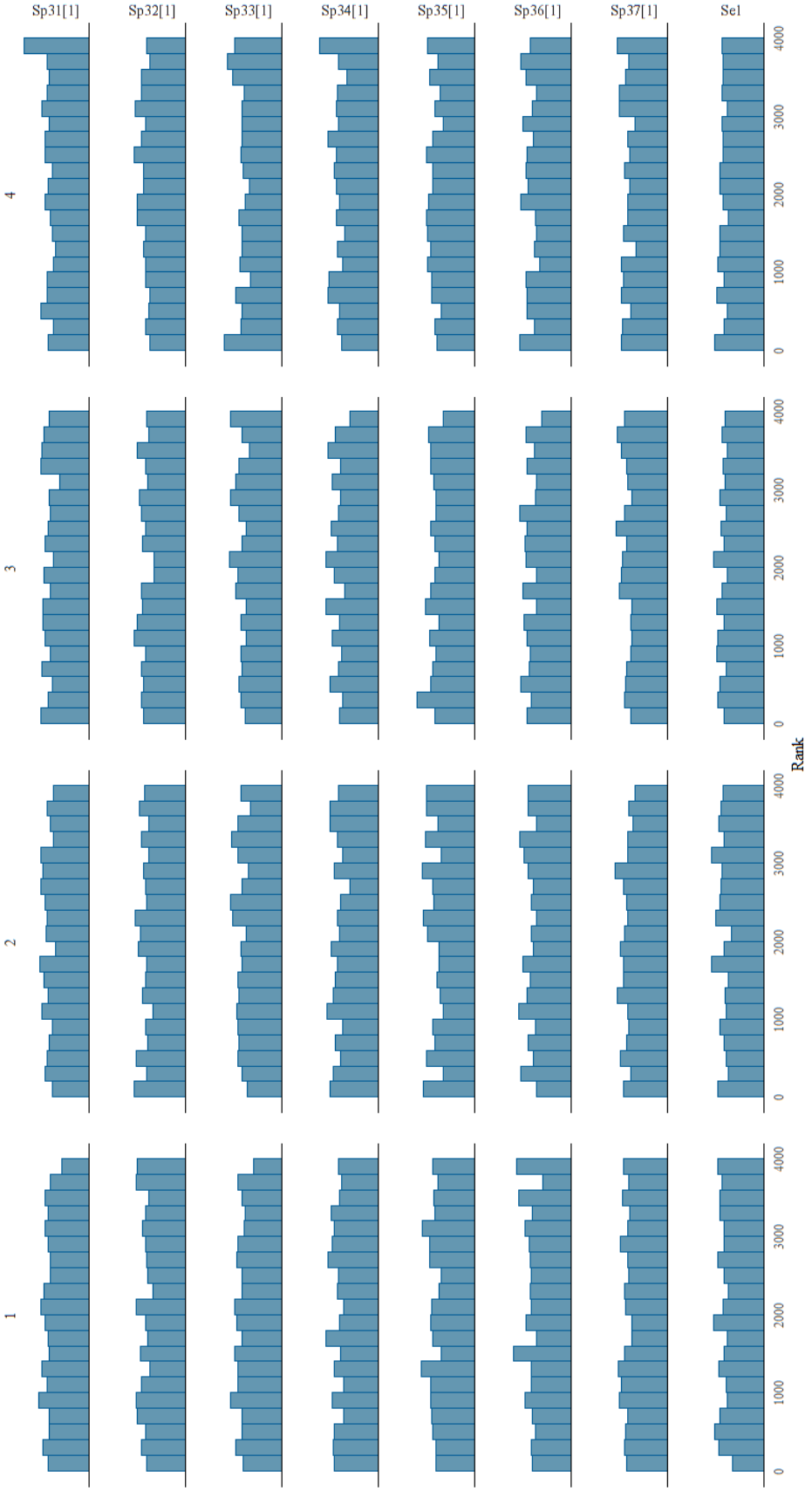

aeti[1,17]
aeti[1,18]
aeti[1,19]
aeti[1,20]
aeti[1,21]
aeti[1,22]
aeti[1,23]
aeti[1,24]
4
3
2
1
0
1000
2000
3000
4000
Rank

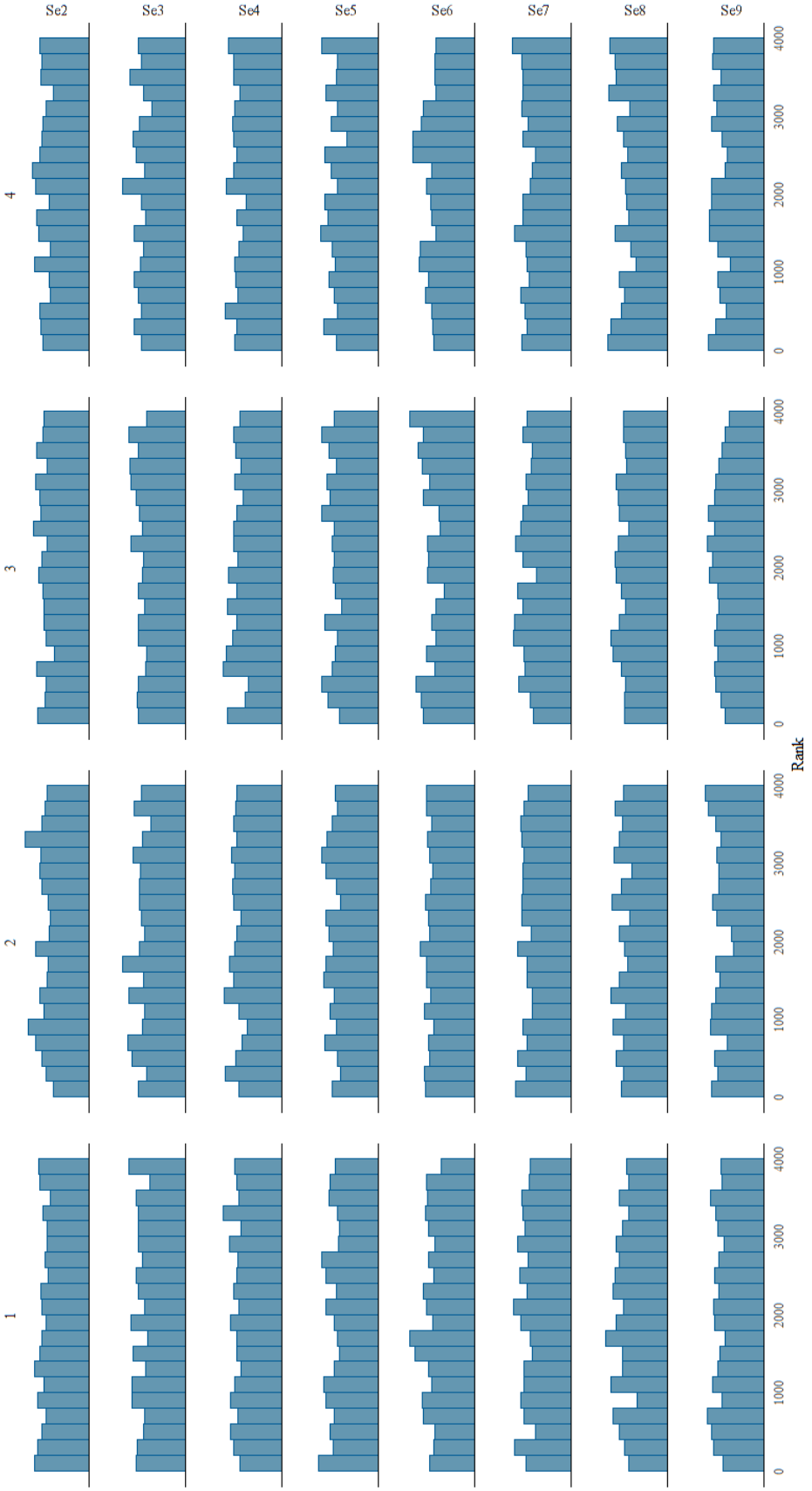

aeti[1,25]
aeti[1,26]
Sp1[1]
Sp2[1]
Sp3[1]
Sp4[1]
Sp5[1]
Sp6[1]
4
3
2
1
0
1000
2000
3000
4000
Rank

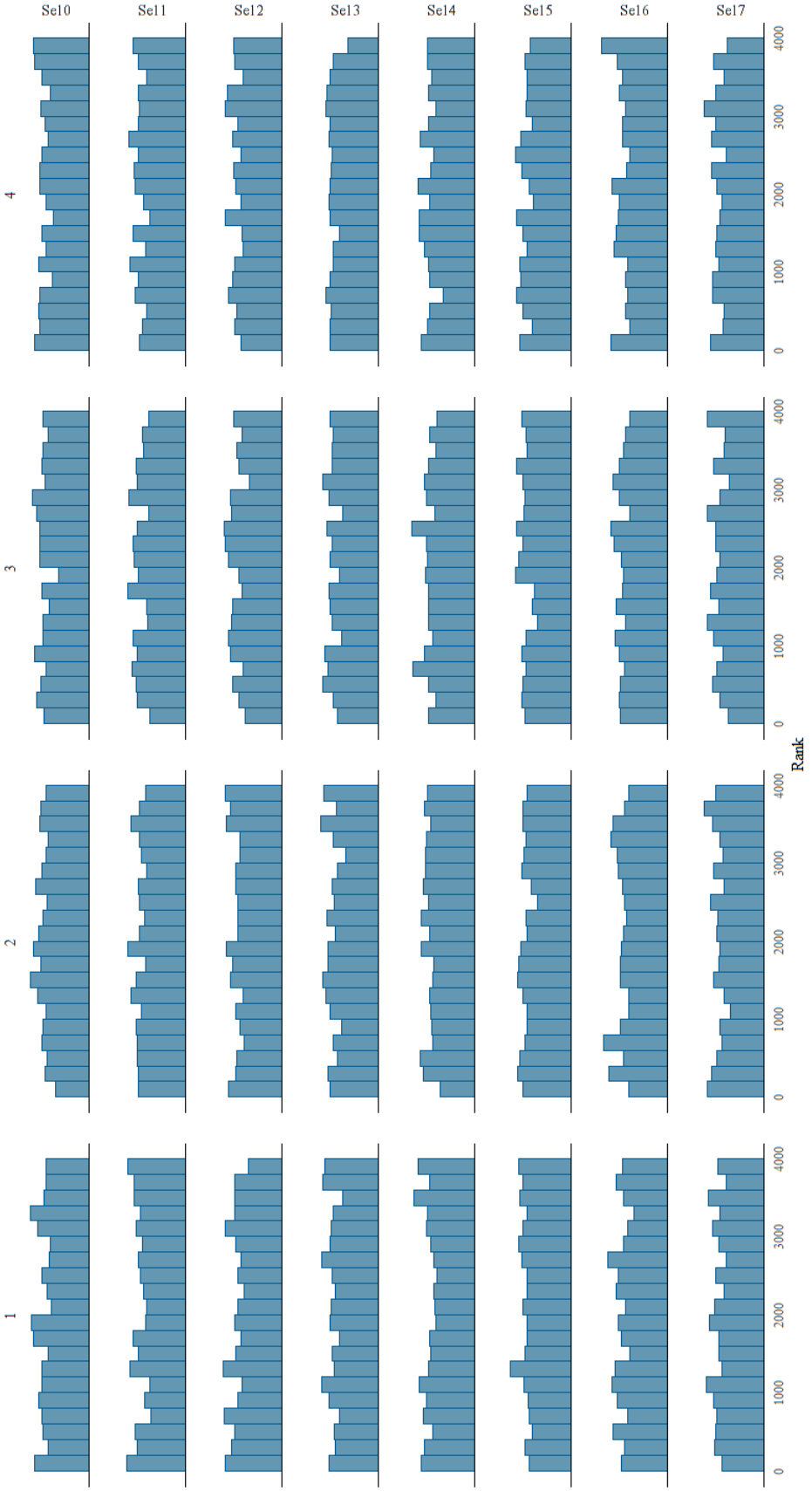

Sp7[1]
Sp8[1]
Sp9[1]
Sp10[1]
Sp11[1]
Sp12[1]
Sp13[1]
Sp14[1]
4
3
2
1
0
1000
2000
3000
4000
Rank

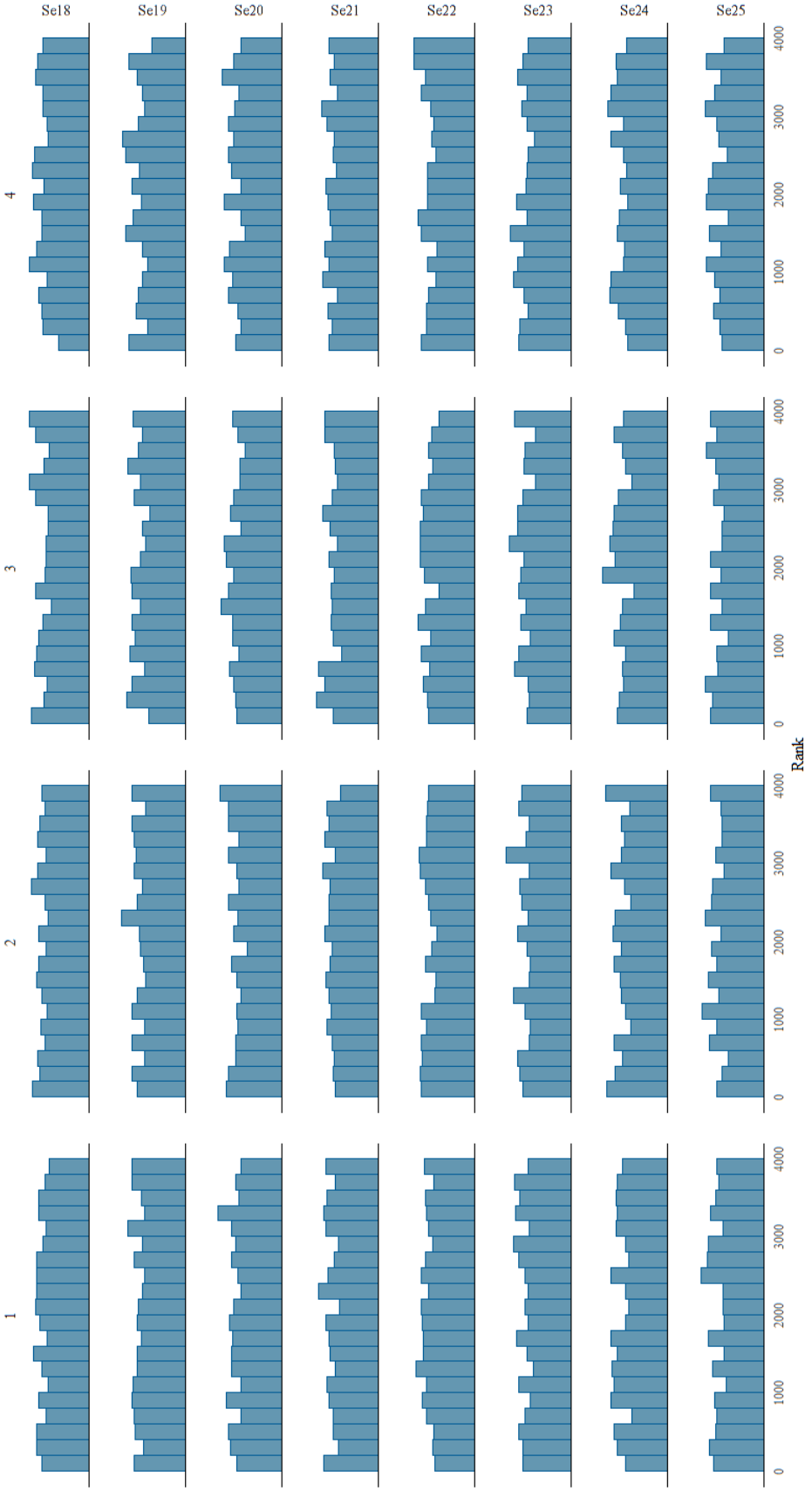

Sp15[1]
Sp16[1]
Sp17[1]
Sp18[1]
Sp19[1]
Sp20[1]
Sp21[1]
Sp22[1]
4
3
2
1
0
1000
2000
3000
4000
Rank

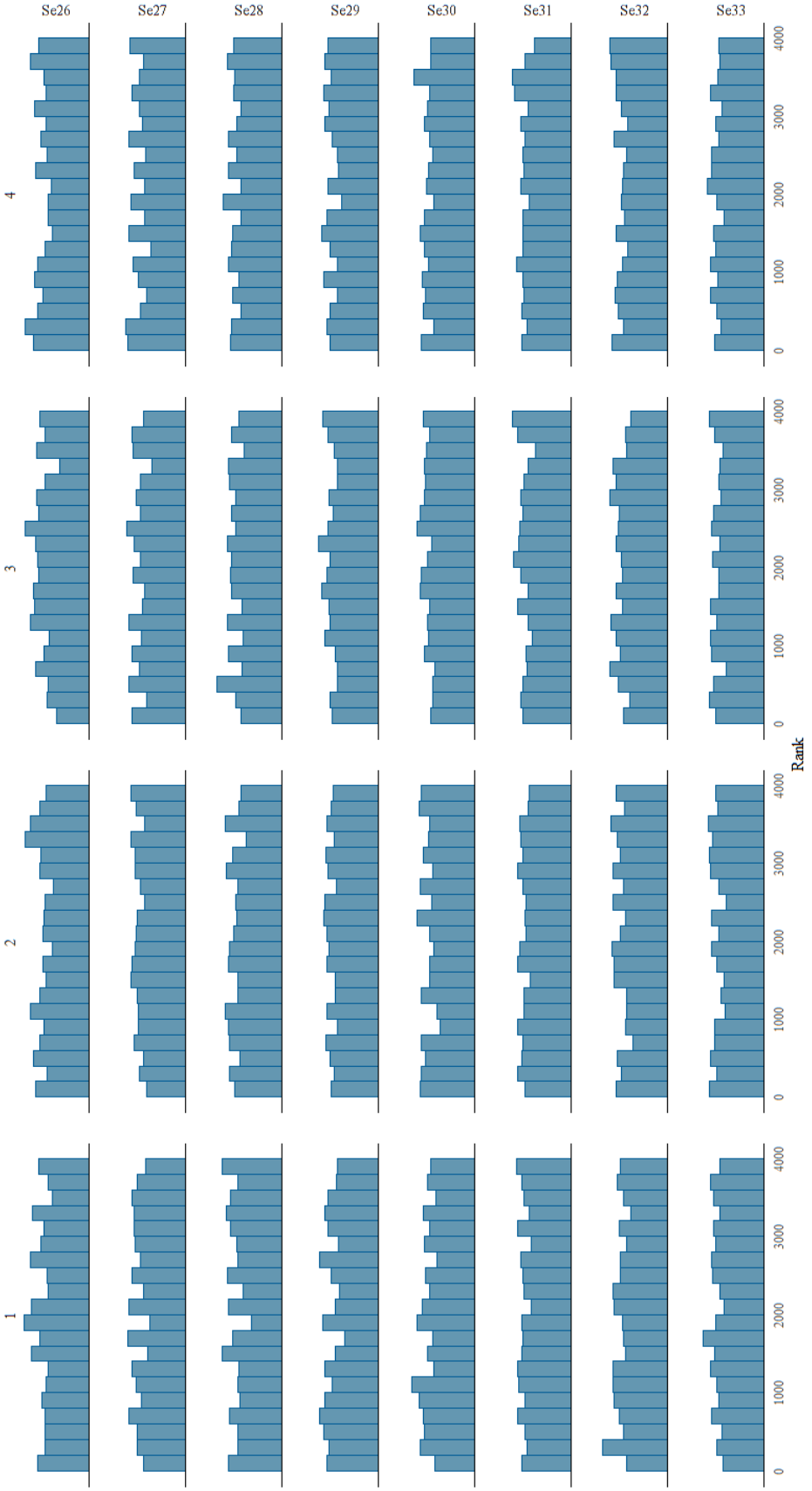

Sp23[1]
Sp24[1]
Sp25[1]
Sp26[1]
Sp27[1]
Sp28[1]
Sp29[1]
Sp30[1]
4
3
2
1
0
1000
2000
3000
4000
Rank

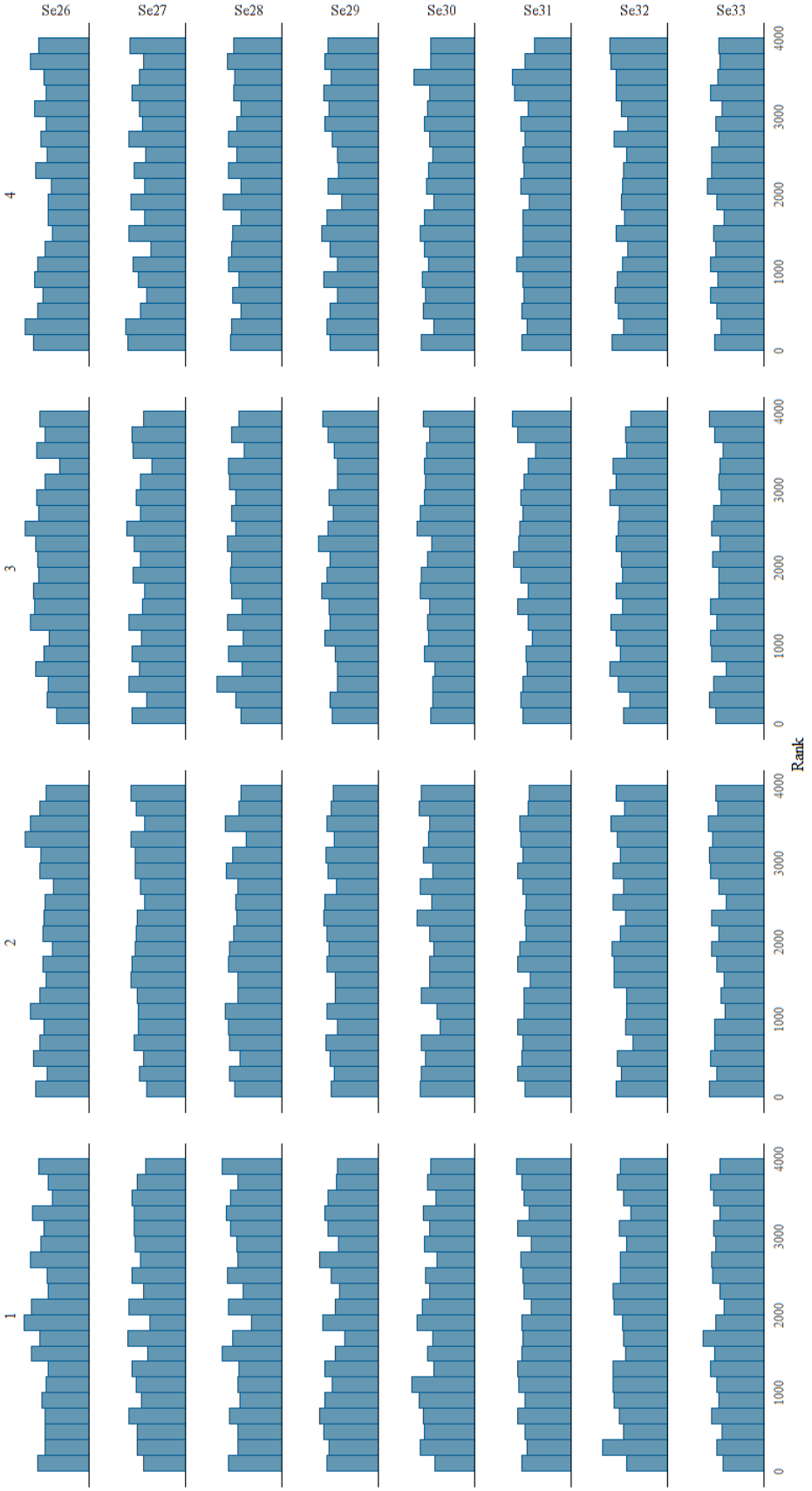

Sp31[1]
Sp32[1]
Sp33[1]
Sp34[1]
Sp35[1]
Sp36[1]
Sp37[1]
Se1
4
3
2
1
0
1000
2000
3000
4000
Rank


Se2
Se3
Se4
Se5
Se6
Se7
Se8
Se9
4
3
2
1
0
1000
2000
3000
4000
Rank


Sel0
Sel1
Sel2
Sel3
Sel4
Sel5
Sel6
Sel7
4
3
2
1
0
1000
2000
3000
4000
Rank


Se18
Se19
Se20
Se21
Se22
Se23
Se24
Se25
4
3
2
1
0
1000
2000
3000
4000
Rank


Se26
Se27
Se28
Se29
Se30
Se31
Se32
Se33
4
3
2
1
0
1000
2000
3000
4000
Rank


Se26
Se27
Se28
Se29
Se30
Se31
Se32
Se33
4
3
2
1
0
1000
2000
3000
4000
Rank

Bibliography

[1] N. Prasad et al. "Community prevalence of fever and relationship with malaria among infants and children in low-resource areas". In: *Am J Trop Med Hyg* 93.1 (2015), pp. 178–80. ISSN: 1476-1645 (Electronic) 0002-9637 (Print) 0002-9637 (Linking). DOI: 10.4269/ajtmh.14-0646. URL: https://www.ncbi.nlm.nih.gov/pubmed/25918207.

[2] NHS England. *High temperature in children*. 2020. URL: https://www.nhs.uk/conditions/fever-in-children/ (visited on 12/05/2023).

[3] P. Roddy et al. "Quantifying the incidence of severe-febrile-illness hospital admissions in sub-Saharan Africa". In: *PLoS One* 14.7 (2019), e0220371. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0220371. URL: https://www.ncbi.nlm.nih.gov/pubmed/31344116.

[4] GBD 2019 Diseases and Injuries Collaborators. "Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019". In: *Lancet* 396.10258 (2020), pp. 1204–1222. ISSN: 1474-547X (Electronic) 0140-6736 (Print) 0140-6736 (Linking). DOI: 10.1016/S0140-6736(20)30925-9. URL: https://www.ncbi.nlm.nih.gov/pubmed/33069326.

[5] N. Prasad et al. "Etiology of Severe Febrile Illness in Low- and Middle-Income Countries: A Systematic Review". In: *PLoS One* 10.6 (2015), e0127962. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0127962. URL: https://www.ncbi.nlm.nih.gov/pubmed/26126200.

[6] K. Wangdi et al. "Diversity of infectious aetiologies of acute undifferentiated febrile illnesses in south and Southeast Asia: a systematic review". In: *BMC Infect Dis* 19.1 (2019), p. 577. ISSN: 1471-2334 (Electronic) 1471-2334 (Linking). DOI: 10.1186/s12879-

019-4185-y. URL: https://www.ncbi.nlm.nih.gov/pubmed/31272417.

[7] J. Moreira et al. "Epidemiology of acute febrile illness in Latin America". In: *Clin Microbiol Infect* 24.8 (2018), pp. 827–835. ISSN: 1469-0691 (Electronic) 1198-743X (Print) 1198-743X (Linking). DOI: 10.1016/j.cmi.2018.05.001. URL: https://www.ncbi.nlm.nih.gov/pubmed/29777926.

[8] N. J. Saad et al. "Seasonal dynamics of typhoid and paratyphoid fever". In: *Sci Rep* 8.1 (2018), p. 6870. ISSN: 2045-2322 (Electronic) 2045-2322 (Linking). DOI: 10.1038/s41598-018-25234-w. URL: https://www.ncbi.nlm.nih.gov/pubmed/29720736.

[9] D. Fisman. "Seasonality of viral infections: mechanisms and unknowns". In: *Clin Microbiol Infect* 18.10 (2012), pp. 946–54. ISSN: 1469-0691 (Electronic) 1198-743X (Linking). DOI: 10.1111/j.1469-0691.2012.03968.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/22817528.

[10] V. D'Acremont et al. "Beyond malaria–causes of fever in outpatient Tanzanian children". In: *N Engl J Med* 370.9 (2014), pp. 809–17. ISSN: 1533-4406 (Electronic) 0028-4793 (Linking). DOI: 10.1056/NEJMoa1214482. URL: https://www.ncbi.nlm.nih.gov/pubmed/24571753.

[11] N. Boillat-Blanco et al. "Causes of fever in Tanzanian adults attending outpatient clinics: a prospective cohort study". In: *Clin Microbiol Infect* 27.6 (2021), 913 e1–913 e7. ISSN: 1469-0691 (Electronic) 1198-743X (Print) 1198-743X (Linking). DOI: 10.1016/j.cmi.2020.08.031. URL: https://www.ncbi.nlm.nih.gov/pubmed/32896654.

[12] T. K. Nguyen et al. "Etiologies of fever of unknown origin in HIV/AIDS patients, Hanoi, Vietnam". In: *BMC Infect Dis* 22.1 (2022), p. 61. ISSN: 1471-2334 (Electronic) 1471-2334 (Linking). DOI: 10.1186/s12879-022-07049-3. URL: https://www.ncbi.nlm.nih.gov/pubmed/35042469.

[13] N. Le-Viet et al. "Prospective case-control analysis of the aetiologies of acute undifferentiated fever in Vietnam". In: *Emerg Microbes Infect* 8.1 (2019), pp. 339–352. ISSN: 2222-1751 (Electronic) 2222-1751 (Linking). DOI: 10.1080/22221751.2019.1580539. URL: https://www.ncbi.nlm.nih.gov/pubmed/30866787.

[14] Neil K. Kaneshiro. *Fever - Symptoms and Causes*. 2022. URL: https://www.pennmedicine.org/for-patients-and-visitors/patient-information/conditions-treated-a-to-z/fever#:~:text=Almost%20any%20infection%20can%20cause,%2C%20bronchitis%2C%20pneumonia%2C%20and%20tuberculosis (visited on 01/02/2024).

[15] Piero Olliaro et al. "Advancing Access to Diagnostic Tools Essential for Universal Health Coverage and Antimicrobial Resistance Prevention: An Overview of Trials in Sub-Saharan Africa". In: *Clinical Infectious Diseases* 77.Supplement$_2$ (July 2023), S125–S133. ISSN: 1058-4838. DOI: 10.1093/cid/ciad326. eprint: https://academic.oup.com/cid/article-pdf/77/Supplement_2/S125/50960775/ciad326.pdf.

[16] World Health Organisation. *World Malaria Report 2022*. Report. 2022.

[17] Anthony Moody. "Rapid Diagnostic Tests for Malaria Parasites". In: *Clinical Microbiology Reviews* 15.1 (2002), pp. 66–78. DOI: 10.1128/cmr.15.1.66-78.2002. eprint: https://journals.asm.org/doi/pdf/10.1128/cmr.15.1.66-78.2002. URL: https://journals.asm.org/doi/abs/10.1128/cmr.15.1.66-78.2002.

[18] John A. Crump et al. "Etiology of Severe Non-malaria Febrile Illness in Northern Tanzania: A Prospective Cohort Study". In: *PLOS Neglected Tropical Diseases* 7 (July 2013), pp. 1–8. DOI: 10.1371/journal.pntd.0002324. URL: https://doi.org/10.1371/journal.pntd.0002324.

[19] Narain H. Punjabi et al. "Etiology of acute, non-malaria, febrile illnesses in Jayapura, northeastern Papua, Indonesia". In: *The American journal of tropical medicine and hygiene* 86.1 (2012), pp. 46–51. URL: https://doi.org/10.4269/ajtmh.2012.10-0497.

[20] H. Hopkins et al. "Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE): protocol for a multisite prospective observational study of the causes of fever in Africa and Asia". In: *BMJ Open* 10.7 (2020), e035632. ISSN: 2044-6055 (Electronic) 2044-6055 (Linking). DOI: 10.1136/bmjopen-2019-035632. URL: https://www.ncbi.nlm.nih.gov/pubmed/32699131.

[21] Nandini Dendukuri. “Commentary on On the robustness of latent class models for diagnostic testing with no gold-standard by Schofield et al.” In: *Statistics in Medicine* 40.22 (2021), pp. 4766–4769. DOI: https://doi.org/10.1002/sim.9086.

[22] S. H. Keddie et al. “Estimating sensitivity and specificity of diagnostic tests using latent class models that account for conditional dependence between tests: a simulation study”. In: *BMC Medical Research Methodology* 23.58 (2023). DOI: https://doi.org/10.1186/s12874-023-01873-0. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-01873-0#citeas.

[23] P. M. Vacek. “The effect of conditional dependence on the evaluation of diagnostic tests”. In: *Biometrics* 41.4 (1985). Vacek, P M eng 14212/PHS HHS/ Research Support, U.S. Gov’t, P.H.S. Biometrics. 1985 Dec;41(4):959-68., pp. 959–68. ISSN: 0006-341X (Print) 0006-341X (Linking). URL: https://www.ncbi.nlm.nih.gov/pubmed/3830260.

[24] Todd A. Alonzo Pepe and Margaret S. “Using a combination of reference tests to assess the accuracy of a new diagnostic test”. In: *Statistics in Medicine* 18 (1999), pp. 2987–3003.

[25] I. Schiller et al. “Bias due to composite reference standards in diagnostic accuracy studies”. In: *Stat Med* 35.9 (2016), pp. 1454–70. ISSN: 1097-0258 (Electronic) 0277-6715 (Linking). DOI: 10.1002/sim.6803. URL: https://www.ncbi.nlm.nih.gov/pubmed/26555849.

[26] Y Takwoingi et al. “Chapter 10: Undertaking meta-analysis”. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Ed. by J.J Deeks et al. Cochrane 2023, 2023.

[27] T. B. Gravelle et al. “Estimating the size of "anti-vax" and vaccine hesitant populations in the US, UK, and Canada: comparative latent class modeling of vaccine attitudes”. In: *Hum Vaccin Immunother* 18.1 (2022), p. 2008214. ISSN: 2164-554X (Electronic) 2164-5515 (Print) 2164-5515 (Linking). DOI: 10.1080/21645515.2021.2008214. URL: https://www.ncbi.nlm.nih.gov/pubmed/35349385.

[28] S. M. Stevens et al. “A latent class analysis approach to the identification of doctoral students at risk of attrition”. In: *PLoS One* 18.1 (2023), e0280325. ISSN: 1932-6203 (Electronic) 1932-

6203 (Linking). DOI: 10.1371/journal.pone.0280325. URL: https://www.ncbi.nlm.nih.gov/pubmed/36638114.

[29] L. Yue et al. "Patterns of sleep problems and internalizing and externalizing problems among Chinese adolescents: A latent class analysis". In: *Sleep Med* 95 (2022), pp. 47–54. ISSN: 1878-5506 (Electronic) 1389-9457 (Linking). DOI: 10.1016/j.sleep.2022.04.008. URL: https://www.ncbi.nlm.nih.gov/pubmed/35561474.

[30] Jacques A Hagenaars and Allan L McCutcheon. *Applied latent class analysis*. Cambridge University Press, 2002. ISBN: 1139439235.

[31] Delucchi KL Sinha P Calfee CS. "Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls". In: *Crit Care Med* 49.1 (2021), pp. 63–79. DOI: 10.1097/CCM.0000000000004710.

[32] Theresa W. Gyorkos Lawrence Joseph and Louis Coupal. "Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard". In: *American Journal of Epidemiology* 141 (1995), pp. 263–272.

[33] S.L.Hui and S.D.Walter. "Estimating the error rates of diagnostic tests". In: *Biometrics* 36 (1980), pp. 167–171.

[34] C. G. Lindholz et al. "Study of diagnostic accuracy of Helmintex, Kato-Katz, and POC-CCA methods for diagnosing intestinal schistosomiasis in Candeal, a low intensity transmission area in northeastern Brazil". In: *PLoS Negl Trop Dis* 12.3 (2018), e0006274. ISSN: 1935-2735 (Electronic) 1935-2727 (Print) 1935-2727 (Linking). DOI: 10.1371/journal.pntd.0006274. URL: https://www.ncbi.nlm.nih.gov/pubmed/29518081.

[35] H. N. Le et al. "The accuracy of tuberculous meningitis diagnostic tests using Bayesian latent class analysis". In: *J Infect Dev Ctries* 14.5 (2020), pp. 479–487. ISSN: 1972-2680 (Electronic) 1972-2680 (Linking). DOI: 10.3855/jidc.11862. URL: https://www.ncbi.nlm.nih.gov/pubmed/32525834.

[36] M. M. Mweu et al. "Accuracy of diagnostic tests for respiratory syncytial virus infection within a paediatric hospital population in Kilifi County, Kenya". In: *Wellcome Open Res* 5 (2020), p. 155. ISSN: 2398-502X (Print) 2398-502X (Electronic) 2398-502X (Linking). DOI: 10.12688/wellcomeopenres.16067.1. URL: https://www.ncbi.nlm.nih.gov/pubmed/32984548.

[37] M. Deloria Knoll et al. “Bayesian Estimation of Pneumonia Etiology: Epidemiologic Considerations and Applications to the Pneumonia Etiology Research for Child Health Study”. In: *Clin Infect Dis* 64.suppl$_3$ (2017), S213–S227. ISSN: 1537-6591 (Electronic) 1058-4838 (Print) 1058-4838 (Linking). DOI: 10.1093/cid/cix144. URL: https://www.ncbi.nlm.nih.gov/pubmed/28575370.

[38] O. S. Levine et al. “The Pneumonia Etiology Research for Child Health Project: a 21st century childhood pneumonia etiology study”. In: *Clin Infect Dis* 54 Suppl 2.Suppl 2 (2012), S93–101. ISSN: 1537-6591 (Electronic) 1058-4838 (Print) 1058-4838 (Linking). DOI: 10.1093/cid/cir1052. URL: https://www.ncbi.nlm.nih.gov/pubmed/22403238.

[39] R. Symons et al. “A statistical framework to estimate diagnostic test performance for COVID-19”. In: *Clin Radiol* 76.1 (2021), 75.e1–75.e3. ISSN: 0009-9260 (Print) 0009-9260. DOI: 10.1016/j.crad.2020.10.004. URL: https://www.clinicalradiologyonline.net/article/S0009-9260(20)30445-1/pdf.

[40] S. Hartnack et al. “Estimating diagnostic test accuracies for Brachyspira hyodysenteriae accounting for the complexities of population structure in food animals”. In: *PLoS One* 9.6 (2014), e98534. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0098534. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4048188/pdf/pone.0098534.pdf.

[41] Madhukar Pai et al. “Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis”. In: *BMC Infectious Diseases* 4.1 (2004), p. 6. ISSN: 1471-2334. DOI: 10.1186/1471-2334-4-6. URL: https://doi.org/10.1186/1471-2334-4-6%20https://bmcinfectdis.biomedcentral.com/track/pdf/10.1186/1471-2334-4-6.pdf.

[42] C Mayers and K Baker. *Impact of false-positives and false-negatives in the UK’s COVID-19 RT-PCR testing programme*. Generic. 2020.

[43] FIND. *SARS-COV-2 TEST TRACKER*. URL: https://www.finddx.org/tools-and-resources/dxconnect/test-directories/covid-19-test-tracker/ (visited on 08/01/2023).

[44] Harika Yadav et al. “Availability of essential diagnostics in ten low-income and middle-income countries: results from national health facility surveys”. In: *The Lancet Global Health* 9.11 (2021). doi: 10.1016/S2214-109X(21)00442-3, e1553–e1560. ISSN: 2214-109X.

DOI: 10.1016/S2214-109X(21)00442-3. URL: https://doi.org/10.1016/S2214-109X(21)00442-3.

[45] Michael L. Wilson et al. "Access to pathology and laboratory medicine services: a crucial gap". In: *The Lancet* 391.10133 (2018). doi: 10.1016/S0140-6736(18)30458-6, pp. 1927–1938. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(18)30458-6. URL: https://doi.org/10.1016/S0140-6736(18)30458-6.

[46] Kenneth A. Fleming et al. "The Lancet Commission on diagnostics: transforming access to diagnostics". In: *The Lancet* 398.10315 (2021). doi: 10.1016/S0140-6736(21)00673-5, pp. 1997–2050. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(21)00673-5. URL: https://doi.org/10.1016/S0140-6736(21)00673-5.

[47] World Health Organization. *The selection and use of essential in vitro diagnostics: report of the third meeting of the WHO Strategic Advisory Group of Experts on In Vitro Diagnostics*. Tech. rep. Geneva, World Health Organization, 2020.

[48] Hermann Brenner. "Correcting for Exposure Misclassification Using an Alloyed Gold Standard". In: *Epidemiology* 7.4 (1996), pp. 406–410. ISSN: 10443983. URL: http://www.jstor.org/stable/3702059 (visited on 07/31/2023).

[49] Ruth H. Keogh et al. "STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1Basic theory and simple methods of adjustment". In: *Statistics in Medicine* 39.16 (2020), pp. 2197–2231. DOI: https://doi.org/10.1002/sim.8532.

[50] S.D Walter and L.M Irwig. "Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review". In: *Journal of clinical epidemiology* 41.9 (1988), pp. 923–937. DOI: https://doi.org/10.1016/0895-4356(88)90110-2.

[51] Chinyereugo M. Umemneku Chikere et al. "Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard An update". In: *PLOS ONE* 14.10 (Oct. 2019), pp. 1–25. DOI: 10.1371/journal.pone.0223832. URL: https://doi.org/10.1371/journal.pone.0223832.

[52] Maurice Staquet et al. "Methodology for the assessment of new dichotomous diagnostic tests". In: *Journal of Chronic Diseases* 34.12 (1981), pp. 599–610. ISSN: 0021-9681. DOI: https://doi.org/10.1016/0021-9681(81)90059-X.

[53] Douglas M. Hawkins, James A. Garrett, and Betty Stephenson. "Some issues in resolution of diagnostic tests using an imperfect gold standard". In: *Statistics in Medicine* 20.13 (2001), pp. 1987–2001. DOI: https://doi.org/10.1002/sim.819.

[54] M. van Smeden et al. "Latent class models in diagnostic studies when there is no reference standard–a systematic review". In: *Am J Epidemiol* 179.4 (2014), pp. 423–31. ISSN: 0002-9262. DOI: 10.1093/aje/kwt286.

[55] Johannes B. Reitsma et al. "A review of solutions for diagnostic accuracy studies with an imperfect orxa0;missing reference standard". In: *Journal of Clinical Epidemiology* 62.8 (2009). doi: 10.1016/j.jclinepi.2009.02.005, pp. 797–806. ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2009.02.005. URL: https://doi.org/10.1016/j.jclinepi.2009.02.005.

[56] Nandini Dendukuri, Alula Hadgu, and Liangliang Wang. "Modeling conditional dependence between diagnostic tests: A multiple latent variable model". In: *Statistics in Medicine* 28.3 (2009), pp. 441–461. DOI: https://doi.org/10.1002/sim.3470. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3470. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3470.

[57] M. Boelaert et al. "The potential of latent class analysis in diagnostic test validation for canine Leishmania infantum infection". In: *Epidemiol Infect* 123.3 (1999). useful, pp. 499–506. ISSN: 0950-2688 (Print) 0950-2688. DOI: 10.1017/s0950268899003040. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2810786/pdf/10694163.pdf.

[58] Yang Xie and Bradley Carlin. "Measures of Bayesian learning and identifiability in hierarchical models". In: *Journal of Statistical Planning and Inference* 136 (Oct. 2006), pp. 3458–3477. DOI: 10.1016/j.jspi.2005.04.003.

[59] Franz-Georg Wieland et al. "On structural and practical identifiability". In: *Current Opinion in Systems Biology* 25 (2021), pp. 60–69. ISSN: 2452-3100. DOI: https://doi.org/10.1016/j.coisb.2021.03.005. URL: https://www.sciencedirect.com/science/article/pii/S245231002100007X.

[60] Paul Gustafson et al. "On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables [with Comments and Re-

joinder]". In: *Statistical Science* 20.2 (2005), pp. 111–140. ISSN: 08834237. URL: http://www.jstor.org/stable/20061166 (visited on 07/31/2023).

[61] Paul Gustafson. "Bayesian Inference for Partially Identified Models". In: *The International Journal of Biostatistics* 6.2 (2010). DOI: doi:10.2202/1557-4679.1206. URL: https://doi.org/10.2202/1557-4679.1206.

[62] Emily L. MacLean et al. "Bayesian latent class analysis produced diagnostic accuracy estimates that were more interpretable than composite reference standards for extrapulmonary tuberculosis tests". In: *Diagnostic and Prognostic Research* 6.1 (2022), p. 11. ISSN: 2397-7523. DOI: 10.1186/s41512-022-00125-x. URL: https://doi.org/10.1186/s41512-022-00125-x.

[63] Jonah Gabry et al. "Visualization in Bayesian Workflow". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 182.2 (Jan. 2019), pp. 389–402. ISSN: 0964-1998. DOI: 10.1111/rssa.12378. eprint: https://academic.oup.com/jrsssa/article-pdf/182/2/389/49429057/jrsssa_182_2_389.pdf. URL: https://doi.org/10.1111/rssa.12378.

[64] Bruce D. Spencer. "When Do Latent Class Models Overstate Accuracy for Diagnostic and Other Classifiers in the Absence of a Gold Standard?" In: *Biometrics* 68.2 (2012), pp. 559–566. DOI: https://doi.org/10.1111/j.1541-0420.2011.01694.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1541-0420.2011.01694.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2011.01694.x.

[65] M. S. Pepe and H. Janes. "Insights into latent class analysis of diagnostic test performance". In: *Biostatistics* 8.2 (2007). useful, pp. 474–84. ISSN: 1465-4644 (Print) 1465-4644. DOI: 10.1093/biostatistics/kxl038.

[66] Trinh Huu Khanh Dong et al. "A novel diagnostic model for tuberculous meningitis using Bayesian latent class analysis". In: *BMC Infectious Diseases* 24.1 (2024), p. 163. ISSN: 1471-2334. DOI: 10.1186/s12879-024-08992-z. URL: https://doi.org/10.1186/s12879-024-08992-z.

[67] N. Dendukuri and L. Joseph. "Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests". In: *Biometrics* 57.1 (2001), pp. 158–67. ISSN: 0006-341X (Print) 0006-

341X (Linking). DOI: 10.1111/j.0006-341x.2001.00158.x. URL: https://www.ncbi.nlm.nih.gov/pubmed/11252592.

[68] Dean Harper. "Local dependence latent structure models". In: *Psychometrika* 37.1 (1972), pp. 53–59. ISSN: 1860-0980. DOI: 10.1007/BF02291412. URL: https://doi.org/10.1007/BF02291412.

[69] Y. Qu, M. Tan, and M. H. Kutner. "Random effects models in latent class analysis for evaluating accuracy of diagnostic tests". In: *Biometrics* 52.3 (1996), pp. 797–810. ISSN: 0006-341X (Print) 0006-341x.

[70] S. G. Schumacher et al. "Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis". In: *Am J Epidemiol* 184.9 (2016), pp. 690–700. ISSN: 0002-9262 (Print) 0002-9262. DOI: 10.1093/aje/kww094.

[71] J. Menten, M. Boelaert, and E. Lesaffre. "Bayesian latent class models with conditionally dependent diagnostic tests: a case study". In: *Stat Med* 27.22 (2008), pp. 4469–88. ISSN: 0277-6715 (Print) 0277-6715. DOI: 10.1002/sim.3317. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3317.

[72] M. R. Schofield et al. "On the robustness of latent class models for diagnostic testing with no gold standard". In: *Stat Med* (2021). ISSN: 1097-0258 (Electronic) 0277-6715 (Linking). DOI: 10.1002/sim.8999.

[73] Bridget E. Weller, Natasha K. Bowen, and Sarah J. Faubert. "Latent Class Analysis: A Guide to Best Practice". In: *Journal of Black Psychology* 46.4 (2020), pp. 287–311. DOI: 10.1177/0095798420930932.

[74] David J. Spiegelhalter et al. "Bayesian Measures of Model Complexity and Fit". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64.4 (Oct. 2002), pp. 583–639. ISSN: 1369-7412. DOI: 10.1111/1467-9868.00353. URL: https://doi.org/10.1111/1467-9868.00353.

[75] Direk Limmathurotsakul et al. "Defining the True Sensitivity of Culture for the Diagnosis of Melioidosis Using Bayesian Latent Class Models". In: *PLOS ONE* 5.8 (Aug. 2010), pp. 1–6. DOI: 10.1371/journal.pone.0012485. URL: https://doi.org/10.1371/journal.pone.0012485.

[76] A. Gelman et al. *Bayesian Data Analysis (3rd ed.)* Chapman and Hall/CRC, 2013.

[77] Hirotugu Akaike. “Information Theory and an Extension of the Maximum Likelihood Principle”. In: 1973. URL: https://api.semanticscholar.org/CorpusID:64903870.

[78] Aki Vehtari et al. “Pareto smoothed importance sampling”. In: *arXiv preprint arXiv:1507.02646* (2015).

[79] Aki Vehtari, Andrew Gelman, and Jonah Gabry. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC”. In: *Statistics and computing* 27.5 (2017), pp. 1413–1432. ISSN: 1573-1375.

[80] Aki Vehtari et al. *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. R package version 2.6.0. 2023. URL: https://mc-stan.org/loo/.

[81] Martyn Plummer. “JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling”. In: *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria* 124 (Apr. 2003).

[82] David J. Lunn et al. “WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility”. In: *Statistics and Computing* 10.4 (2000), pp. 325–337. ISSN: 1573-1375. DOI: 10.1023/A:1008929526011. URL: https://doi.org/10.1023/A:1008929526011.

[83] MF Huque. “Experiences with meta-analysis in NDA submissions”. In: *Proceedings of the Biopharmaceutical Section of the American Statistical Association*. Vol. 2, pp. 28–33.

[84] L. Irwig et al. “Guidelines for meta-analyses evaluating diagnostic tests”. In: 0003-4819 (Print) ().

[85] Deeks J.J, Bossuyt P.M., and Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Cochrane 2013, 2013.

[86] Deeks J.J et al. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Cochrane 2023, 2023.

[87] L. E. Moses, D. Shapiro, and B. Littenberg. “Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations”. In: *Stat Med* 12.14 (1993), pp. 1293–316. DOI: 10.1002/sim.4780121403. URL: https://www.ncbi.nlm.nih.gov/pubmed/8210827%20https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780121403.

[88] C. M. Rutter and C. A. Gatsonis. “A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations”. In: *Stat Med* 20.19 (2001), pp. 2865–84. DOI: 10.1002/sim.942.

[89] N. Dendukuri et al. “Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference”. In: *Biometrics* 68.4 (2012), pp. 1285–93. DOI: 10.1111/j.1541-0420.2012.01773.x.

[90] Chukwuemeka Onwuchekwa et al. “Underascertainment of Respiratory Syncytial Virus Infection in Adults Due to Diagnostic Testing Limitations: A Systematic Literature Review and Meta-analysis”. In: *The Journal of Infectious Diseases* 228.2 (Jan. 2023), pp. 173–184. DOI: 10.1093/infdis/jiad012. eprint: https://academic.oup.com/jid/article-pdf/228/2/173/50880678/jiad012.pdf. URL: https://doi.org/10.1093/infdis/jiad012.

[91] Eleni Karlafti et al. “The Diagnostic Accuracy of SARS-CoV-2 Nasal Rapid Antigen Self-Test: A Systematic Review and Meta-Analysis”. In: *Life* 13.2 (2023). DOI: 10.3390/life13020281. URL: https://www.mdpi.com/2075-1729/13/2/281.

[92] Matthew J Page et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. In: *BMJ* 372 (2021). DOI: 10.1136/bmj.n71. eprint: https://www.bmj.com/content/372/bmj.n71.full.pdf. URL: https://www.bmj.com/content/372/bmj.n71.

[93] “QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies”. In: *Annals of Internal Medicine* 155.8 (2011), pp. 529–536. DOI: 10.7326/0003-4819-155-8-201110180-00009. URL: https://www.acpjournals.org/doi/abs/10.7326/0003-4819-155-8-201110180-00009.

[94] L. Irwig et al. “Meta-analytic methods for diagnostic test accuracy”. In: *J Clin Epidemiol* 48.1 (1995), 119–30, discussion 131–2. DOI: 10.1016/0895-4356(94)00099-c. URL: https://www.ncbi.nlm.nih.gov/pubmed/7853038.

[95] C. M. Rutter and C. A. Gatsonis. “Regression methods for meta-analysis of diagnostic test data”. In: *Acad Radiol* 2 Suppl 1 (1995), S48–56, discussion S65–7, S70–1 pas.

[96] J. B. Reitsma et al. “Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews”. In: *J Clin Epidemiol* 58.10 (2005), pp. 982–90. DOI: 10.1016/j.jclinepi.2005.02.022. URL: https://www.ncbi.nlm.nih.gov/pubmed/16168343%20https://www.jclinepi.com/article/S0895-4356(05)00162-9/fulltext.

[97] R. M. Harbord et al. “A unification of models for meta-analysis of diagnostic accuracy studies”. In: *Biostatistics* 8.2 (2007), pp. 239–51. DOI: 10.1093/biostatistics/kxl004.

[98] M. Sadatsafavi et al. “A statistical method was used for the meta-analysis of tests for latent TB in the absence of a gold standard, combining random-effect and latent-class methods to estimate test accuracy”. In: *J Clin Epidemiol* 63.3 (2010), pp. 257–69. DOI: 10.1016/j.jclinepi.2009.04.008. URL: https://www.jclinepi.com/article/S0895-4356(09)00129-2/fulltext.

[99] Enzo Cerullo et al. “Meta-analysis of dichotomous and polytomous diagnostic tests without a gold standard”. In: *Research Synthesis Methods* (2022), pp. 595–611.

[100] Antonia Zapf et al. “Meta-analysis of diagnostic accuracy studies with multiple thresholds: Comparison of different approaches”. In: *Biometrical Journal* 63.4 (2021), pp. 699–711. DOI: https://doi.org/10.1002/bimj.202000091. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202000091%20https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/bimj.202000091?download=true.

[101] E. Timsit et al. “Diagnostic accuracy of clinical illness for bovine respiratory disease (BRD) diagnosis in beef cattle placed in feedlots: A systematic literature review and hierarchical Bayesian latent-class meta-analysis”. In: *Preventive Veterinary Medicine* 135 (2016), pp. 67–73. ISSN: 0167-5877. DOI: https://doi.org/10.1016/j.prevetmed.2016.11.006. URL: https://www.sciencedirect.com/science/article/pii/S0167587716305372.

[102] Polychronis Kostoulas, Paolo Eusebi, and Sonja Hartnack. “Diagnostic accuracy estimates for COVID-19 real-time polymerase chain reaction and lateral flow immunoassay tests with bayesian latent-class models”. In: *American journal of epidemiology* 190.8 (2021), pp. 1689–1695.

[103] Hiroaki Ogawa et al. “Sputum Gram Stain for Bacterial Pathogen Diagnosis in Community-acquired Pneumonia: A Systematic Review and Bayesian Meta-analysis of Diagnostic Accuracy and Yield”. In: *Clinical Infectious Diseases* 71.3 (Sept. 2019), pp. 499–513. ISSN: 1058-4838. DOI: 10.1093/cid/ciz876. eprint: https://academic.oup.com/cid/article-pdf/71/3/499/33534769/ciz876.pdf. URL: https://doi.org/10.1093/cid/ciz876.

[104] John S. Uebersax. “Probit Latent Class Analysis with Dichotomous or Ordered Category Measures: Conditional Independence/Dependence Models”. In: *Applied Psychological Measurement* 23 (1999), pp. 283–297. URL: https://api.semanticscholar.org/CorpusID:120497324.

[105] Huiping Xu and Bruce A. Craig. “A Probit Latent Class Model with General Correlation Structures for Evaluating Accuracy of Diagnostic Tests”. In: *Biometrics* 65.4 (2009), pp. 1145–1155. DOI: https://doi.org/10.1111/j.1541-0420.2008.01194.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2008.01194.x.

[106] Deeks J Macaskill P Takwoingi Y and Gatsonis C. “Chapter 9: Understanding meta-analysis”. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. useful. 2023. Chap. 9. URL: https://training.cochrane.org/9-understanding-meta-analysis.

[107] J.J Deeks et al. “Chapter 11: Presenting findings”. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Ed. by J.J Deeks et al. Cochrane 2023, 2023.

[108] Tamalee Roberts et al. “Accuracy of the direct agglutination test for diagnosis of visceral leishmaniasis: a systematic review and meta-analysis”. In: *BMC Infectious Diseases* 23.1 (2023), p. 782. ISSN: 1471-2334. DOI: 10.1186/s12879-023-08772-1. URL: https://doi.org/10.1186/s12879-023-08772-1.

[109] United Nations. *Standard country or area codes for statistical use (M49)*. 2023. URL: https://unstats.un.org/unsd/methodology/m49/ (visited on 02/16/2024).

[110] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press, 2006. DOI: 10.1017/CBO9780511790942.

[111] P. Macaskill et al. “Chapter 10: Analysing and Presenting Results”. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. useful. 2010. Chap. 10. URL: http://srdta.cochrane.org/.

[112] Julian P. T. Higgins, Simon G. Thompson, and Jonathon J. Deeks. “Measuring inconsistency in a meta-analyses”. In: *BMJ* 327.7414 (2003), pp. 557–560. DOI: 10.1136/bmj.327.7414.557.

[113] William G Cochran. “The comparison of percentages in matched samples”. In: *Biometrika* 37.3/4 (1950), pp. 256–266.

[114] Julian P. T. Higgins and Simon G. Thompson. “Quantifying heterogeneity in a meta-analysis”. In: *Statistics in Medicine* 21.11 (2002), pp. 1539–1558. DOI: https://doi.org/10.1002/sim.1186. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1186.

[115] Paul T. von Hippel. “The heterogeneity statistic I2 can be biased in small meta-analyses”. In: *BMC Medical Research Methodology* 15.1 (2015), p. 35. DOI: 10.1186/s12874-015-0024-z. URL: https://doi.org/10.1186/s12874-015-0024-z.

[116] Michael Borenstein et al. “Basics of meta-analysis: I2 is not an absolute measure of heterogeneity”. In: *Research Synthesis Methods* 8.1 (2017), pp. 5–18. DOI: https://doi.org/10.1002/jrsm.1230. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1230.

[117] Joanna IntHout et al. “Plea for routinely presenting prediction intervals in meta-analysis”. In: *BMJ Open* 6.7 (2016). DOI: 10.1136/bmjopen-2015-010247. eprint: https://bmjopen.bmj.com/content/6/7/e010247.full.pdf. URL: https://bmjopen.bmj.com/content/6/7/e010247.

[118] Enzo Cerullo et al. “MetaBayesDTA: codeless Bayesian meta-analysis of test accuracy, with or without a gold standard”. In: *BMC Medical Research Methodology* 23.1 (2023), p. 127. ISSN: 1471-2288. DOI: 10.1186/s12874-023-01910-y. URL: https://doi.org/10.1186/s12874-023-01910-y.

[119] Geoffrey Jones et al. “Identifiability of Models for Multiple Diagnostic Testing in the Absence of a Gold Standard”. In: *Biometrics* 66.3 (2010), pp. 855–863. DOI: https://doi.org/10.1111/j.1541-0420.2009.01330.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2009.01330.x.

[120] John Collins and Minh Huynh. "Estimation of diagnostic test accuracy without full verification: a review of latent class methods". In: *Statistics in Medicine* 33.24 (2014), pp. 4141–4169. DOI: https://doi.org/10.1002/sim.6218. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6218.

[121] Matthew Stephens. "Dealing with label switching in mixture models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62.4 (2000), pp. 795–809. DOI: https://doi.org/10.1111/1467-9868.00265. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9868.00265. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00265.

[122] Jonah Gabry, Rok enovar, and Andrew Johnson. *cmdstanr: R Interface to 'CmdStan'*. https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org. 2023.

[123] Aki Vehtari et al. "Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC". English. In: *Bayesian Analysis* 16.2 (June 2021), pp. 667–718. ISSN: 1931-6690. DOI: 10.1214/20-BA1221.

[124] Andrew Gelman and Donald B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (1992), pp. 457–472. DOI: 10.1214/ss/1177011136. URL: https://doi.org/10.1214/ss/1177011136.

[125] Stan Development Team. *Stan User's Guide*. Version 2.34. 2022. URL: https://mc-stan.org.

[126] Sophie Jullien et al. "Diagnostic accuracy of multiplex respiratory pathogen panels for influenza or respiratory syncytial virus infections: systematic review and meta-analysis". In: *BMC Infectious Diseases* 22.1 (2022), p. 785. ISSN: 1471-2334. DOI: 10.1186/s12879-022-07766-9. URL: https://doi.org/10.1186/s12879-022-07766-9.

[127] Catriona Macrae et al. "Diagnostic performance of the IMMY cryptococcal antigen lateral flow assay on serum and cerebrospinal fluid for diagnosis of cryptococcosis in HIV-negative patients: a systematic review". In: *BMC Infectious Diseases* 23.1 (2023), p. 209. ISSN: 1471-2334. DOI: 10.1186/s12879-023-08135-w. URL: https://doi.org/10.1186/s12879-023-08135-w.

[128] Marta Valente et al. “Diagnosis of human leptospirosis: systematic review and meta-analysis of the diagnostic accuracy of the Leptospira microscopic agglutination test, PCR targeting Lfb1, and IgM ELISA to Leptospira fainei serovar Hurstbridge”. In: *BMC Infectious Diseases* 24.1 (2024), p. 168. ISSN: 1471-2334. DOI: 10.1186/s12879-023-08935-0. URL: https://doi.org/10.1186/s12879-023-08935-0.

[129] Yemisi Takwoingi et al. “Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data”. In: *Statistical Methods in Medical Research* 26.4 (2017), pp. 1896–1911. DOI: 10.1177/0962280215592269. URL: https://journals.sagepub.com/doi/abs/10.1177/0962280215592269%20https://journals.sagepub.com/doi/pdf/10.1177/0962280215592269.

[130] Solomon Faine, World Health Organization, et al. *Guidelines for the control of leptospirosis*. World Health Organization, 1982.

[131] World Health Organization. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition*. 2009. URL: https://www.ncbi.nlm.nih.gov/books/NBK143156/ (visited on 12/10/2023).

[132] D. Limmathurotsakul et al. “Fool’s gold: Why imperfect reference tests are undermining the evaluation of novel diagnostics: a reevaluation of 5 diagnostic tests for leptospirosis”. In: 1537-6591 (Electronic) ().

[133] Michael J. Maze et al. “Prevalence and risk factors for human leptospirosis at a hospital serving a pastoralist community, Endulen, Tanzania”. In: *PLOS Neglected Tropical Diseases* 17.12 (Dec. 2023), pp. 1–15. DOI: 10.1371/journal.pntd.0011855. URL: https://doi.org/10.1371/journal.pntd.0011855.

[134] Eliana L. Parra Barrera et al. “Demographic and clinical risk factors associated with severity of lab-confirmed human leptospirosis in Colombia, 20152020”. In: *PLOS Neglected Tropical Diseases* 17.7 (July 2023), pp. 1–17. DOI: 10.1371/journal.pntd.0011454. URL: https://doi.org/10.1371/journal.pntd.0011454.

[135] UK clinical research collaboration. *Aetiology*. URL: https://hrcsonline.net/research-activities/2-aetiology/ (visited on 02/27/2024).

[136] Merel van Diepen et al. “Prediction versus aetiology: common pitfalls and how to avoid them”. In: *Nephrology Dialysis Transplantation* 32.suppl$_2$ (Feb. 2017), pp. ii1–ii5. ISSN: 0931-0509. DOI: 10.1093/ndt/gfw459. eprint: https://academic.oup.com/ndt/article-pdf/32/suppl_2/ii1/11145847/gfw459.pdf. URL: https://doi.org/10.1093/ndt/gfw459.

[137] A. B. Doll R Fau - Hill and A. B. Hill. “Smoking and carcinoma of the lung; preliminary report”. In: 0007-1447 (Print) (). DOI: D-CLML:5120:2117:173:378:570OTO-NLM.

[138] M.L. Levin. “The occurrence of lung cancer in man”. In: *Acta - Unio Internationalis Contra Cancrum* 9.3 (1953), pp. 531–541. ISSN: 0365-3056. URL: https://eurekamag.com/research/050/682/050682788.php.

[139] The American Cancer Society. *Lung cancer*. URL: https://canceratlas.cancer.org/the-burden/lung-cancer/ (visited on 02/27/2024).

[140] R. G. Feachem and M. A. Koblinsky. “Interventions for the control of diarrhoeal diseases among young children: measles immunization”. In: *Bull World Health Organ* 61.4 (1983), pp. 641–52. ISSN: 0042-9686 (Print) 0042-9686.

[141] F. Calvo et al. “Diagnoses, syndromes, and diseases: a knowledge representation problem”. In: *AMIA Annu Symp Proc* 2003 (2003), p. 802. ISSN: 1559-4076.

[142] Karen L Kotloff et al. “Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study”. In: *The Lancet* 382.9888 (2013), pp. 209–222. ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(13)60844-2. URL: https://www.sciencedirect.com/science/article/pii/S0140673613608442.

[143] R. Palmer Beasley et al. “HEPATOCELLULAR CARCINOMA AND HEPATITIS B VIRUS: A Prospective Study of 22 707 Men in Taiwan”. In: *The Lancet* 318.8256 (1981). doi: 10.1016/S0140-6736(81)90585-7, pp. 1129–1133. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(81)90585-7. URL: https://doi.org/10.1016/S0140-6736(81)90585-7.

[144] S. D. Walter. “The Estimation and Interpretation of Attributable Risk in Health Research”. In: *Biometrics* 32.4 (1976), pp. 829–849. ISSN: 0006341X, 15410420. URL: http://www.jstor.org/stable/2529268 (visited on 10/06/2023).

[145] Jerome Cornfield. "A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix." In: *Journal of the National Cancer Institute* 11 6 (1951), pp. 1269–75. URL: https://api.semanticscholar.org/CorpusID:35687964.

[146] O. S. Miettinen. "Proportion of disease caused or prevented by a given exposure, trait or intervention". In: *Am J Epidemiol* 99.5 (1974), pp. 325–32. ISSN: 0002-9262 (Print) 0002-9262. DOI: 10.1093/oxfordjournals.aje.a121617.

[147] LAURA RODRIGUES and BETTY R KIRKWOOD. "Case-Control Designs in the Study of Common Diseases: Updates on the Demise of the Rare Disease Assumption and the Choice of Sampling Scheme for Controls". In: *International Journal of Epidemiology* 19.1 (Mar. 1990), pp. 205–213. ISSN: 0300-5771. DOI: 10.1093/ije/19.1.205. eprint: https://academic.oup.com/ije/article-pdf/19/1/205/1803590/19-1-205.pdf. URL: https://doi.org/10.1093/ije/19.1.205.

[148] Beverly Rockhill, Beth Newman, and Clarice Weinberg. "Use and misuse of population attributable fractions." In: *American journal of public health* 88.1 (1998), pp. 15–19.

[149] P Cole and B MacMahon. "Attributable risk percent in case-control studies." In: *Journal of Epidemiology & Community Health* 25.4 (1971), pp. 242–244. ISSN: 0007-1242. DOI: 10.1136/jech.25.4.242. eprint: https://jech.bmj.com/content/25/4/242.full.pdf. URL: https://jech.bmj.com/content/25/4/242.

[150] David E. Lilienfeld, Paul D. Stolley, and Abraham M. Lilienfeld. *Foundations of epidemiology*. 3rd. New York: Oxford University Press New York, 1994. ISBN: 0195050355; 9780195050356; 0195050363; 9780195050363.

[151] Zhenke Wu and Irena Chen. "Probabilistic cause-of-disease assignment using case-control diagnostic tests: A latent variable regression approach". In: *Statistics in Medicine* 40.4 (2021), pp. 823–841. DOI: https://doi.org/10.1002/sim.8804. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8804. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8804.

[152] K.J. Jager et al. "Confounding: What it is and how to deal with it". In: *Kidney International* 73.3 (2008), pp. 256–260. ISSN: 0085-2538. DOI: https://doi.org/10.1038/sj.ki.5002650. URL:

https://www.sciencedirect.com/science/article/pii/S0085253815529748.

[153] N. E. Breslow. "Statistics in Epidemiology: The Case-Control Study". In: *Journal of the American Statistical Association* 91.433 (1996), pp. 14–28. DOI: 10.1080/01621459.1996.10476660. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476660.

[154] Joan Acam et al. "Antimicrobial prescription patterns in East Africa: a systematic review". In: *Systematic Reviews* 12.1 (2023), p. 18. ISSN: 2046-4053. DOI: 10.1186/s13643-022-02152-7. URL: https://doi.org/10.1186/s13643-022-02152-7.

[155] Centers for disease control and prevention. *Serologic Tests for Dengue Virus*. URL: https://www.cdc.gov/dengue/healthcare-providers/testing/serologic-tests.html (visited on 03/25/2024).

[156] N. Raafat, S. D. Blacksell, and R. J. Maude. "A review of dengue diagnostics and implications for surveillance and control". In: *Trans R Soc Trop Med Hyg* 113.11 (2019). 1878-3503 Raafat, Nader Blacksell, Stuart D Maude, Richard J Wellcome Trust/United Kingdom Journal Article Research Support, Non-U.S. Gov't Review England 2019/08/01 Trans R Soc Trop Med Hyg. 2019 Nov 1;113(11):653-660. doi: 10.1093/trstmh/trz068., pp. 653–660. ISSN: 0035-9203 (Print) 0035-9203. DOI: 10.1093/trstmh/trz068.

[157] Brian K. Kigozi et al. "Investigating the etiology of acute febrile illness: a prospective clinic-based study in Uganda". In: *BMC Infectious Diseases* 23.1 (2023), p. 411. ISSN: 1471-2334. DOI: 10.1186/s12879-023-08335-4. URL: https://doi.org/10.1186/s12879-023-08335-4.

[158] Faisal Nooh et al. "Prevalence of fever of unidentified aetiology in East African adolescents and adults: a systematic review and meta-analysis". In: *Infectious Diseases of Poverty* 12.1 (2023), p. 55. ISSN: 2049-9957. DOI: 10.1186/s40249-023-01105-z. URL: https://doi.org/10.1186/s40249-023-01105-z.

[159] M. J. Maze et al. "The epidemiology of febrile illness in sub-Saharan Africa: implications for diagnosis and management". In: *Clinical Microbiology and Infection* 24.8 (2018). doi: 10.1016/j.cmi.2018.02.011, pp. 808–814. ISSN: 1198-743X. DOI: 10.1016/j.cmi.2018.02.011. URL: https://doi.org/10.1016/j.cmi.2018.02.011.

[160] Myron M. Levine et al. "The Global Enteric Multicenter Study (GEMS): Impetus, Rationale, and Genesis". In: *Clinical Infectious Diseases* 55.suppl$_4$ (Dec. 2012), S215–S224. ISSN: 1058-4838. DOI: 10.1093/cid/cis761. eprint: https://academic.oup.com/cid/article-pdf/55/suppl_4/S215/20911252/cis761.pdf. URL: https://doi.org/10.1093/cid/cis761.

[161] K. L. Kotloff et al. "The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study". In: 1537-6591 (Electronic) ().

[162] Tamer H. Farag et al. "Some Epidemiologic, Clinical, Microbiologic, and Organizational Assumptions That Influenced the Design and Performance of the Global Enteric Multicenter Study (GEMS)". In: *Clinical Infectious Diseases* 55.suppl$_4$ (Dec. 2012), S225–S231. ISSN: 1058-4838. DOI: 10.1093/cid/cis787. eprint: https://academic.oup.com/cid/article-pdf/55/suppl_4/S225/20911805/cis787.pdf. URL: https://doi.org/10.1093/cid/cis787.

[163] Sandra Panchalingam et al. "Diagnostic Microbiologic Methods in the GEMS-1 Case/Control Study". In: *Clinical Infectious Diseases* 55.suppl$_4$ (Dec. 2012), S294–S302. ISSN: 1058-4838. DOI: 10.1093/cid/cis754. eprint: https://academic.oup.com/cid/article-pdf/55/suppl_4/S294/20911223/cis754.pdf. URL: https://doi.org/10.1093/cid/cis754.

[164] William C. Blackwelder et al. "Statistical Methods in the Global Enteric Multicenter Study (GEMS)". In: *Clinical Infectious Diseases* 55.suppl$_4$ (Dec. 2012), S246–S253. ISSN: 1058-4838. DOI: 10.1093/cid/cis788. eprint: https://academic.oup.com/cid/article-pdf/55/suppl_4/S246/20911826/cis788.pdf. URL: https://doi.org/10.1093/cid/cis788.

[165] Laura L. Hammitt et al. "Addressing the Analytic Challenges of Cross-Sectional Pediatric Pneumonia Etiology Data". In: *Clinical Infectious Diseases* 64 (2017), S197–S204. ISSN: 10584838, 15376591. URL: http://www.jstor.org/stable/26374787 (visited on 10/13/2023).

[166] M. Deloria-Knoll et al. "Identification and selection of cases and controls in the Pneumonia Etiology Research for Child Health project". In: 1537-6591 (Electronic) ().

[167] Melissa M. Higdon et al. “Should Controls With Respiratory Symptoms Be Excluded From Case-Control Studies of Pneumonia Etiology? Reflections From the PERCH Study”. In: *Clinical Infectious Diseases* 64 (2017), S205–S212. ISSN: 10584838, 15376591. URL: http://www.jstor.org/stable/26374788 (visited on 10/13/2023).

[168] Z. Wu et al. “Partially latent class models for case-control studies of childhood pneumonia aetiology”. In: *J R Stat Soc Ser C Appl Stat* 65.1 (2016), pp. 97–114. ISSN: 0035-9254 (Print) 0035-9254. DOI: 10.1111/rssc.12101.

[169] David B. Dunson. “Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data”. In: *American Journal of Epidemiology* 153.12 (June 2001), pp. 1222–1226. ISSN: 0002-9262. DOI: 10.1093/aje/153.12.1222. eprint: https://academic.oup.com/aje/article-pdf/153/12/1222/287771/1222.pdf. URL: https://doi.org/10.1093/aje/153.12.1222.

[170] D. R. Murdoch et al. “Laboratory methods for determining pneumonia etiology in children”. In: 1537-6591 (Electronic) ().

[171] S. Bhatt et al. “The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015”. In: *Nature* 1476-4687 (Electronic) ().

[172] Heidi Hopkins. personal communication. Aug. 3, 2023.

[173] Zhenke Wu, Maria Deloria-Knoll, and Scott Zeger. “Nested Partially-Latent Class Models for Dependent Binary Data; Estimating Disease Etiology”. In: *Biostatistics (Oxford, England)* 18 (Oct. 2015). DOI: 10.1093/biostatistics/kxw037.

[174] Lawrence Joseph and Patrick Belisle. *BetaParmsFromQuantiles R Function*. 2017. URL: https://www.medicine.mcgill.ca/epidemiology/Joseph/PBelisle/BetaParmsFromQuantiles.html.

[175] Bogdan Grigore et al. “A comparison of two methods for expert elicitation in health technology assessments”. In: *BMC Medical Research Methodology* 16.1 (2016), p. 85. ISSN: 1471-2288. DOI: 10.1186/s12874-016-0186-3. URL: https://doi.org/10.1186/s12874-016-0186-3.

[176] Elizabeth S. Garrett and Scott L. Zeger. "Latent Class Model Diagnosis". In: *Biometrics* 56.4 (2000), pp. 1055–1067. DOI: https://doi.org/10.1111/j.0006-341X.2000.01055.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0006-341X.2000.01055.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.2000.01055.x.

[177] Neil Pearce. "Analysis of matched case-control studies". In: *BMJ* 352 (2016). DOI: 10.1136/bmj.i969. eprint: https://www.bmj.com/content/352/bmj.i969.full.pdf. URL: https://www.bmj.com/content/352/bmj.i969.

[178] Michael Betancourt. *Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo*. 2016. arXiv: 1604.00695 [stat.ME].

[179] V. Pyrgos et al. "Epidemiology of cryptococcal meningitis in the US: 1997-2009". In: 1932-6203 (Electronic) ().

[180] Etienne Simon-Loriere et al. "Autochthonous Japanese Encephalitis with Yellow Fever Coinfection in Africa". In: *New England Journal of Medicine* 376.15 (2017), pp. 1483–1485. DOI: 10.1056/NEJMc1701600. eprint: https://doi.org/10.1056/NEJMc1701600. URL: https://doi.org/10.1056/NEJMc1701600.

[181] Natalie E Hofmann et al. "Diagnostic Performance of Conventional and Ultrasensitive Rapid Diagnostic Tests for Malaria in Febrile Outpatients in Tanzania". In: *The Journal of Infectious Diseases* 219.9 (Nov. 2018), pp. 1490–1498. ISSN: 0022-1899. DOI: 10.1093/infdis/jiy676. eprint: https://academic.oup.com/jid/article-pdf/219/9/1490/28460077/jiy676.pdf. URL: https://doi.org/10.1093/infdis/jiy676.

[182] K. MARSH and S. KINYANJUI. "Immune effector mechanisms in malaria". In: *Parasite Immunology* 28.1-2 (2006), pp. 51–60. DOI: https://doi.org/10.1111/j.1365-3024.2006.00808.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3024.2006.00808.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3024.2006.00808.x.

[183] Luminex Corporation. personal communication. Jan. 15, 2024.

[184] T. J. Meerhoff et al. "Detection of multiple respiratory pathogens during primary respiratory infection: nasal swab versus nasopharyngeal aspirate using real-time polymerase chain reaction". In: *European Journal of Clinical Microbiology Infectious Diseases* 29.4 (2010), pp. 365–371. ISSN: 1435-4373. DOI: 10.1007/s10096-

009-0865-7. URL: https://doi.org/10.1007/s10096-009-0865-7.

[185] Katherine L. O'Brien et al. "Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study". In: *The Lancet* 394.10200 (2019). doi: 10.1016/S0140-6736(19)30721-4, pp. 757–779. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(19)30721-4. URL: https://doi.org/10.1016/S0140-6736(19)30721-4.

[186] Jj Auid-Orcid Nutor et al. "Predictors of high HIV+ prevalence in Mozambique: A complex samples logistic regression modeling and spatial mapping approaches". In: 1932-6203 (Electronic) ().

[187] Josée Mayxay M Fau - Castonguay-Vanier et al. "Causes of non-malarial fever in Laos: a prospective study". In: 2214-109X (Electronic) ().

[188] M. Sugeno et al. "Association between environmental factors and dengue incidence in Lao People's Democratic Republic: a nationwide time-series study". In: *BMC Public Health* 1471-2458 (Electronic) ().

[189] Simaly Phongmany et al. "Rickettsial Infections and Fever, Vientiane, Laos". In: *Emerg Infect Dis. 12(2):256-262.* 12.2 (2006). URL: https://stacks.cdc.gov/view/cdc/16020.

[190] Onanong Mayxay M Fau - Sengvilaipaseuth et al. "Causes of Fever in Rural Southern Laos". In: 1476-1645 (Electronic) ().

[191] T. Phanitchat et al. "Spatial and temporal patterns of dengue incidence in northeastern Thailand 2006-2016". In: *BMC Infectious Diseases* 1471-2334 (Electronic) ().

[192] K. Baltzell et al. "'Not all fevers are malaria': a mixed methods study of non-malarial fever management in rural southern Malawi". In: 1445-6354 (Electronic) ().

[193] Remco P. H. Peters et al. "A prospective study of bloodstream infections as cause of fever in Malawi: clinical predictors and implications for management". In: *Tropical Medicine & International Health* 9.8 (2004), pp. 928–934. DOI: https://doi.org/10.1111/j.1365-3156.2004.01288.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3156.2004.01288.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3156.2004.01288.x.

[194] World Health Organization. *Guidelines for The Diagnosis, Prevention and Management of Cryptococcal Disease in HIV-Infected Adults, Adolescents and Children: Supplement to the 2016 Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection*. URL: https://www.ncbi.nlm.nih.gov/books/NBK531449/ (visited on 03/07/2024).

[195] Hammad S. Ncho et al. "Typhoid Fever Outbreak Harare, Zimbabwe, October 2017February 2018". In: *MMWR Weekly* 68 (2019).

[196] Gaetan Thilliez et al. "Population structure of Salmonella enterica Typhi in Harare, Zimbabwe (2012x2013;19) before typhoid conjugate vaccine roll-out: a genomic epidemiology study". In: *The Lancet Microbe* 4.12 (2023). doi: 10.1016/S2666-5247(23)00214-8, e1005–e1014. ISSN: 2666-5247. DOI: 10.1016/S2666-5247(23)00214-8. URL: https://doi.org/10.1016/S2666-5247(23)00214-8.

[197] World Health Organisation. *Malaria vaccines (RTS,S and R21)*. 2024. URL: https://www.who.int/news-room/questions-and-answers/item/q-a-on-rts-s-malaria-vaccine (visited on 02/13/2024).

[198] Rui Gutierrez et al. "Aetiology of non-malaria acute febrile illness fever in children in rural Guinea-Bissau: a prospective cross-sectional investigation". In: *Frontiers in Epidemiology* 4 (2024). ISSN: 2674-1199. DOI: 10.3389/fepid.2024.1309149. URL: https://www.frontiersin.org/articles/10.3389/fepid.2024.1309149.

[199] Christopher Troeger et al. "Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 19902016: a systematic analysis for the Global Burden of Disease Study 2016". In: *The Lancet Infectious Diseases* 18.11 (2018), pp. 1191–1210. ISSN: 1473-3099. DOI: https://doi.org/10.1016/S1473-3099(18)30310-4. URL: https://www.sciencedirect.com/science/article/pii/S1473309918303104.

[200] Justice NA Jain H Schweitzer JW. *Respiratory Syncytial Virus Infection in Children*. 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK459215/ (visited on 02/13/2024).

[201] Bahaa Abu-Raya et al. "Why has the epidemiology of RSV changed during the COVID-19 pandemic?" In: *eClinicalMedicine* 61 (2023). doi: 10.1016/j.eclinm.2023.102089. ISSN: 2589-5370. DOI: 10.1016/

j.eclinm.2023.102089. URL: https://doi.org/10.1016/j.eclinm.2023.102089.

[202] Charles Poole. "Controls who experienced hypothetical causal intermediates should not be excluded from case-control studies". In: *American journal of epidemiology* 150.6 (1999), pp. 547–551.

[203] Rebecca Unterborn et al. "False-positive rapid diagnostic test for malaria in new world cutaneous leishmaniasis: a tale of two travelers". In: *Therapeutic Advances in Infectious Disease* 9 (2022), p. 20499361221097791. DOI: 10.1177/20499361221097791. eprint: https://doi.org/10.1177/20499361221097791. URL: https://doi.org/10.1177/20499361221097791.

[204] Ruth H. Keogh and D. R. Cox. *Case-Control Studies*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139094757.

[205] Lamine Ceesay Sj Fau - Koivogui et al. "Malaria Prevalence among Young Infants in Different Transmission Settings, Africa". In: 1080-6059 (Electronic) ().

[206] Yibing Ruan et al. "A comparison of approaches for estimating combined population attributable risks (PARs) for multiple risk factors". In: *Epidemiologic Methods* 9.1 (2020), p. 20190021. DOI: doi:10.1515/em-2019-0021. URL: https://doi.org/10.1515/em-2019-0021.

[207] A. B. Doll R Fau - Hill and A. B. Hill. "A study of the aetiology of carcinoma of the lung". In: 0007-1447 (Print) (). DOI: D-CLML:5323:22288:318:552OTO-NLM.

[208] Nicholas Midzi et al. "Accuracy of different diagnostic techniques for Schistosoma haematobium to estimate treatment needs in Zimbabwe: Application of a hierarchical Bayesian egg count model". In: *PLOS Neglected Tropical Diseases* 14 (Aug. 2020), e0008451. DOI: 10.1371/journal.pntd.0008451.

[209] Irena B Chen et al. *baker: An R package for Nested Partially-Latent Class Models*. 2022. arXiv: 2202.11778 [stat.ME].