



UNIVERSITAT^{DE}
BARCELONA

Aplicación de modelos bayesianos para estimar la prevalencia de enfermedad y la sensibilidad y especificidad de tests de diagnóstico clínico *sin gold standard*

Hélio Doyle Pereira da Silva



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution 3.0. Spain License.**



UNIVERSITAT DE
BARCELONA

Programa de Doctorado en Estadística
Métodos Estadísticos en Investigación Biomédica

Tesis Doctoral

**Aplicación de modelos bayesianos para estimar la prevalencia
de enfermedad y la sensibilidad y especificidad de tests de
diagnóstico clínico sin *gold standard***

Memoria presentada por:

Hélio Doyle Pereira da Silva

para optar al grado de:

Doctor por la Universitat de Barcelona

Visto bueno de los directores de la tesis:

Dra. Rosa Mari Abellana Sangrà

Dr. Carlos Ascaso Terrén

Departament de Fonaments Clínic

Facultat de Medicina

Barcelona, noviembre de 2016

Aplicación de modelos bayesianos para estimar la prevalencia de enfermedad y la sensibilidad y especificidad de tests de diagnóstico clínico sin *gold standard*

Esta es la versión definitiva de tesis elaborada por
Hélio Doyle Pereira da Silva, aprobada por la
comisión juzgadora y el tribunal en 17/11/2016,
con la valoración definitiva: “EXCELLENT”

Tribunal:

- Dr. Albert Sorribas Tello (President) - Universitat de Lleida
- Dr. Joan Pau Millet Vilanova (Vocal) - Agència de Salut Pública de Catalunya
- Dr. Josep LLuis Carrasco Jordan (Secretari) - Universitat de Barcelona
- Dr. Tomas Martinez Terrer (President Suplent) - Universitat de Zaragoza
- Dra. Montserrat Girabent Farrés (Vocal Suplent) - Universitat Internacional de Catalunya
- Dr. Jaume Canela Soler (Secretari Suplent) - Universitat de Barcelona



VALORACIÓ DEFINITIVA TESI DOCTORAL

HÉLIO DOYLE PEREIRA DA SILVA

Data de lectura: 17 de novembre de 2016

Títol: "Aplicación de modelos bayesianos para estimar la prevalencia de enfermedad y la sensibilidad y especificidad de tests diagnóstico clínico sin gold standard".

Segons el resultat de l'escrutini dels vots secrets emesos pels membres del tribunal de la tesi, la valoració definitiva és:

"EXCEL. LEIUT"

El President de la Comissió de Doctorat
Josep M. Grau Junyent

Barcelona, 17 de novembre de 2016



*Dedicado a
mi familia*

Agradecimientos

A mi directora de tesis, Dra. Rosa Abellana por el apoyo personal y técnico en todas las etapas de mis estudios de doctorado. Por la disposición en la resolución de las dudas y por exponer ideas y consejos indispensables para el desarrollo de los artículos y de la tesis de doctorado.

A mi director, Dr. Carlos Ascaso, por el aprendizaje, por el estímulo constante debido a su espíritu dinámico y principalmente por sus palabras de motivación que fueron decisivas para la continuidad y la finalización del proceso de estudios en este doctorado.

Al personal del Departamento de Salud Pública de la Universidad de Barcelona (UB), por la participación directa o indirecta especialmente agradezco al Dr. Jaume por el aporte moral.

A Dr. Tomàs y Dra. Alessandra por haber proporcionado los datos indispensables para los artículos que componen esta tesis y por el apoyo a los respectivos temas específicos de tuberculosis y parásitos intestinales.

A la Dra. Ana Maria Nogales Vasconcelos, mi adorable profesora de Demografía del Departamento de Estadística de la Universidad de Brasília, UnB - Brasil, y su esposo Dr. David Duarte Lima, profesor del Departamento de Medicina de la misma universidad, que fueron los primeros a depositar apoyo en mis sueños de estudiar en una Universidad fuera de Brasil.

A mi amigo y profesor Alan Ricardo da Silva, por su voluntad de enseñarme a trabajar con SAS en las aplicaciones estadísticas y por la confianza en mis capacidades.

A mi mejor jefa Laura Bernardes da Silva, coordinadora del INEP-MEC, Brasil.

A mis amigos de estudios en los primeros pasos con estadística en la Universidad de Brasília.

A mi familia que vive en Brasil, hace 10 años que no los veo y siempre han sido comprensivos por mi ausencia. Especialmente agradezco a mi madre Anísia, mi hermano Hilber, a mis sobrinas Alanis y Camila y a su madre Carmen, a mi hermana Hilma y a su marido Valdir y a mis sobrinos William y Liliane.

A la familia de mi esposa que vive en Irán, especialmente a mis suegros Mohammad Shams Abadi y Moazam Shams Abadi por el apoyo incondicional necesario para la finalización

de mi tesis.

Especialmente al grupo de “Nenas” (Lunna, Elsa y Aiko) y a los padres, nuestros grandes amigos y nuestra pequeña familia en Barcelona: César Rozas y Maria Angeles Cruz, Pilar Cornejo y Lluís Puche que me apoyaron en distintas fases de esta aventura.

A mis queridos amigos Jimmie Matsson, Nancy Jacome, Jeaneth Lopez y Grevy Guzman que participaron y me apoyaron de diferentes formas y en momentos especiales.

A mi esposa Safa Shams Abadi, por creer siempre en que todo es posible. Por las indispensables charlas necesarias para el equilibrio familiar y por no perder el entusiasmo y la fuerza, por incentivar mis capacidades personales para que sobrepasara mis límites.

A mi linda hija Lunna Doyle Abadi (con 8 años en esta fecha) que siempre tiene una novedad que me hace muy feliz y me ayuda para seguir adelante en el camino de la vida.

Esta tesis se realizó con base en dos artículos publicados (vide Apéndices **A** y **B**)

Artículo I (A) Publicación en la revista *Statistics in Medicine*

Hélio Doyle Pereira da Silva, Carlos Ascaso, Alessandra Queiroga Gonçalves, Patricia Puccinelli Orlandi and Rosa Abellana. (2017).

A Bayesian approach to model the conditional correlation between several diagnostic tests and various replicated subjects measurements.

Statistics in Medicine

DOI:10.1002/sim.7339

<https://www.ncbi.nlm.nih.gov/pubmed/28543307>

Artículo II (B) Publicación en la revista *PlosONE*

Perez-Porcuna TM, Pereira-da-Silva HD, Ascaso C, Malheiro A, Bühner S, Martinez-Espinosa F, et al. (2016).

Prevalence and Diagnosis of Latent Tuberculosis Infection in Young Children in the Absence of a Gold Standard.

PLoS ONE 11(10): e0164181.

DOI:10.1371/journal.pone.0164181

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0164181>

Curriculum Vitae

<http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4444481A1>

<http://heliodyle.wixsite.com/analysestdata>

Resumen

Dos objetivos claves de la investigación diagnóstica son estimar con exactitud y precisión la prevalencia de la enfermedad y la sensibilidad y especificidad de tests diagnósticos. Se han desarrollado modelos de clases latentes que tienen en cuenta la correlación entre las medidas de los individuos determinadas con diferentes tests con el fin de diagnosticar enfermedades para las cuales no están disponibles tests *gold standard*. En algunos estudios clínicos, se hacen varias medidas del mismo individuo con el mismo test en las mismas condiciones y, por tanto, las mediciones replicadas para cada individuo no son independientes. En esta tesis se propone una extensión Bayesiana del modelo de clases latentes de efectos aleatorios de Gauss para ajustar a los datos de tests con resultados binarios y con medidas replicadas por individuo. Se describe una aplicación que utiliza los datos recogidos de personas infectadas por parásitos intestinales *Hookworm* llevada a cabo en el municipio de Presidente Figueiredo, Estado de Amazonas en Brasil. Además, a través de un estudio de simulación se comparó el desempeño del modelo propuesto con los modelos actuales (el modelo de efectos aleatorios individuo y modelos de dependencia e independencia condicional). Como era de esperar, el modelo propuesto presenta una mayor exactitud y precisión en las estimaciones de prevalencia, sensibilidad y especificidad.

Para un control adecuado de las enfermedades la Organización Mundial de la Salud ha propuesto el diagnóstico y tratamiento de la infección tuberculosa latente (LTBI) en grupos de riesgo de desarrollar la enfermedad, como los niños. No existe un test *gold standard* para el diagnóstico de la infección latente. Los modelos estadísticos basados en la estimación de clases latentes permiten la evaluación de la prevalencia de la infección y la validez de los tests utilizados en ausencia de un *gold standard*. Se realizó un estudio transversal con niños de hasta 6 años de edad que habían sido vacunados con la BCG en Manaus, Amazonas-Brasil. El objetivo de dicho estudio fue estimar la prevalencia de la infección latente en los niños pequeños en contacto con un caso índice de tuberculosis en el hogar (TB-HCC) y determinar la validez y la seguridad del test cutáneo de tuberculina (TST) y QuantiFERON-TB Gold-in-tube (QFT), utilizando modelos de clases latentes. Para las estimaciones, en una primera fase se consideró la correlación entre los dos tests, y en la segunda fase se consideró la prevalencia en función

de la intensidad y del tiempo de exposición. El cincuenta por ciento de los niños con TB-HCC tenía LTBI, con la prevalencia en función del tiempo y la intensidad de la exposición al caso índice. La sensibilidad y la especificidad de TST fueron del 73 % [intervalo de confianza del 95 % (IC): 53-91] y el 97 % (IC del 95 %: 89-100), respectivamente, frente al 53 % (IC del 95 %: 41-66) y el 81 % (IC del 95 %: 71-90) para QFT. El valor predictivo positivo de TST en niños con TB-HCC fue del 91 % (IC del 95 %: 61-99), y para QFT fue del 74 % (IC del 95 %: 47-95). Este es uno de los primeros estudios que usa modelos de clases latentes para estimar la prevalencia de la infección por *M. tuberculosis* en niños y los parámetros de sus principales tests diagnósticos. Los resultados sugieren que los niños en contacto con un caso índice tienen un alto riesgo de infección. La validez y los valores predictivos no mostraron diferencias significativas según el test aplicado. El uso combinado de los dos tests en nuestro estudio mostró una sutil mejoría en el diagnóstico de la LTBI.

Keywords Bayesian analysis; diagnostic tests; diagnosis; gold standard; random effects model; sensitivity; specificity; prevalence; accuracy; latent tuberculous infection; children; latent class model; tuberculin skin test; QuantiFERON-TB Gold in-tube.

Abstract

Two key aims of diagnostic research are to accurately and precisely estimate disease prevalence and test sensitivity and specificity. Latent class models have been proposed that consider the correlation between subject measures determined by different tests in order to diagnose diseases for which gold standard tests are not available. In some clinical studies, several measures of the same subject are made with the same test under the same conditions (replicated measurements) and thus, replicated measurements for each subject are not independent. In the present study, we propose an extension of the Bayesian latent class Gaussian random effects model to fit the data with binary outcomes for tests with replicated subject measures. We describe an application using data collected on hookworm infection carried out in the municipality of Presidente Figueiredo, Amazonas State, Brazil. In addition, the performance of the proposed model was compared with that of current models (the subject random effects model and the conditional (in)dependent model) through a simulation study. As expected, the proposed model presented better accuracy and precision in the estimations of prevalence, sensitivity and specificity.

For adequate disease control the World Health Organization has proposed the diagnosis and treatment of latent tuberculous infection (LTBI) in groups of risk of developing the disease such as children. There is no gold standard (GS) test for the diagnosis of LTBI. Statistical models based on the estimation of latent class allow evaluation of the prevalence of infection and the accuracy of the tests used in the absence of a GS. We conducted a cross-sectional study with children up to 6 years of age who had been vaccinated with the BCG in Manaus, Amazonas-Brazil. The objective of this study was to estimate the prevalence of LTBI in young children in contact with a household case of tuberculosis (TB-HCC) and determine the accuracy and precision of the Tuberculin Skin Test (TST) and QuantiFERON-TB Gold in-tube (QFT) using the latent class model. Fifty percent of the children with TB-HCC had LTBI, with the prevalence depending on the intensity and length of exposure to the index case. The sensitivity and specificity of TST were 73 % [95 % confidence interval (CI): 53-91] and 97 % (95 % CI: 89-100), respectively, versus 53 % (95 % CI: 41-66) and 81 % (95 % CI: 71-90) for QFT. The positive predictive value of TST in children with TB-HCC was 91 % (95 % CI: 61-99), and for

QFT was 74 % (95 % CI: 47-95). This is one of the first studies to estimate the prevalence of *M. tuberculosis* infection in children and the parameters of its main diagnostic tests by latent class model. The results suggest that children in contact with an index case have a high risk of infection. The accuracy and the predictive values did not show significant differences according to the test applied. Combined use of the two tests in our study showed scarce improvement in the diagnosis of LTBI.

Keywords Bayesian analysis; diagnostic tests; diagnosis; gold standard; random effects model; sensitivity; specificity; prevalence; accuracy; latent tuberculous infection; children; latent class model; tuberculin skin test; QuantiFERON-TB Gold in-tube.

Índice general

Índice de figuras	XIX
Índice de tablas	XXI
1 Introducción	25
1.1 Test diagnóstico	25
1.1.1 Validez y seguridad del test diagnóstico	27
1.1.2 Test de diagnóstico sin <i>gold standard</i>	30
1.2 Bases teóricas de los métodos Bayesianos	32
1.2.1 Teorema de Bayes	32
1.2.2 Modelo Bayesiano	33
1.2.3 Algoritmos MCMC: Muestreo de Gibbs y Metropolis-Hastings	35
1.2.4 El diagnóstico de convergencia	38
1.3 Estimación de la prevalencia y validez de tests diagnósticos en ausencia de <i>gold standard</i> mediante inferencia Bayesiana	43
1.3.1 Caso de un solo test diagnóstico	46
1.3.1.1 Verosimilitud para el caso de un solo test diagnóstico	46
1.3.1.2 Definición de las distribuciones prior para un solo test diagnóstico	46
1.3.1.3 Distribución posterior para un test diagnóstico	47
1.3.2 Caso de dos tests diagnósticos independientes	50
1.3.2.1 Verosimilitud para el caso de dos tests diagnósticos	50
1.3.2.2 Definición de las distribuciones prior para dos tests diagnósticos	53
1.3.2.3 Distribución posterior para dos tests diagnósticos	53

1.3.3	Caso de dos tests diagnósticos dependientes	54
1.3.3.1	Modelo dependiente con efectos fijos	55
1.3.3.2	Verosimilitud para el modelo con efectos fijos	56
1.3.3.3	Prior para los parámetros del modelo con efectos fijos	56
1.3.3.4	Distribución posterior para el modelo con efectos fijos	58
1.3.3.5	Modelo dependiente con efectos aleatorios individuo	61
1.3.3.6	Verosimilitud para el modelo con efectos aleatorios individuo	62
1.3.3.7	Prior para los parámetros del modelo con efectos aleatorios	62
1.3.3.8	Distribución posterior para el modelo con efectos aleatorios	66
1.4	Inclusión de covariables en el modelo	67
2	Objetivos	69
3	Objetivo 1. Modelo individuo-test	71
3.1	Material y métodos	71
3.1.1	Objetivo 1.(a) Extensión del modelo bayesiano con efectos aleatorios para estimar la prevalencia, la sensibilidad y la especificidad de dos o más tests diagnósticos en ausencia de <i>gold standard</i> para el caso que existan medidas repetidas de los tests por individuo	71
3.1.1.1	Modelo con efecto aleatorio Individuo-test	72
3.1.1.2	Estimación de los parámetros del modelo	74
3.1.1.3	Distribución prior	74
3.1.1.4	Distribución posterior	75
3.1.2	Objetivo 1.(b) Aplicación de la extensión para evaluar el funcionamiento de dos tests para diagnosticar la presencia de parásitos intestinales	76
3.1.2.1	Información prior	76
3.1.3	Objetivo 1.(c) Simulación de datos para evaluar la calidad de los modelos que estiman la validez de los tests diagnósticos	78
3.2	Resultados	83
3.2.1	Datos <i>Hookworm</i>	83
3.2.2	Simulación	85
3.3	Discusión	92
4	Objetivo 2. Estudio de la prevalencia y validez de los tests (sin gold-standard) en el diagnóstico de la Tuberculosis en niños	95
4.1	Material y Métodos	96

4.2	Resultados	101
4.3	Discusión	107
5	Conclusiones	111
	Apéndice	113
A	Paper I	113
B	Paper II	131
C	Código WinBUGS para el modelo propuesto (SUB×TST)	145
D	Código WinBUGS para el modelo de efectos fijos con estratificación de la población (dos prevalencias)	146
E	Código WinBUGS para el modelo de efecto aleatorio individuo incluyendo covariables en la prevalencia	148
F	Funciones R utilizadas	151
	Bibliografía	155

Índice de figuras

1.1	Zona amarilla: región sujeta a mala clasificación de individuos	26
1.2	El diagnóstico de convergencia de Gelman-Rubin	40
1.3	El diagnóstico de convergencia - autocorrelación	40
1.4	El histórico de las cadenas generadas con el <i>Trace-plot</i>	41
1.5	El diagnóstico de convergencia con el <i>Dinamic Trace-Plot</i>	41
1.6	Densidad de los datos finales generados a partir de tres cadenas convergentes .	42
1.7	Densidad de los datos finales generados a partir de tres cadenas no convergentes	42
1.8	Covarianzas para la combinación de dos tests condicionalmente dependientes. .	57
1.9	Gráfico de contorno para S_1 en el plano dado por sus respectivas coordenadas (a_{11}, b_1)	65
3.1	Sesgo relativo calculado en los escenarios I, II, III y IV de acuerdo con los cuatro modelos; SUB×TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio indi- viduo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente.	86
3.2	Desviación típica calculada en los escenarios I, II, III y IV de acuerdo con los cuatro modelos; SUB×TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio indi- viduo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente.	87

3.3	Error cuadrático medio calculado en los escenarios I, II, III y IV de acuerdo con los cuatro modelos; $SUB \times TST$, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB , tiene en cuenta sólo el efecto aleatorio individuo; DEP , modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND , modelo independiente condicional al verdadero estado del paciente.	89
3.4	Resultados de la simulación para los escenarios I, II, III and IV de acuerdo con los cuatro modelos; $SUB \times TST$, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB , tiene en cuenta sólo el efecto aleatorio individuo; DEP , modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND , modelo independiente condicional al verdadero estado del paciente. $Prev$, prevalencia; S_1 , sensibilidad del test 1; S_2 , sensibilidad del test 2; C_1 , especificidad del test 1; C_2 , especificidad del test 2. El verdadero valor del parámetro está en la línea discontinua.	91
4.1	Diagrama de flujo de la inclusión de los niños en el estudio.	97
4.2	Prevalencia en los tiempos de exposición de 0, 1, 3, 6 y 12 meses y MTC-score con intensidades 4, 8, 13 y 15 de exposición.	105
4.3	Valores predictivos positivos y negativos para QFT en los tiempos de exposición de 0, 1, 3, 6 y 12 meses y MTC-score con intensidades 4, 8, 13 y 15 de exposición.	106

Índice de tablas

1.1	Clasificaciones de individuos cuando es posible verificar su verdadero estado de salud	27
1.2	Datos latentes y observados para el caso de un test de diagnóstico en ausencia de <i>gold standard</i>	49
3.1	Verdaderos valores de los parámetros de acuerdo con los escenarios (I-IV) . . .	79
3.2	Distribución prior y opinión de los expertos considerados en los escenarios . .	82
3.3	Resultados de las estimaciones de los parámetros para los datos <i>Hookworm</i> . . .	84
3.4	Sesgo relativo, desviación típica y error cuadrático medio de las estimaciones. .	88
3.5	Tasa de cobertura para los cuatro escenarios	90
4.1	Análisis descriptivo y comparaciones entre los resultados de los tests QFT y TST.	102
4.2	Resultados de la prevalencia de LTBI de acuerdo con la exposición a un caso índice de TB y la sensibilidad y la especificidad de los tests QFT y TST (primera fase).	103
4.3	Probabilidad de LTBI de acuerdo con los resultados de los dos tests en el modelo de clases latentes (primera fase).	104
4.4	Probabilidad de no LTBI (LTBI ^c) de acuerdo con los resultados de los dos tests en el modelo de clases latentes (primera fase).	104

Introducción

1.1 Test diagnóstico

El estudio diagnóstico se realiza para prevenir, detectar o confirmar una enfermedad, una infección o cualquier otra condición que modifica el estado de salud de un paciente. Los médicos usan criterios objetivos y subjetivos a partir de la observación de los signos y síntomas del individuo, y se apoyan en los diferentes exámenes ante la sospecha clínica para finalmente decidir acerca del estado de salud del paciente. La exploración física y las exploraciones complementarias obtenidas de diversas fuentes de información constituyen el conjunto de procedimientos de diagnóstico. Un test diagnóstico se puede crear a partir de la información obtenida de la historia clínica, del análisis bioquímico y genético de individuos.

El test diagnóstico es un instrumento importante en la práctica médica y en la investigación clínica por ser un mecanismo con el objetivo de obtener una respuesta realista y objetiva del estado de salud del paciente. Por tanto, constituye una herramienta para tomar una decisión médica y se espera identificar correctamente el estado de salud de un individuo. El conocimiento sobre la calidad de los tests de diagnóstico es de vital importancia para valorar el estado del paciente. Esta calidad se evalúa mediante los parámetros de validez y de seguridad.

El resultado del test diagnóstico se puede medir en escala numérica y posteriormente reducirlo a categorías de fácil interpretación, capaz de revelar de forma práctica el estado de los pacientes. En la figura 1.1 se muestra la distribución de los valores de una característica de

los pacientes diferenciando entre enfermos y no enfermos. El punto de corte, P_c , o valor de discriminación genera dos regiones definidas en color amarillo que corresponden a los pacientes mal clasificados por el test. Cuanto más se solapan la distribución de valores entre enfermos y no enfermos, más difícil es de crear un test con buena calidad de diagnóstico.

Los tests con resultado binario (dos categorías), aplicados a una muestra aleatoria de la población evaluada, sirven para hacer un diagnóstico diferencial, es decir, clasificar entre enfermo o no enfermo, infectado o no infectado, o para identificar presencia o ausencia de un factor biológico asociado a la condición de salud del paciente.

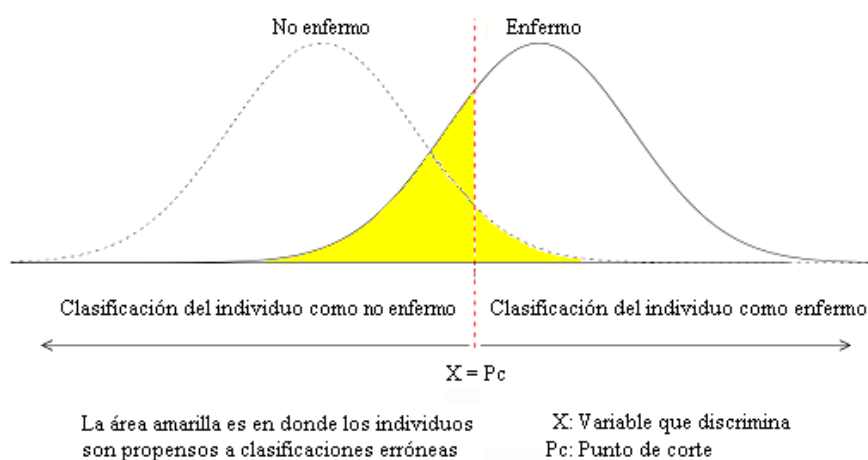


Figura 1.1: Zona amarilla: región sujeta a mala clasificación de individuos

Los pacientes con signos o síntomas de una enfermedad generalmente son sometidos al “mejor” test de referencia (*reference standard*) disponible y preferiblemente a un test libre de errores de clasificación conocido como *gold standard*.

La prevalencia de una enfermedad es la proporción de individuos enfermos en la población de estudio. Se denota prevalencia $\pi = P(G = 1)$, donde G es el test *gold standard* binario con valor 1 que indica enfermo y 0 que indica no enfermo.

La clasificación de los individuos de una población está sujeta a errores en ausencia de *gold standard*. Los errores que interfieren en la correcta clasificación pueden ser aleatorios o sistemáticos. Error aleatorio ocurre cuando alguna característica desconocida (o no contemplada) del individuo interacciona o no reacciona debido a la intrínseca calidad tecnológica del test, o cuando algún equívoco involuntario es provocado por el evaluador. El error sistemático ocurre, por ejemplo, cuando el aparato usado para recoger la información del individuo no esté cali-

brado. En este caso, el resultado de la medición estará equivocado sistemáticamente para todos los individuos sometidos al test. Las propiedades del test que permiten conocer su calidad son la validez y la seguridad.

1.1.1 Validez y seguridad del test diagnóstico

Pocos tests diagnósticos identifican con certeza si el paciente tiene o no una enfermedad y para la mayoría de los estudios de diagnósticos es poco probable clasificar correctamente a los individuos como verdaderamente enfermos o no enfermos [Reitsma et~al., 2009, Rujes et~al., 2007, Valenstein, 1990]. La validez y la seguridad de un test diagnóstico dependen de su capacidad para detectar correctamente la presencia o ausencia de la enfermedad que se estudia, lo que se expresa principalmente con los parámetros sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo.

Validez

La validez del test diagnóstico es el grado en que un test mide efectivamente la enfermedad, o una medida que permite evaluar cuánto el resultado del test se aproxima al valor real. La validez se mide mediante los parámetros de sensibilidad y especificidad. La sensibilidad es la probabilidad que un enfermo sea clasificado como positivo por el test. La especificidad es la probabilidad que un no enfermo sea clasificado como negativo por el test. En la tabla 1.1 se presenta la clasificación de los individuos en función del test.

		<i>Gold standard</i>	
		Enfermo	No enfermo
Test	Positivo	Verdadero Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadero Negativo (VN)

Tabla 1.1: Clasificaciones de individuos cuando es posible verificar su verdadero estado de salud

Se puede observar que los individuos enfermos se dividen entre los verdaderos positivos,

VP, (enfermos clasificados correctamente por el test) y los falsos negativos, FN, (los enfermos clasificados incorrectamente por el test). Los individuos no enfermos se clasifican entre los verdaderos negativos, VN, (los no enfermos clasificados correctamente) y los falsos positivos, FP, (no enfermos clasificados incorrectamente). La sensibilidad es la proporción de verdaderos positivos del total de individuos enfermos, $(S = \frac{VP}{VP+FN})$. Otra manera de definir la sensibilidad de forma matemática es a partir de las probabilidades condicionadas. Si se denota el resultado del test por T con valor 1 si el resultado es positivo y 0 si es negativo, entonces la sensibilidad es igual a $S = P(T = 1|G = 1)$. La especificidad es la proporción de verdaderos negativos del total de individuos no enfermos, $(C = \frac{VN}{VN+FP})$ o $C = P(T = 0|G = 0)$.

Seguridad

La certeza de que un test predica la presencia o ausencia de la enfermedad en un paciente es lo que define la seguridad del test. La seguridad viene determinada por el valor predictivo de un resultado positivo o negativo.

El valor predictivo positivo, **VPP**, se define como la proporción de verdaderos positivos del total de individuos con resultado positivo en el test $(VPP = \frac{VP}{VP+FP})$. El VPP se define matemáticamente como la probabilidad de ser verdaderamente enfermo condicionada al resultado positivo en el test, denotado por $VPP = P(G = 1|T = 1)$. En términos de los parámetros π, S, C , la expresión de los valores predictivos positivos es:

$$\begin{aligned}
 VPP &= P(G = 1|T = 1) \\
 &= \frac{P(G = 1)P(T = 1|G = 1)}{P(G = 1)P(T = 1|G = 1) + P(G = 0)P(T = 1|G = 0)} \\
 &= \frac{P(G = 1)P(T = 1|G = 1)}{P(G = 1)P(T = 1|G = 1) + (1 - P(G = 1))(1 - P(T = 0|G = 0))} \\
 &= \frac{\pi S}{\pi S + (1 - \pi)(1 - C)}
 \end{aligned} \tag{1.1}$$

En cuanto al valor predictivo negativo, **VPN**, es la proporción de verdaderos negativos entre el total de individuos con resultado negativo $(VPN = \frac{VN}{VN+FN})$. Matemáticamente, es la probabilidad del paciente ser verdaderamente no enfermo condicionada al resultado negativo en el, denotado por $VPN = P(G = 0|T = 0)$. En términos de los parámetros π, S, C , la

expresión de los valores predictivos negativos es:

$$\begin{aligned}
 VPN &= P(G = 0|T = 0) \\
 &= \frac{P(G = 0)P(T = 0|G = 0)}{P(G = 1)P(T = 0|G = 1) + P(G = 0)P(T = 0|G = 0)} \\
 &= \frac{P(G = 0)P(T = 0|G = 0)}{P(G = 1)(1 - P(T = 1|G = 1)) + (1 - P(G = 1))P(T = 0|G = 0)} \\
 &= \frac{(1 - \pi)C}{\pi(1 - S) + (1 - \pi)C}
 \end{aligned} \tag{1.2}$$

Los valores predictivos positivos y negativos se pueden interpretar como riesgos [Bossuyt, 2008]: el riesgo de enfermedad en los que tienen resultados positivos y el riesgo en los que tienen resultados negativos.

Los parámetros de validez son menos útiles en la práctica clínica que los parámetros de seguridad. Interesa a la práctica clínica saber que probabilidad hay de que un paciente esté enfermo si es que el test da positivo. No que probabilidad habría de que el test salga positivo o negativo si el paciente está enfermo o no. Se podría decir que la sensibilidad y especificidad se utilizan en estudios de validación de nuevo test diagnóstico.

Los valores predictivos por el contrario van en el mismo sentido en que va el razonamiento clínico, es decir, al aplicar un test, con resultado positivo entonces ¿qué probabilidad hay de que el paciente realmente esté enfermo? o con resultado negativo, ¿qué probabilidad hay de que realmente el paciente no esté enfermo?

Otra diferencia importante es que la sensibilidad y la especificidad son parámetros intrínsecos del test, es decir, no deberían cambiar si es que el test es de la misma calidad y este es aplicado de la misma manera en diferentes poblaciones. Por el contrario, los valores predictivos positivos y negativos se modifican en función de la prevalencia de la enfermedad en la población.

En ausencia de *gold standard* se recomienda preferiblemente el uso de dos o más tests combinados, que en conjunto pueden disminuir los errores en la clasificación de los individuos [Moons et al., 2004, 1999]. En estos casos puede ser de interés calcular un valor predictivo positivo y un valor predictivo negativo resultante de tests combinados. Las expresiones siguientes

representan el cálculo de los valores predictivos para el caso de dos tests combinados:

para los valores predictivos positivos,

$$\begin{aligned}
 VPP_{12} = & \frac{\pi S_1 S_2}{\pi S_1 S_2 + (1 - \pi)(1 - C_1)(1 - C_2)} + \frac{\pi S_1(1 - S_2)}{\pi S_1(1 - S_2) + (1 - \pi)(1 - C_1)C_2} \\
 & + \frac{\pi(1 - S_1)S_2}{\pi(1 - S_1)S_2 + (1 - \pi)C_1(1 - C_2)} + \frac{\pi(1 - S_1)(1 - S_2)}{\pi(1 - S_1)(1 - S_2) + (1 - \pi)C_1C_2}
 \end{aligned} \tag{1.3}$$

y para los valores predictivos negativos,

$$\begin{aligned}
 VPN_{12} = & \frac{(1 - \pi)C_1C_2}{\pi(1 - S_1)(1 - S_2) + (1 - \pi)C_1C_2} + \frac{(1 - \pi)C_1(1 - C_2)}{\pi(1 - S_1)S_2 + (1 - \pi)C_1(1 - C_2)} \\
 & + \frac{(1 - \pi)(1 - C_1)C_2}{\pi S_1(1 - S_2) + (1 - \pi)(1 - C_1)C_2} + \frac{(1 - \pi)(1 - C_1)(1 - C_2)}{\pi S_1S_2 + (1 - \pi)(1 - C_1)(1 - C_2)}
 \end{aligned}$$

1.1.2 Test de diagnóstico sin *gold standard*

El paradigma clásico para la evaluación de la validez de un test es comparar los resultados del test índice con un test de referencia considerado *gold standard*. Cuando el test de referencia no refleja la “verdad” de manera adecuada, las estimaciones de sensibilidad y especificidad son sesgadas.

Se puede decir que la ausencia de *gold standard* ocurre en 3 ocasiones:

1. cuando no hay ningún test de referencia para una determinada enfermedad;
2. en los casos de error de medición del test de referencia;
3. cuando el procedimiento no se puede realizar en la práctica porque es invasivo o costoso aplicar a todos los pacientes del estudio, o cuando algunos pacientes rechazan ser sometidos al test.

Cuando el test de referencia no refleja la “verdad” de manera adecuada, la validez y la seguridad del test están sujetos a errores. De hecho, cuanto mayor es la desviación de la “verdad” del *reference standard*, peores son las estimaciones de la sensibilidad y especificidad del test índice. Esto es conocido por sesgo de *reference standard* [Bossuyt, 2008, Rujes et-al., 2007, Whiting et-al., 2004].

Es importante para la práctica clínica y salud pública estimar la prevalencia de las enfermedades y conocer la validez (la sensibilidad y especificidad) y la seguridad (valor predictivo positivo y negativo) de las técnicas o tests diagnósticos a utilizar. Pero cuando se evalúa un nuevo test diagnóstico respecto a otros existentes sin poder considerar que ninguno de ellos es el test de referencia, *gold standard*, la metodología estadística clásica presenta limitaciones que en general introducen sesgo en las estimaciones de las propiedades de los tests y en la prevalencia de la enfermedad a causa de la mala clasificación.

Joseph et al. [1995] plantea aproximaciones Bayesianas para la estimación de la prevalencia y de los parámetros de validez y seguridad de tests de diagnóstico. Los métodos bayesianos permiten obtener la distribución de los parámetros y capturar la incertidumbre en las estimaciones. Con el fin de simplificar el análisis estadístico, se asumía que los resultados de cada individuo a los diferentes tests de diagnóstico eran independientes. Pero varios autores demostraron la importancia de tener en cuenta esta dependencia para obtener estimaciones insesgadas de la prevalencia y de la validez de los tests [Torrance-Rynard & Walter, 1997, Vacek, 1985, Walter et al., 2012]. Qu et al. [1996] propuso un modelo de clases latentes con efectos aleatorios que tiene en cuenta la dependencia condicional entre los tests de diagnóstico estimado mediante máxima verosimilitud. El modelo de efectos aleatorios determina la correlación entre tests mediante la medición de la “intensidad” o “severidad” de la enfermedad en cada individuo. Más tarde, Dendukuri & Joseph [2001] proponen una aproximación bayesiana para este modelo, y definen dos propuestas de modelos para tener en cuenta la correlación entre los tests. Una propuesta modela la dependencia entre los tests utilizando la covarianza entre tests, modelo de efecto fijo, y la otra modela la dependencia mediante un efecto aleatorio individuo.

Paralelamente, algunos autores han propuesto modelos que permiten incorporar variables asociadas a la prevalencia, sensibilidad y especificidad de los tests, tales como la historia clínica o características demográficas de los pacientes. Por ejemplo, Martinez et al. [2008] aplicaron esta propuesta a tres tests usados para el cribado de cáncer cervical en función de la edad y del estado de embarazo de la mujer. Martinez et al. [2005] han propuesto métodos para escenarios donde se mezclan pacientes evaluados por el *gold standard* y pacientes evaluados sin *gold standard*. En este caso se denomina análisis de datos semi-latentes.

Este capítulo introductorio cubre los conceptos básicos detrás de los tests diagnósticos sin *gold standard*, también contempla una breve introducción de la aproximación bayesiana aplicada en esta tesis.

1.2 Bases teóricas de los métodos Bayesianos

Durante las últimas décadas, el análisis estadístico bayesiano ha llegado a representar una alternativa importante a la escuela de pensamiento clásica o frecuentista. Con la metodología frecuentista, el parámetro desconocido es considerado fijo y esta basado en la muestra aleatoria representativa de la población de interés. Se espera que el intervalo de 95 % de confianza para el parámetro pueda contener en 95 % de las veces el verdadero parámetro poblacional, a través de repetidos experimentos con distintas muestras de la misma población. En contrapartida, en la perspectiva bayesiana [Box & Tiao, 2011, Ntzoufras, 2011] el parámetro desconocido es una variable aleatoria que si puede asignar una distribución de probabilidad. Esta distribución refleja el conocimiento que el investigador tiene del parámetro. Entonces, la inferencia bayesiana combina la información de los datos que es recorrido mediante una función de verosimilitud, con la distribución inicial del parámetro. Esto implica que el conocimiento del parámetro es modificado con los datos mediante aplicación del teorema de Bayes que permite la utilización de toda la información disponible acerca del parámetro.

1.2.1 Teorema de Bayes

El teorema de Bayes que aparece en los manuscritos del matemático británico Thomas Bayes¹ fueron publicados dos años después de su muerte (1763) por el filósofo moral Richard Price². En un fragmento del texto, Richard literalmente escribe: *“In an introduction which he has written to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that,*

¹en.wikipedia.org/wiki/Thomas_Bayes

²en.wikipedia.org/wiki/Richard_Price

under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.” [Bayes & Price, 1763].

En términos de probabilidad, considerando dos posibles eventos A y B , además asumiendo que $A = A_1 \cup \dots \cup A_n$ tal que $A_i \cap A_j = \emptyset$ para cada $i \neq j$. Luego, el *teorema de Bayes* proporciona una expresión para la probabilidad condicional de A_i dado B , que es igual a

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B | A_i) P(A_i)}{P(B)} = \frac{P(B | A_i) P(A_i)}{\sum_{i=1}^n P(B | A_i) P(A_i)},$$

que para cualquier resultado A y B , de forma simple y más general se puede escribir la ecuación conocida como *regla de Bayes*³

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B | A) P(A)}{P(B)}.$$

1.2.2 Modelo Bayesiano

La regla de Bayes ofrece un mecanismo probabilístico de aprendizaje a partir de los datos, que es la base de la inferencia bayesiana. El interés es inferir acerca del vector de parámetros desconocidos, θ , que caracteriza la distribución de los datos, Y , con una función de densidad o función de probabilidad $f(y | \theta)$. La **función de verosimilitud**, a partir de una muestra de datos $y = [y_1, \dots, y_n]^T$ independientes e idénticamente distribuida (i.i.d.) de tamaño n es la distribución conjunta:

$$\mathcal{L} = f(y | \theta) = \prod_{i=1}^n f(y_i | \theta).$$

En la teoría Bayesiana, a diferencia de la frecuentista clásica, θ es considerado como variable aleatoria con distribución $f(\theta)$. La función $f(\theta)$ es la conocida como función de **distribución a priori** o **prior** que expresa la información disponible sobre el parámetro θ antes de disponer de la muestra de datos. Consecuentemente, las conclusiones estadísticas acerca de θ son expresadas en términos de probabilidad de θ condicionada a los valores observados, $f(\theta | y)$, denominada **distribución a posteriori** o **posterior**. De acuerdo con el teorema de Bayes, la

³Algunos autores afirman que la *regla de Bayes* fue primeramente encontrada por el matemático francés Pierre-Simon de Laplace [Hoffman-Jørgensen, 1994a,b].

distribución posterior es igual a

$$f(\theta | y) = \frac{f(\theta, y)}{f(y)} = \frac{f(y | \theta) f(\theta)}{f(y)} \propto f(y | \theta) f(\theta).$$

El factor $f(y)$ se puede omitir porque no depende de θ , y se considera como una constante de normalización, es decir, la aplicación del Teorema de Bayes produce una distribución de probabilidad posterior que es proporcional al producto de la función de verosimilitud de los datos observados por la función prior del vector de parámetros desconocidos θ .

La especificación de la distribución prior es importante en inferencia bayesiana ya que influencia en la distribución posterior. El problema es la elección de la distribución prior porque en función de la prior se puede llevar a diferentes funciones posterior y posiblemente a diferentes conclusiones. El hecho de que el parámetro desconocido depende de la especificación de una distribución prior significa que se interpreta subjetivamente la probabilidad [Berger, 1985, Gelman et-al., 2014]. La media (mediana o moda) de la distribución prior informa acerca de las creencias prior de los parámetros y la varianza de esta distribución contempla la incertidumbre acerca de estas creencias. Por tanto en caso de tener una fuerte creencia prior, la varianza se definirá con un valor pequeño y si la creencia es débil, la varianza será grande. En ausencia de información disponible, o cuando hay información débil o poco creíble, normalmente se especifica una distribución prior que no influya en la distribución posterior, dejando los datos “*hablar por si solos*”. Tales distribuciones son conocidas como **prior no informativa** o **distribución prior difusa** o **vaga**.

La distribución posterior es el resultado final del análisis bayesiano y toda la información sobre los parámetros se obtiene de ella. En la mayoría de problemas hay más de un parámetro de interés, y entonces hay que diferenciar entre la distribución posterior conjunta y la distribución posterior marginal de cada parámetro. El primer objetivo es encontrar la distribución conjunta posterior de todos los parámetros, y luego obtener la distribución marginal de cada uno de los parámetros mediante la integración de la función conjunta respecto del resto de parámetros. Así, si consideramos que θ consiste en tres parámetros $\theta = (\theta_1, \theta_2, \theta_3)$, la distribución conjunta

posterior será

$$f(\theta_1, \theta_2, \theta_3 | y) \propto \mathcal{L}(y, \theta_1, \theta_2, \theta_3) f(\theta_1, \theta_2, \theta_3).$$

La distribución marginal posterior para θ_1 se obtiene a partir de la integración múltiple

$$f(\theta_1 | y) = \int_{\theta_2} \int_{\theta_3} f(\theta_1, \theta_2, \theta_3 | y) d\theta_2 d\theta_3,$$

el cálculo sigue análogo para los parámetros θ_2 y θ_3 . Por tanto, la estimación de los parámetros en el análisis bayesiano se centra en la distribución marginal posterior para cada parámetro de interés y la estimación puntual del parámetro desconocido θ_1 se obtiene a partir de la esperanza matemática de su distribución marginal posterior, $E(\theta_1 | y)$. También se puede obtener el intervalo de probabilidad del parámetro denominado intervalo de credibilidad. Este intervalo se calcula con la distribución marginal posterior del parámetro θ_1 definido por los cuantiles inferior, $q/2$, y superior, $1 - q/2$, con $(1 - q)100\%$ de credibilidad.

Generalmente, la distribución marginal posterior no es analíticamente tratable o es de difícil solución analítica por involucrar integrales de elevadas dimensiones [Ntzoufras, 2011]. Por eso se utilizan métodos de Monte Carlo vía cadenas de Markov (MCMC) [Gelfand & Smith, 1990] para simular la función de distribución posterior de los parámetros. La inferencia se hace condicionando a los datos observados y cuantificando la incertidumbre con probabilidades.

1.2.3 Algoritmos MCMC: Muestreo de Gibbs y Metropolis-Hastings

La metodología MCMC fue introducida por físicos en 1953 en una versión simplificada de Nicholas Metropolis⁴ [Metropolis et~al., 1953] que después fue generalizada por Hastings [1970] y finalmente utilizada por científicos Bayesianos [Gelfand et~al., 1990, Gelfand & Smith, 1990, Tanner & Wong, 1987]. El método MCMC más popular es el muestreo de Gibbs (Josiah Willard Gibbs⁵) introducido por Geman & Geman [1984] que es un caso especial del algoritmo Metropolis-Hastings [Chib & Greenberg, 1995]. Estos algoritmos son una clase de procedimientos basados en la integración de Monte Carlo, usando cadenas de Markov para generar

⁴https://en.wikipedia.org/wiki/Nicholas_Metropolis

⁵https://en.wikipedia.org/wiki/Josiah_Willard_Gibbs

muestras de la distribución de interés sin la necesidad de conocer la función de densidad posterior exacta. Las cadenas deben converger a una distribución posterior estacionaria (equilibrada), mediante un método iterativo en que en cada paso se genera un valor condicionado al valor del paso anterior. El proceso de funcionamiento del algoritmo usado para generar muestras y estimar el parámetro θ parte de un valor inicial $\theta^{(0)}$. [Gelman & Rubin \[1992\]](#) aconsejan generar diferentes muestras (cadenas) usando distintos valores iniciales y generar un gran número de iteraciones. Las primeras iteraciones se aconseja que se descarten para evitar la influencia a los valores iniciales fijados, conocido como fase de calentamiento (*burning phase*). Debido a que los valores de la muestra final generados no son independientes, es necesario controlar la autocorrelación de estos valores. La forma de controlar la autocorrelación es seleccionar valores a cada k iteraciones descartando los demás.

Algoritmo Muestreo de Gibbs

El muestreador de Gibbs es una de las técnicas para generar cadenas de Markov ampliamente utilizado para hacer inferencias sobre los parámetros de modelos bayesianos. El método consiste en el uso de distribuciones condicionales completas de los parámetros como la distribución de transición de la cadena de Markov. El muestreo de Gibbs es usado cuando la distribución condicional posterior presenta la forma de una densidad de probabilidad con un padrón conocido.

Supongamos que tenemos y datos observados y un vector con tres parámetros desconocidos a estimar $[\theta_1, \theta_2, \theta_3]$ con distribuciones condicionadas

$$f(\theta_1 | y, \theta_2, \theta_3), f(\theta_2 | y, \theta_1, \theta_3), f(\theta_3 | y, \theta_1, \theta_2),$$

y unos valores iniciales escogidos de forma arbitraria para cada parámetro $[\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}]$. El primer paso del muestreo de Gibbs es generar una nueva muestra $[\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}]$ a partir de las distribuciones condicionadas a los datos observados y a los valores iniciales. El procedimiento comienza con la generación de la observación aleatoria $\theta_1^{(1)}$ a partir de la distribución condicionada de θ_1 fijando todos los otros parámetros en el valor inicial $\theta_1^{(1)} \sim f(\theta_1 | y, \theta_2^{(0)}, \theta_3^{(0)})$,

luego se genera la observación aleatoria de $\theta_2^{(1)}$ a partir de $f(\theta_2 | y, \theta_1^{(0)}, \theta_3^{(0)})$ y del parámetro $\theta_3^{(1)}$ a partir de $f(\theta_3 | y, \theta_1^{(0)}, \theta_2^{(0)})$. Una vez se han actualizado todos los parámetros se obtiene una secuencia de Gibbs $[\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}]$ y el proceso vuelve a empezar, pero en lugar de condicionar los parámetros a los valores iniciales, se condicionan a los valores generados en la última iteración. El proceso se puede repetir hasta obtener k muestras de los parámetros desconocidos, por lo que se genera la cadena de Gibbs siguiente:

$$[\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}], [\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}], \dots, [\theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)}].$$

Algoritmo Metropolis-Hastings

El algoritmo de Metropolis-Hastings es un método de simulación estocástica que permite el muestreo de cualquier función de densidad $f(\theta)$, sin necesidad de conocer las distribuciones condicionales [Chib & Greenberg, 1995]. Por ejemplo, si la distribución condicional completa de interés no tiene la forma de una densidad, no es posible generar directamente los valores de esta distribución, el algoritmo Metropolis-Hastings convierte en una opción. El método consiste en la generación de valores candidatos de una densidad auxiliar que puede ser muestreada. Esta densidad auxiliar debe tener dos condiciones: 1º permitir la generación de muestras aleatorias con facilidad y rapidez y, 2º, deberá estar en el mismo dominio que la densidad objetivo. Estos valores candidatos que fueron simulados de esta densidad auxiliar se aceptan o no, dependiendo de cierta probabilidad. Según Chib & Greenberg [1995], hay varias familias de densidades auxiliares que se pueden elegir para generar valores candidatos [Hastings, 1970]

Supongamos que dada una distribución de probabilidad posterior $f(\theta | y)$ se desea generar una muestra de tamaño T para el parámetro de interés desconocido θ , entonces el algoritmo Metropolis-Hastings implementado bajo la perspectiva bayesiana se resume como sigue:

1. establecer el valor inicial $\theta^{(0)}$.
2. Para $t = 1, \dots, T$ repetir los siguientes pasos:
 - a. Establecer $\theta = \theta^{(t-1)}$.

b. Generar un nuevo valor del parámetro candidato θ' de una propuesta distribución condicional $q(\theta' | \theta)$.

c. Calcular

$$\alpha = \min \left(1, \frac{f(\theta' | y)q(\theta | \theta')}{f(\theta | y)q(\theta' | \theta)} \right).$$

d. Atualizar $\theta^{(t)} = \theta'$ con probabilidad α ; caso contrario establecer $\theta^{(t)} = \theta$.

El algoritmo convergerá a su distribución de equilibrio independientemente de cualquiera que sea la distribución $q(\cdot)$. Una característica importante del algoritmo es que no necesitamos evaluar la normalización constante $f(y)$ involucrada en $f(\theta | y)$ ya que se anula en α . De ahí que la probabilidad de aceptación se simplifica a

$$\alpha = \min \left(1, \frac{f(y | \theta')f(\theta')q(\theta | \theta')}{f(y | \theta)f(\theta)q(\theta' | \theta)} \right).$$

Estos métodos están implementados en el software WinBUGS [Lunn et al., 2000].

Una vez finalizado el procedimiento, es importante asegurarse de que la técnica ha funcionado correctamente. Para cada cadena de MCMC se debe verificar su convergencia y si la cantidad de valores obtenidos es suficiente para terminar de forma segura. En otras palabras, se debe asegurar de que las muestras obtenidas provienen de la verdadera distribución estacionaria.

1.2.4 El diagnóstico de convergencia

El algoritmo converge si la distribución posterior llega al equilibrio (estacionaria). Controlar la convergencia del algoritmo es esencial para obtener la distribución posterior correcta. Diferentes técnicas de diagnóstico son propuestos para controlar la convergencia, como el diagnóstico de Gelman-Rubin, los *dinamics trace plots*, el histórico de convergencia, el error de Monte Carlo, entre otros. A continuación se presentan algunos métodos de convergencia usando gráficos obtenidos con el software WinBUGS.

Gelman-Rubin

Gelman & Rubin [1992] proponen una aproximación que permite monitorizar la convergencia cuando más de una cadena son generadas en paralelo ($k > 1$). La convergencia se alcanza cuando las cadenas se han “olvidado” de sus valores iniciales y no se puede distinguir los resultados de todas las cadenas. El diagnóstico Gelman-Rubin [Gelman & Rubin, 1992] es un test estadístico de convergencia del tipo-ANOVA en el cual se calcula y compara la variabilidad entre cadenas (*between-sample*) y dentro de la cadena (*within-sample*). El estadístico R puede ser estimado por

$$\hat{R} = \frac{\hat{V}}{W} = \frac{T-1}{T} + \frac{B/T}{W} \frac{k+1}{k}$$

donde k es el número cadenas generadas. T es el número de iteraciones de cada cadena. B/T es la varianza de los valores medios posterior de todas las cadenas generadas (varianza entre cadenas), W es la media de las varianzas dentro de cada cadena, y la estimación

$$\hat{V} = \frac{T-1}{T}W + \frac{B}{T} \frac{k+1}{k}$$

es la varianza posterior combinada. Cuando la convergencia es alcanzada y con una cantidad elevada de datos generados, entonces $\hat{R} \rightarrow 1$ (a efectos prácticos, se podría suponer la convergencia si $\hat{R} < 1,05$, y para valores sustancialmente por encima de 1 indican falta de convergencia). Brooks & Gelman [1998] adoptaron una versión corregida de esta estadística dada por

$$\hat{R}_c = \frac{d+3}{d+1} \hat{R}$$

donde d son los grados de libertad estimados de la estimación de la varianza posterior combinada \hat{V} . Brooks & Gelman [1998] advierten de la importancia de asegurar no sólo que \hat{R} ha convergido a 1, sino también que B y W se han estabilizado, como ilustra la figura 1.2 que es la evolución de la varianza posterior combinada \hat{V} (en color verde), la media de la varianza W (en azul) y su razón \hat{R} (en rojo). En esta figura se observa que el ratio \hat{R} se aproxima al valor de referencia uno (línea negra discontinua) a partir de la iteración 200 y las varianzas se estabilizan también a partir de este valor, indicando convergencia del algoritmo. En WinBUGS

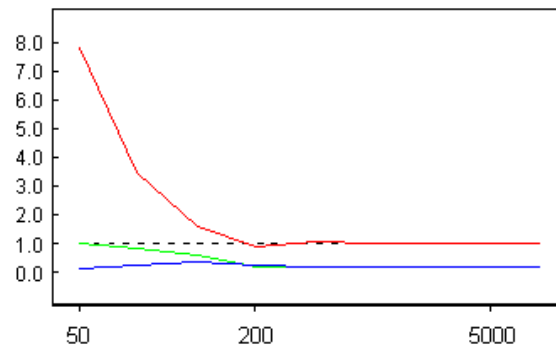


Figura 1.2: El diagnóstico de convergencia de Gelman-Rubin

para poder ser comparados en la misma figura, los valores \hat{V} y W son normalizados dividiendo por su máximo ($\max(\hat{V}, W)$).

Autocorrelación

En la figura 1.3 se presenta el gráfico de autocorrelación de las cadenas. Bajos valores de autocorrelación indican rápida convergencia y valores elevados indican convergencia débil. En este ejemplo la autocorrelación de las primeras iteraciones es alta, este es el proceso natural cuando se empieza a generar las cadenas, por eso las primeras iteraciones se descartan. Para solucionar el problema de una tasa débil de convergencia, se puede especificar un *lag* (o *thinning interval*) suficientemente grande (20, 50 . . .) dependiendo del valor de la autocorrelación y así obtener las observaciones finales independientes.

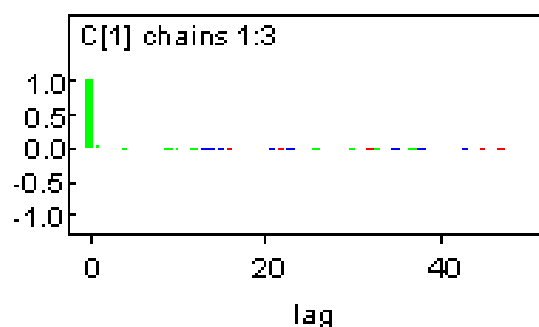


Figura 1.3: El diagnóstico de convergencia - autocorrelación

Trace-plots

En la figura 1.4 se muestran las iteraciones con los valores que se han generado. En el

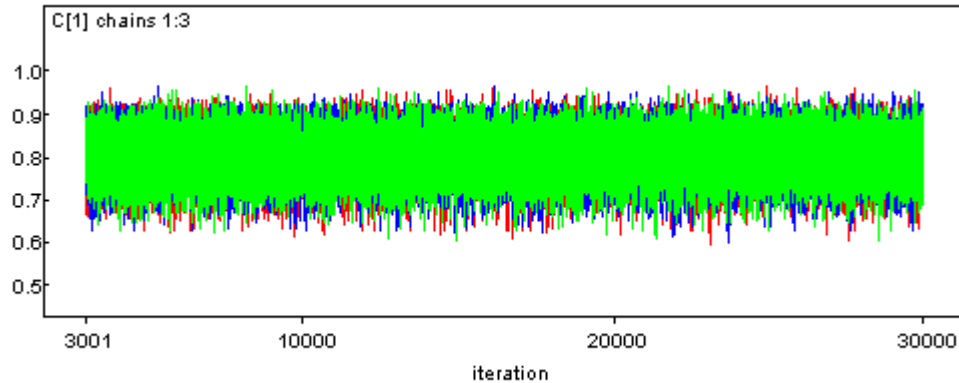


Figura 1.4: El histórico de las cadenas generadas con el Trace-plot

trace plot se puede mostrar las iteraciones finales descartando las primeras 3000 iteraciones y también se muestra las iteraciones una vez aplicado el *thinning interval* definido en el apartado de autocorrelación. La convergencia se obtiene cuando los valores están dentro de la zona sin grandes periodicidades y especialmente sin tendencias o paralelismos entre las cadenas. Es decir, cuando, por ejemplo el estadístico Gelman-Rubin indica convergencia, la distribución posterior es única para las cadenas generadas con los valores dispersos aleatoriamente por todo el espacio muestral y no se pueden distinguir las diferentes cadenas. El *Dynamic Trace-Plot* (figura 1.5) representa la visualización de las cadenas en el momento de la generación.

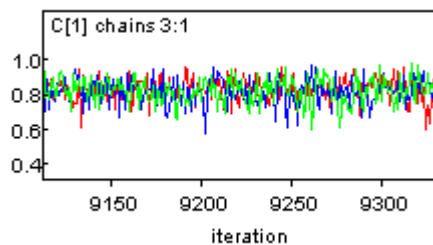


Figura 1.5: El diagnóstico de convergencia con el *Dynamic Trace-Plot*

En la figura 1.6 se muestra un caso de función de densidad de cadenas convergentes. En cambio, en la figura 1.7 se muestra el caso de no convergencia.

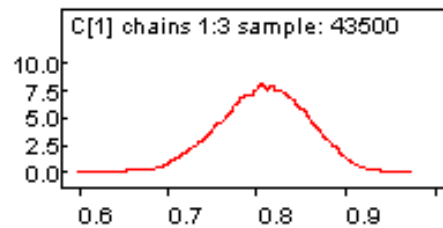


Figura 1.6: Densidad de los datos finales generados a partir de tres cadenas convergentes

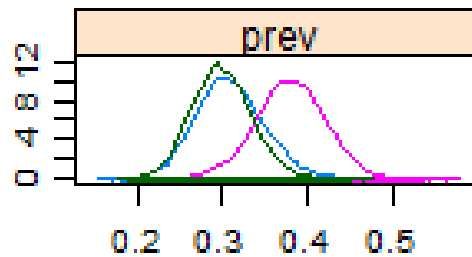


Figura 1.7: Densidad de los datos finales generados a partir de tres cadenas no convergentes

Error de Monte Carlo

El error de Monte Carlo (MCError) mide la variabilidad de la cadena debido a la simulación y es similar al error estándar de la media, pero ajustado para la muestra autocorrelacionada. El MCError debe ser bajo para aumentar la precisión de la estimación del parámetro de interés. Este es proporcional al inverso del tamaño de la muestra generada que puede ser controlado por el usuario. Por tanto, para un número suficiente de iteraciones, la cantidad de interés puede estimarse con mayor precisión.

También es utilizado para guiar la decisión en cuanto a la cantidad de iteraciones que necesita realizar debido a que se hace más pequeño a medida que se ejecutan más iteraciones. Una estrategia es definir un valor mínimo de corte y comparar el MCError con la desviación típica de las simulaciones, por ejemplo: 5 % del valor de la desviación típica del parámetro simulado debe ser inferior al MCError. Así se puede tener en cuenta el tiempo necesario de simulación para el parámetro convergir con la precisión deseada.

1.3 Estimación de la prevalencia y validez de tests diagnósticos en ausencia de *gold standard* mediante inferencia Bayesiana

En ocasiones hay interés en medir variables que no son observadas directamente, denominadas variables latentes. Los análisis realizados usando dichos datos fueron referidos como Análisis de Clases Latentes (ACL) en [Walter & Irwig \[1988\]](#) y [Kaldor & Clayton \[1985\]](#). El ACL [[McCutcheon & Hagenars, 2002](#)] tienen raíces en las ciencias sociales en que [Lazarsfeld \[1950\]](#) utilizó la técnica como una herramienta para la construcción de tipologías basadas en las variables dicotómicas observadas. Algunos de los primeros trabajos con el uso del ACL se llevaron a cabo en la Segunda Guerra Mundial, en donde a partir de los datos de una encuesta de perfiles se pretendía hacer una evaluación psicológica de los soldados del ejército estadounidense acerca de ítems tales como sus sentimientos y actitudes hacia el ejército, el enemigo y la guerra, además de sus condiciones de vida, entretenimiento y salud mental, entre otros.

El uso de ACL en el contexto de tests diagnósticos se realiza cuando los resultados del test de referencia no están libres de errores o cuando no existe un *gold standard* y entonces el estado de salud es un dato latente [[Joseph et al., 1995](#), [Tanner & Wong, 1987](#)]. Particularmente, las clases latentes usadas modelan la condición no observable como una variable latente categórica. En ausencia de *gold standard* se puede asumir que los resultados de dos tests aplicados al mismo individuo son independientes o dependientes, condicionados a conocer el verdadero estado de la enfermedad.

Según la teoría básica de probabilidad, cuando dos eventos son independientes la probabilidad de que ocurran de forma conjunta viene dada por el producto de sus probabilidades individuales. Dos acontecimientos A y B, dado un tercer evento C, son condicionalmente independientes si la ocurrencia o no ocurrencia de A junto con la de B se da de forma independiente dado C, y se representa matemáticamente por $P(A \cap B | C) = P(A | C)P(B | C)$. En nuestro caso, estaríamos interesados en los resultados de dos tests diagnósticos ($T_j, j = 1, 2$) condicionados al estado de la enfermedad. Suponemos que D representa el verdadero y latente estado

de salud y que tiene dos categorías, en donde $d = 1$ representa la enfermedad y $d = 0$ no enfermedad, es decir $P(T_1 = t_1 | D = d)$ y $P(T_2 = t_2 | D = d)$. Si los dos tests son independientes, eso quiere decir que

$$P(T_1 = t_1, T_2 = t_2 | D = d) \stackrel{ind}{=} P(T_1 = t_1 | D = d)P(T_2 = t_2 | D = d), \forall t_1, t_2 \quad (1.4)$$

La probabilidad de que J tests tengan resultados positivos considerando que los tests son independientes es igual a:

$$\begin{aligned} P(T_1 = 1, T_2 = 1, \dots, T_j = 1) &= \\ &= P(D = 1)P(T_1 = 1 | D = 1)P(T_2 = 1 | D = 1) \dots P(T_j = 1 | D = 1) \\ &+ P(D = 0)P(T_1 = 1 | D = 0)P(T_2 = 1 | D = 0) \dots P(T_j = 1 | D = 0) = \\ &= \pi S_1 S_2 \dots S_j + (1 - \pi)(1 - C_1)(1 - C_2) \dots (1 - C_j). \end{aligned} \quad (1.5)$$

Sin embargo, la suposición de independencia condicional no siempre es realista, ya que los resultados de los tests pueden estar correlacionados dentro de cada clase de la condición del enfermo. Cuando el supuesto de independencia condicional no es válido, podría dar lugar a estimaciones sesgadas de los parámetros de interés.

El primer trabajo realizado para modelar la probabilidad de detección de un test de diagnóstico en ausencia de *gold standard*, con el uso de metodología estadística bayesiana, se debe a [Joseph et~al. \[1995\]](#). En su metodología, el objetivo fue estimar la prevalencia de la infección por Strongyloides, las sensibilidades y especificidades de dos tests diagnósticos (examen de heces y serología) y consecuentemente sus respectivos valores predictivos positivos y negativos en ausencia de *gold standard*.

Para definir el modelo para la estimación de la prevalencia, sensibilidad y especificidad, se denota D_i como el verdadero y latente estado de salud de individuo i con dos categorías, $d = 1$ enfermo y $d = 0$ no enfermo. La variable T_{ji} denota el conjunto de J tests con resultado t_{ji} , $j = 1, \dots, J$, para cada individuo i . Si se asume que T_{ji} tiene 2 posibles resultados, entonces $t_{ji} = 1$ cuando el resultado del test j es positivo y $t_{ji} = 0$ cuando el resultado es negativo. Sea \mathbf{T}_i el vector con el conjunto de los resultados del individuo i para los J tests $(T_{1i}, T_{2i}, \dots, T_{Ji})^T$.

La contribución del i th individuo para la verosimilitud es dada por la expresión:

$$P(\mathbf{T}_i = \mathbf{t}_i) = \sum_{d=0}^1 P(D_i = d) P(\mathbf{T}_i = \mathbf{t}_i | D_i = d) \quad (1.6)$$

La ecuación (1.6) es flexible y general y puede ser usada para estimar los parámetros de uno o varios tests de diagnóstico, y considerar correlación entre los tests.

Asumiendo que J tests sean mutuamente independientes, condicionado al estado real, la expresión para los parámetros de validez de los tests es:

$$P(\mathbf{T}_i = \mathbf{t}_i | D_i = d) = \prod_{j=1}^J P(T_{ji} = t_{ji} | D_i = d) \quad (1.7)$$

Combinando las ecuaciones (1.6) con (1.7) se obtiene la expresión:

$$P(\mathbf{T}_i = \mathbf{t}_i) = \sum_{d=0}^1 P(D_i = d) \prod_{j=1}^J P(T_{ji} = t_{ji} | D_i = d) \quad (1.8)$$

La ecuación (1.8) se puede parametrizar en términos de sensibilidad $S_j = P(T_j = 1 | D = 1)$, especificidad $C_j = P(T_j = 0 | D = 0)$ y prevalencia $\pi = P(D = 1)$. La función de verosimilitud (\mathcal{L}) es igual a

$$\mathcal{L} = \prod_{i=1}^N P(\mathbf{T}_i = \mathbf{t}_i) \quad (1.9)$$

1.3.1 Caso de un solo test diagnóstico

1.3.1.1 Verosimilitud para el caso de un solo test diagnóstico

A partir de la ecuación (1.8), la probabilidad de obtener un resultado positivo con un solo test diagnóstico es

$$\begin{aligned}
 P(T_i = 1) &= \sum_{d=1}^0 P(D_i = d_i)P(T_i = 1|D_i = d_i) \\
 &= P(D_i = 1)P(T_i = 1|D_i = 1) + P(D_i = 0)P(T_i = 1|D_i = 0) \quad (1.10) \\
 &= P(D_i = 1)P(T_i = 1|D_i = 1) + (1 - P(D_i = 1))(1 - P(T_i = 0|D_i = 0)) \\
 &= \pi S + (1 - \pi)(1 - C)
 \end{aligned}$$

De similar manera, la probabilidad de un test negativo ($P(T_i = 0) = 1 - P(T_i = 1)$) es dada por

$$\begin{aligned}
 P(T_i = 0) &= \sum_{d=1}^0 P(D_i = d_i)P(T_i = 0|D_i = d_i) \\
 &= P(D_i = 1)P(T_i = 0|D_i = 1) + P(D_i = 0)P(T_i = 0|D_i = 0) \quad (1.11) \\
 &= P(D_i = 1)(1 - P(T_i = 1|D_i = 1)) + P(D_i = 0)P(T_i = 0|D_i = 0) \\
 &= \pi(1 - S) + (1 - \pi)C
 \end{aligned}$$

Asumiendo distribución Bernoulli para la variable resultado del test T y con probabilidad del suceso dada por (1.10), la función de verosimilitud es:

$$\mathcal{L}(\theta) = \prod_{i=1}^N [\pi S + (1 - \pi)(1 - C)]^{t_i} [\pi(1 - S) + (1 - \pi)C]^{1-t_i} \quad (1.12)$$

donde θ es el vector de parámetros (π, S, C) .

1.3.1.2 Definición de las distribuciones prior para un solo test diagnóstico

Se asume que la prevalencia y los parámetros de validez del test tengan distribución $Beta(\alpha, \beta)$ por que es conveniente debido a la flexibilidad de la distribución que es definida

entre 0 y 1. Los parámetros de la distribución prior Beta son denominados hiperparámetros α y β para distinguirlos de los parámetros del modelo. Estos hiperparámetros aportan las creencias que deben ser fijadas con base en anteriores informaciones a los datos observados. Esta información debe ser cuidadosamente estudiada, en base a estudios científicos previos o a opiniones subjetivas de los expertos. Por tanto, para los parámetros θ , se asume la distribución prior:

$$\begin{aligned}\pi &\sim \text{Beta}(\alpha_\pi, \beta_\pi) \\ S &\sim \text{Beta}(\alpha_s, \beta_s) \\ C &\sim \text{Beta}(\alpha_c, \beta_c)\end{aligned}\tag{1.13}$$

Las creencias previas de los parámetros pueden ser resumidas con el intervalo de creencias delimitado por l (límite inferior) y u (límite superior). Por tanto, α and β se calculan fijando la media de la distribución Beta $\left(\frac{\alpha}{\alpha+\beta}\right)$ al punto medio del intervalo de creencias y fijando la desviación de la distribución Beta $\left(\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}\right)$ a una cuarta parte de la amplitud del intervalo de creencias. Por tanto, α y β , en función de las creencias serán igual a:

$$\begin{aligned}\alpha &= -\frac{1}{2} \frac{(5l^3 + 11l^2u - 8l^2 + 11lu^2 - 16lu + 5u^3 - 8u^2)}{(l-u)^2} \\ \beta &= \alpha \frac{(2-l-u)}{l+u}\end{aligned}\tag{1.14}$$

1.3.1.3 Distribución posterior para un test diagnóstico

Asumiendo que D_{ik} (estado real de la enfermedad) dado el resultado del test $T_i = |k - 2|$ para $k = 1, 2$ es una distribución Bernoulli con probabilidad g_{ki} igual a

$$g_{1i} = P(D_i|T_i = 1) = \frac{\pi S}{\pi S + (1 - \pi)(1 - C)}\tag{1.15}$$

para $k=1$, y

$$g_{2i} = P(D_i|T_i = 0) = \frac{\pi(1 - S)}{\pi(1 - S) + (1 - \pi)C}\tag{1.16}$$

para $k=2$, entonces la distribución conjunta de la variable latente D_i , $i = 1, \dots, n$, cuando el test tiene resultado positivo es:

$$h_1 = f(d_1, \dots, d_n) \propto \prod_{i=1}^n g_{1i}^{t_i d_i} (1 - g_{1i})^{t_i (1-d_i)} \quad (1.17)$$

y la función de distribución conjunta de la variable latente cuando el test tiene resultado negativo es:

$$h_2 = f(d_1, \dots, d_n) \propto \prod_{i=1}^n g_{2i}^{(1-t_i)d_i} (1 - g_{2i})^{(1-t_i)(1-d_i)} \quad (1.18)$$

La función de verosimilitud de los datos observados, \mathbf{T} , y los latentes, \mathbf{D} , se calcula combinando la (1.12) con las ecuaciones (1.17) y (1.18). Para facilitar los cálculos se utilizan las siguientes reparametrizaciones: $u = \pi S$, $v = (1 - \pi)(1 - C)$, $w = \pi(1 - S)$ y $x = (1 - \pi)C$

$$\begin{aligned} L &= \prod_{i=1}^n (u + v)^{t_i} (w + x)^{(1-t_i)} \\ h_1 &= \prod_{i=1}^n \frac{u^{t_i d_i} v^{t_i (1-d_i)}}{(u + v)^{t_i}} \\ h_2 &= \prod_{i=1}^n \frac{w^{(1-t_i)d_i} x^{(1-t_i)(1-d_i)}}{(w + x)^{(1-t_i)}} \\ L \times h_1 &= \prod_{i=1}^n (w + x)^{(1-t_i)} u^{t_i d_i} v^{t_i (1-d_i)} \\ L \times h_1 \times h_2 &= \prod_{i=1}^n u^{t_i d_i} v^{t_i (1-d_i)} w^{(1-t_i)d_i} x^{(1-t_i)(1-d_i)} \end{aligned}$$

por tanto, la función de verosimilitud es:

$$\mathcal{L}(\mathbf{T}, \mathbf{D} \mid \theta) = [\pi S]^{\sum_{i=1}^n t_i d_i} [(1 - \pi)(1 - C)]^{\sum_{i=1}^n t_i (1-d_i)} [\pi(1 - S)]^{\sum_{i=1}^n (1-t_i)d_i} [(1 - \pi)C]^{\sum_{i=1}^n (1-t_i)(1-d_i)}, \quad (1.19)$$

donde: $\sum_{i=1}^n d_i$ es la cantidad total de individuos enfermos; $\sum_{i=1}^n (1 - d_i)$ es la cantidad total de individuos no enfermos; $\sum_{i=1}^n t_i$ es el total de individuos con resultado positivo al test; $\sum_{i=1}^n (1 - t_i)$ es el total de individuos con resultado negativo al test. La tabla 1.2 ilustra la relación entre los

datos latentes, $\sum_{i=1}^n d_i$ y $\sum_{i=1}^n (1 - d_i)$, y los datos observados, $\sum_{i=1}^n t_i$ y $\sum_{i=1}^n (1 - t_i)$, de la siguiente forma:

- la cantidad de enfermos con el resultado positivo en el test es $\sum_{i=1}^n t_i d_i$
- la cantidad de enfermos con el resultado negativo en el test es $\sum_{i=1}^n (1 - t_i) d_i$
- la cantidad de no enfermos con el resultado positivo en el test es $\sum_{i=1}^n t_i (1 - d_i)$
- la cantidad de no enfermos con el resultado negativo en el test es $\sum_{i=1}^n (1 - t_i) (1 - d_i)$

		“Gold Standard” (latente)		
		+	-	Total
Test (observado)	+	$\sum_{i=1}^n t_i d_i$	$\sum_{i=1}^n t_i (1 - d_i)$	$\sum_{i=1}^n t_i$
	-	$\sum_{i=1}^n (1 - t_i) d_i$	$\sum_{i=1}^n (1 - t_i) (1 - d_i)$	$\sum_{i=1}^n (1 - t_i)$
Total		$\sum_{i=1}^n d_i$	$\sum_{i=1}^n (1 - d_i)$	$\sum_{i=1}^n 1$

Tabla 1.2: Datos latentes y observados para el caso de un test de diagnóstico en ausencia de gold standard

Aplicando el Teorema de Bayes, la distribución posterior de cada parámetro es una distribución Beta ya que es proporcional al producto de una verosimilitud de distribución Bernoulli por una distribución prior Beta. La función de verosimilitud (1.19) en función de la variable latente y de los datos observados, multiplicada por la distribuciones prior en (1.13) es la distribución conjunta posterior de los parámetros:

$$\begin{aligned}
 \mathcal{P}(\theta, \mathbf{D}|\mathbf{T}) &\propto \pi^{\alpha_\pi + \sum_{i=1}^n d_i - 1} (1 - \pi)^{\beta_\pi + \sum_{i=1}^n (1 - d_i) - 1} \\
 &\times S^{\alpha_s + \sum_{i=1}^n t_i d_i - 1} (1 - S)^{\beta_s + \sum_{i=1}^n (1 - t_i) d_i - 1} \\
 &\times C^{\alpha_c + \sum_{i=1}^n (1 - t_i)(1 - d_i) - 1} (1 - C)^{\beta_c + \sum_{i=1}^n t_i(1 - d_i) - 1}
 \end{aligned} \tag{1.20}$$

La complejidad del modelo (1.20) no permite deducir de forma trivial la distribución marginal posterior analíticamente. Para solucionar este problema, se puede utilizar el muestreo de Gibbs [Geman & Geman, 1984]. Para cada componente de θ es necesario especificar la

distribución condicional posterior completa, es decir, la distribución de cada parámetro dados los datos observables, los datos latentes y otros parámetros, como sigue:

$$\begin{aligned}
\pi|\mathbf{D}, \alpha_\pi, \beta_\pi &\sim \text{Beta} \left(\alpha_\pi + \sum_{i=1}^n d_i, \beta_\pi + \sum_{i=1}^n (1 - d_i) \right) \\
S|\mathbf{T}, \mathbf{D}, \alpha_s, \beta_s &\sim \text{Beta} \left(\alpha_s + \sum_{i=1}^n t_i d_i, \beta_s + \sum_{i=1}^n (1 - t_i) d_i \right) \\
C|\mathbf{T}, \mathbf{D}, \alpha_c, \beta_c &\sim \text{Beta} \left(\alpha_c + \sum_{i=1}^n (1 - t_i)(1 - d_i), \beta_c + \sum_{i=1}^n t_i(1 - d_i) \right) \\
d_i|\theta, \mathbf{T} &\sim \text{Bernoulli} \left(\frac{\pi S^{t_i} (1 - S)^{1-t_i}}{\pi S^{t_i} (1 - S)^{1-t_i} + (1 - \pi) C^{1-t_i} (1 - C)^{t_i}} \right)
\end{aligned} \tag{1.21}$$

1.3.2 Caso de dos tests diagnósticos independientes

1.3.2.1 Verosimilitud para el caso de dos tests diagnósticos

Para dos tests de diagnóstico independientes, con la verosimilitud definida en (1.9) el procedimiento es semejante al caso de un único test. Considerando que $J = 2$, \mathbf{T}_i el vector con el conjunto de los resultados del individuo i para los 2 tests $(T_{1i}, T_{2i})^T$. La contribución del i th individuo para la verosimilitud es dada por las siguientes expresiones:

$$\begin{aligned}
 P(T_{1i} = 1, T_{2i} = 1) &= \sum_{d=1}^0 P(D_i = d_i)P(T_{1i} = 1, T_{2i} = 1|D_i = d_i) \stackrel{ind}{=} \\
 &= P(D_i = 1)P(T_{1i} = 1|D_i = 1)P(T_{2i} = 1|D_i = 1) \\
 &\quad + P(D_i = 0)P(T_{1i} = 1|D_i = 0)P(T_{2i} = 1|D_i = 0) = \\
 &= \pi S_1 S_2 + (1 - \pi)(1 - C_1)(1 - C_2) \\
 P(T_{1i} = 1, T_{2i} = 0) &= \sum_{d=1}^0 P(D_i = d_i)P(T_{1i} = 1, T_{2i} = 0|D_i = d_i) \stackrel{ind}{=} \\
 &= P(D_i = 1)P(T_{1i} = 1|D_i = 1)P(T_{2i} = 0|D_i = 1) \\
 &\quad + P(D_i = 0)P(T_{1i} = 1|D_i = 0)P(T_{2i} = 0|D_i = 0) = \\
 &= \pi S_1(1 - S_2) + (1 - \pi)(1 - C_1)C_2 \\
 P(T_{1i} = 0, T_{2i} = 1) &= \sum_{d=1}^0 P(D_i = d_i)P(T_{1i} = 0, T_{2i} = 1|D_i = d_i) \stackrel{ind}{=} \\
 &= P(D_i = 1)P(T_{1i} = 0|D_i = 1)P(T_{2i} = 1|D_i = 1) \\
 &\quad + P(D_i = 0)P(T_{1i} = 0|D_i = 0)P(T_{2i} = 1|D_i = 0) = \\
 &= \pi(1 - S_1)S_2 + (1 - \pi)C_1(1 - C_2) \\
 P(T_{1i} = 0, T_{2i} = 0) &= \sum_{d=1}^0 P(D_i = d_i)P(T_{1i} = 0, T_{2i} = 0|D_i = d_i) \stackrel{ind}{=} \\
 &= P(D_i = 1)P(T_{1i} = 0|D_i = 1)P(T_{2i} = 0|D_i = 1) \\
 &\quad + P(D_i = 0)P(T_{1i} = 0|D_i = 0)P(T_{2i} = 0|D_i = 0) = \\
 &= \pi(1 - S_1)(1 - S_2) + (1 - \pi)C_1C_2
 \end{aligned} \tag{1.22}$$

La probabilidad de que el individuo esté enfermo dado los resultados de dos tests independientes es la probabilidad de suceso de la variable latente, con distribución Bernoulli D_i , condicio-

nada a la combinación de los resultados de los tests

$$g_{1i} = P(D_i | T_{1i} = 1, T_{2i} = 1) = \frac{\pi S_1 S_2}{\pi S_1 S_2 + (1 - \pi)(1 - C_1)(1 - C_2)}$$

$$g_{2i} = P(D_i | T_{1i} = 1, T_{2i} = 0) = \frac{\pi S_1(1 - S_2)}{\pi S_1(1 - S_2) + (1 - \pi)(1 - C_1)C_2}$$

$$g_{3i} = P(D_i | T_{1i} = 0, T_{2i} = 1) = \frac{\pi(1 - S_1)S_2}{\pi(1 - S_1)S_2 + (1 - \pi)C_1(1 - C_2)}$$

$$g_{4i} = P(D_i | T_{1i} = 0, T_{2i} = 0) = \frac{\pi(1 - S_1)(1 - S_2)}{\pi(1 - S_1)(1 - S_2) + (1 - \pi)C_1C_2},$$

de forma simplificada, la variable latente

$$D_i \sim \text{Bernoulli} \left(\frac{\pi \prod_{j=1}^2 S_j^{t_{ji}} (1 - S_j)^{1-t_{ji}}}{\pi \prod_{j=1}^2 S_j^{t_{ji}} (1 - S_j)^{1-t_{ji}} + (1 - \pi) \prod_{j=1}^2 C_j^{1-t_{ji}} (1 - C_j)^{t_{ji}}} \right). \quad (1.23)$$

La contribución del i th individuo a la función de verosimilitud es igual a:

$$\pi^{d_i} (1 - \pi)^{(1-d_i)} \prod_{j=1}^2 S_j^{t_{ji} d_i} (1 - S_j)^{(1-t_{ji}) d_i} C_j^{(1-t_{ji})(1-d_i)} (1 - C_j)^{t_{ji}(1-d_i)} \quad (1.24)$$

y la función de verosimilitud será igual a:

$$\begin{aligned} \mathcal{L}(\theta) &= \pi^{\sum_{i=1}^n d_i} (1 - \pi)^{\sum_{i=1}^n (1-d_i)} \\ &\times \prod_{j=1}^2 S_j^{\sum_{i=1}^n t_{ji} d_i} (1 - S_j)^{\sum_{i=1}^n (1-t_{ji}) d_i} C_j^{\sum_{i=1}^n (1-t_{ji})(1-d_i)} (1 - C_j)^{\sum_{i=1}^n t_{ji}(1-d_i)} \end{aligned} \quad (1.25)$$

1.3.2.2 Definición de las distribuciones prior para dos tests diagnósticos

En este caso se considera al igual que el caso de un test, que la distribución prior de los parámetros sea un distribución Beta,

$$\begin{aligned}
 \pi &\sim \text{Beta}(\alpha_\pi, \beta_\pi) \\
 S_1 &\sim \text{Beta}(\alpha_{s_1}, \beta_{s_1}) \\
 S_2 &\sim \text{Beta}(\alpha_{s_2}, \beta_{s_2}) \\
 C_1 &\sim \text{Beta}(\alpha_{c_1}, \beta_{c_1}) \\
 C_2 &\sim \text{Beta}(\alpha_{c_2}, \beta_{c_2})
 \end{aligned} \tag{1.26}$$

1.3.2.3 Distribución posterior para dos tests diagnósticos

La distribución posterior conjunta es proporcional a $\mathcal{L}(\theta) \times \text{prior}$

$$\begin{aligned}
 \mathcal{P}(\theta, \mathbf{D}|\mathbf{T}) &\propto \mathcal{L}(\theta) \times \pi^{\alpha_\pi-1} (1-\pi)^{\beta_\pi-1} \prod_{j=1}^2 S_j^{\alpha_{S_j}-1} (1-S_j)^{\beta_{S_j}-1} C_j^{\alpha_{C_j}-1} (1-C_j)^{\beta_{C_j}-1} \\
 &= \pi^{\sum_{i=1}^n d_i + \alpha_\pi - 1} (1-\pi)^{\sum_{i=1}^n (1-d_i) + \beta_\pi - 1} \\
 &\quad \times \prod_{j=1}^2 S_j^{\sum_{i=1}^n t_{ji} d_i + \alpha_{S_j} - 1} (1-S_j)^{\sum_{i=1}^n (1-t_{ji}) d_i + \beta_{S_j} - 1} \\
 &\quad \times C_j^{\sum_{i=1}^n (1-t_{ji})(1-d_i) + \alpha_{C_j} - 1} (1-C_j)^{\sum_{i=1}^n t_{ji}(1-d_i) + \beta_{C_j} - 1}
 \end{aligned} \tag{1.27}$$

Las distribuciones condicionales posterior para las componentes de θ necesarias para el

algoritmo muestreo de Gibbs son dadas por

$$\begin{aligned}
\pi|\mathbf{D}, \alpha_\pi, \beta_\pi &\sim \text{Beta} \left(\alpha_\pi + \sum_{i=1}^n d_i, \beta_\pi + \sum_{i=1}^n (1 - d_i) \right) \\
S_j|\mathbf{T}, \mathbf{D}, \alpha_{s_j}, \beta_{s_j} &\sim \text{Beta} \left(\alpha_{s_j} + \sum_{i=1}^n t_{ji}d_i, \beta_{s_j} + \sum_{i=1}^n (1 - t_{ji})d_i \right) \\
C_j|\mathbf{T}, \mathbf{D}, \alpha_{c_j}, \beta_{c_j} &\sim \text{Beta} \left(\alpha_{c_j} + \sum_{i=1}^n (1 - t_{ji})(1 - d_i), \beta_{c_j} + \sum_{i=1}^n t_{ji}(1 - d_i) \right) \\
d_i|\theta, \mathbf{T} &\sim \text{Bernoulli} \left(\frac{\pi \prod_{j=1}^2 S_j^{t_{ji}} (1 - S_j)^{1-t_{ji}}}{\pi \prod_{j=1}^2 S_j^{t_{ji}} (1 - S_j)^{1-t_{ji}} + (1 - \pi) \prod_{j=1}^2 C_j^{1-t_{ji}} (1 - C_j)^{t_{ji}}} \right).
\end{aligned} \tag{1.28}$$

La independencia condicional es una hipótesis bastante fuerte y varios autores han señalado que no es una hipótesis realista en muchas situaciones prácticas. Si bien la hipótesis de independencia condicional generalmente simplifica el problema estadístico, recientemente se han desarrollado más métodos que relajen tal hipótesis. En la literatura hay recomendaciones en favor del uso de modelos bajo la hipótesis de dependencia porque presenta resultados más fiables cuando la independencia no es realista [van Smeden et al., 2014]. En la próxima sección son presentados modelos Bayesianos bajo la hipótesis de dependencia entre tests diagnósticos.

1.3.3 Caso de dos tests diagnósticos dependientes

Los modelos de dependencia que directamente captan la variabilidad entre tests son denominados de efectos fijos y los modelos que indirectamente modelan la dependencia entre los tests con la componente de variabilidad inter-sujetos son denominados de efecto aleatorio individuo. Los primeros modelos propuestos para modelar la dependencia condicional entre tests, se destacan los de clases latentes [Yang & Becker, 1997] y los con efectos aleatorios [Hadgu & Qu, 1998, Qu et al., 1996].

En el ámbito de medicina veterinaria Engel et al. [2006] desarrolló metodología bayesiana para ajustar los datos a los modelos con dependencia entre tres tests, aplicados a cuatro poblaciones de rebaños de cerdas para detectar EF-positiva *Streptococcus suis* serotipo 2. Para dos o más tests aplicados para detectar la misma enfermedad (o un fenómeno biológico espe-

cífico) es intuitivo esperar que dichos tests sean correlacionados, cuando ciertas características comunes entre individuos pueden aumentar o disminuir la probabilidad de detección de los tests. [Dendukuri & Joseph \[2001\]](#) proponen dos modelos con enfoque bayesiano. El primer modelo, conocido como de efecto fijo, estima la covarianza entre los tests para el grupo de enfermos y la covarianza entre los tests para el grupo de los no enfermos. El segundo, conocido como modelo de efecto aleatorio, tienen en cuenta la correlación entre las medidas del individuo realizados con los distintos test mediante un efecto aleatorio [[Dendukuri & Joseph, 2001](#), [Hadgu & Qu, 1998](#), [Qu et~al., 1996](#)].

1.3.3.1 Modelo dependiente con efectos fijos

Diversos autores han demostrado que es importante tener en cuenta la dependencia entre tests con el fin de obtener estimaciones insesgadas [[Fryback, 1978](#), [Vacek, 1985](#)]. Se debe de tener en cuenta que dos tests de diagnóstico pueden ser condicionalmente dependientes debido a causas que no sean únicamente el estado real de salud del paciente. En este caso, la probabilidad de detección de un test puede variar en los individuos a causa de que la gravedad de la enfermedad puede hacer que un test sea más sensible en la detección que otro test. El modelo de efectos fijos supone que la dependencia condicional entre tests puede ser estimada a través de parámetros que modelan la covarianza entre los tests en cada clase latente. Para la clase de los individuos enfermos, la componente que representa la covarianza entre dos tests se denomina cov_S y para la clase de los individuos no enfermos, cov_C . En en contexto estadístico, sean X y Y dos variables aleatorias, $E(\cdot)$ es la esperanza matemática y $Cov(\cdot)$ es la covarianza, por definición $Cov(X, Y) = E(X, Y) - E(X)E(Y)$, lo que implica

$$\begin{aligned} cov_S &= E(T_1, T_2|D = 1) - E(T_1|D = 1)E(T_2|D = 1), \text{ y} \\ cov_C &= E(T_1, T_2|D = 0) - E(T_1|D = 0)E(T_2|D = 0) \end{aligned} \tag{1.29}$$

Las probabilidades de la combinación de los resultados con los dos tests se expresan:

$$\begin{aligned}
P(T_1 = 1, T_2 = 1|D = 1) &= S_1 S_2 + cov_S \\
P(T_1 = 1, T_2 = 0|D = 1) &= S_1(1 - S_2) - cov_S \\
P(T_1 = 0, T_2 = 1|D = 1) &= (1 - S_1)S_2 - cov_S \\
P(T_1 = 0, T_2 = 0|D = 1) &= (1 - S_1)(1 - S_2) + cov_S \\
P(T_1 = 1, T_2 = 1|D = 0) &= (1 - C_1)(1 - C_2) + cov_C \\
P(T_1 = 1, T_2 = 0|D = 0) &= (1 - C_1)C_2 - cov_C \\
P(T_1 = 0, T_2 = 1|D = 0) &= C_1(1 - C_2) - cov_C \\
P(T_1 = 0, T_2 = 0|D = 0) &= C_1 C_2 + cov_C
\end{aligned} \tag{1.30}$$

De las ecuaciones (1.30) notase que la probabilidad de cada combinación de resultados entre dos tests, dado la variable latente D , es añadida las cantidades cov_S o cov_C .

1.3.3.2 Verosimilitud para el modelo con efectos fijos

La función de verosimilitud de los datos observados dado los datos latentes para los parámetros $\psi = (\pi, S_1, S_2, C_1, C_2, cov_S, cov_C)$ del modelo dependiente de efectos fijos es

$$\begin{aligned}
\mathcal{L}(\psi) &= \left(\pi^{\sum_{i=1}^n d_i} (1 - \pi)^{\sum_{i=1}^n (1-d_i)} \right) \\
&\times \left[\left(\prod_{j=1}^2 S_j^{\sum_{i=1}^n t_{ji} d_i} (1 - S_j)^{\sum_{i=1}^n (1-t_{ji}) d_i} \right) + cov_S (-1)^{2-(t_{1i}+t_{2i})} \right] \\
&\times \left[\left(\prod_{j=1}^2 C_j^{\sum_{i=1}^n (1-t_{ji})(1-d_i)} (1 - C_j)^{\sum_{i=1}^n t_{ji}(1-d_i)} \right) + cov_C (-1)^{2-(t_{1i}+t_{2i})} \right]
\end{aligned} \tag{1.31}$$

1.3.3.3 Prior para los parámetros del modelo con efectos fijos

Para la elicitación de las distribuciones prior del modelo con efectos fijos se asume la distribución $Beta(\alpha, \beta)$ para la prevalencia, las sensibilidades y las especificidades (ecuaciones (1.26)), ya que el valor de cada parámetros pertenece al intervalo $(0, \dots, 1)$. Los límites l

y u son usados para fijar los valores de los hiperparámetros α y β que son calculados fijando la media de la distribución Beta con el punto medio del intervalo de las opiniones de expertos y también fijando la desviación estándar de la distribución Beta a $1/4$ del rango del mismo intervalo de creencias. Las expresiones dependientes de los límites l y u fueron deducidas en las ecuaciones (1.14). Es natural también que no esté disponible información previa acerca de los parámetros de covarianza entre los tests, pero se puede asumir una distribución prior beta generalizada para cov_S y cov_C considerando el rango ilustrado en la figura 1.8. Los posibles valores para las covarianzas, representados en el gráfico de contornos, son calculados de acuerdo con la combinación de las sensibilidades (o las especificidades) de dos tests.

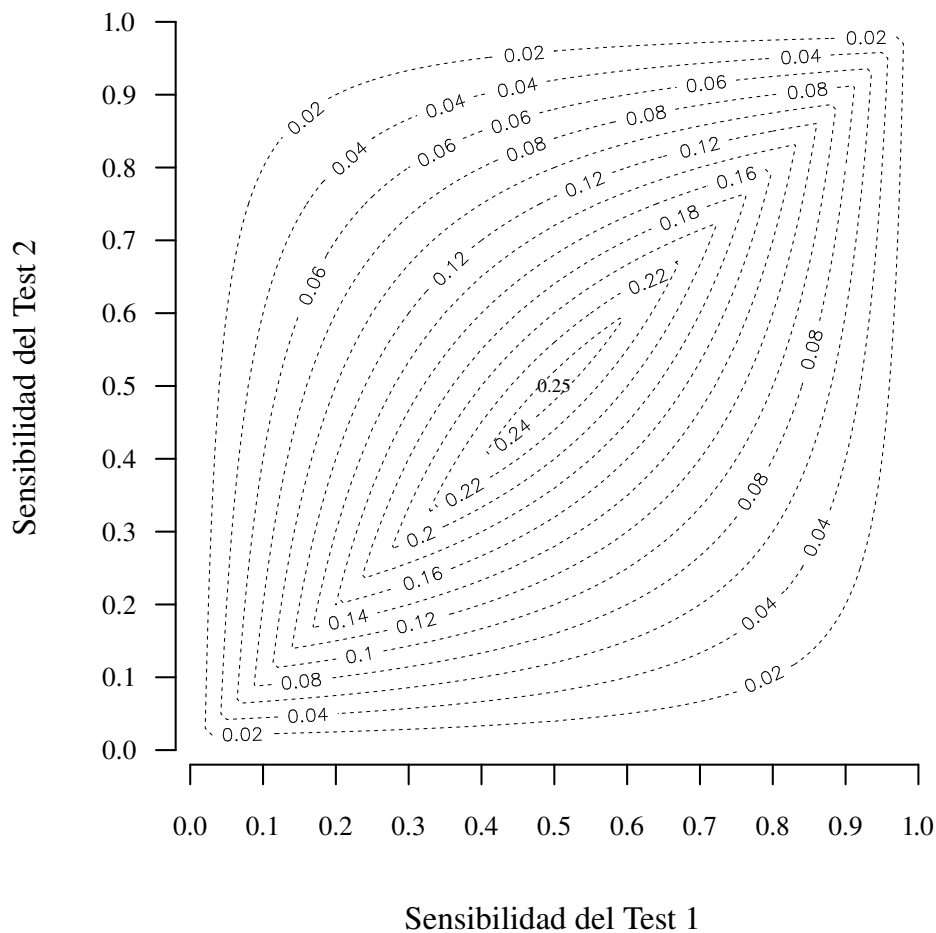


Figura 1.8: Covarianzas para la combinación de dos tests condicionalmente dependientes.

Primeramente, los rangos plausibles para las covarianzas son determinado por:

las sensibilidades entre los individuos enfermos

$$(S_1 - 1)(1 - S_2) \leq cov_S \leq \min(S_1, S_2) - S_1 S_2 \quad (1.32)$$

y las especificidades entre los individuos no enfermos

$$(C_1 - 1)(1 - C_2) \leq cov_C \leq \min(C_1, C_2) - C_1 C_2$$

Los límites inferiores en (1.32) son siempre negativos, pero normalmente las covarianzas son asumidas positivas, dado que en la practica es muy poco probable que dos tests de diagnóstico estén correlacionados negativamente [Vacek, 1985]. De esta forma, las ecuaciones para cov_S y cov_C están limitadas por

$$\begin{aligned} 0 \leq cov_S &\leq \min(S_1, S_2) - S_1 S_2 \\ 0 \leq cov_C &\leq \min(C_1, C_2) - C_1 C_2 \end{aligned} \quad (1.33)$$

Para la combinación de dos tests dependientes, la covarianza mínima es cero y la máxima es de 0,25. La máxima covarianza ocurre cuando las dos sensibilidades o las dos especificidades son iguales a 0,5, debido a que el límite superior en las sensibilidades es $l_s = \min(S_1, S_2) - S_1 S_2$ y el límite superior en las especificidades $l_c = \min(C_1, C_2) - C_1 C_2$ (figura 1.8).

Considerando estos límites para las covarianzas, asumir la distribución Beta generalizada es la forma de mantener la flexibilidad de la distribución estándar Beta, además de la necesidad de controlar el rango entre 0 y 0,25, por tanto, la distribución prior para las covarianzas son:

$$\begin{aligned} cov_S &\sim genBeta(\alpha_{cov_S}, \beta_{cov_S}) \\ cov_C &\sim genBeta(\alpha_{cov_C}, \beta_{cov_C}) \end{aligned} \quad (1.34)$$

1.3.3.4 Distribución posterior para el modelo con efectos fijos

La función de distribución posterior conjunta, es igual a la multiplicación de la verosimilitud en (1.31) por las distribuciones prior Beta en las ecuaciones (1.26) y Beta generalizada en

las ecuaciones (1.34),

$$\begin{aligned}
 \mathcal{P}(\psi, \mathbf{D}|\mathbf{T}) &\propto \mathcal{L}(\psi) \times \pi^{\alpha_\pi-1} (1 - \pi)^{\beta_\pi-1} \\
 &\times \prod_{j=1}^2 S_j^{\alpha_{S_j}-1} (1 - S_j)^{\beta_{S_j}-1} cov_S^{\alpha_{cov_S}-1} (l_s - cov_S)^{\beta_{cov_S}-1} \\
 &\times \prod_{j=1}^2 C_j^{\alpha_{C_j}-1} (1 - C_j)^{\beta_{C_j}-1} cov_C^{\alpha_{cov_C}-1} (l_c - cov_C)^{\beta_{cov_C}-1}
 \end{aligned} \tag{1.35}$$

A partir de la distribución posterior (1.35) se deducen las distribuciones marginales posteriores para cada parámetro necesarias para el algoritmo muestreo de Gibbs (o el Metropolis-Hastings). Algunos parámetros pierden la característica de una distribución conjugada por no tener el formato de una distribución conocida.

$$\begin{aligned}
& \pi | \mathbf{D}, \alpha_\pi, \beta_\pi \sim \text{Beta} \left(\alpha_\pi + \sum_{i=1}^n d_i, \beta_\pi + \sum_{i=1}^n (1 - d_i) \right) \\
p(S_j | S_{3-j}, \mathbf{T}, \mathbf{D}, \alpha_{s_j}, \beta_{s_j}, cov_S, l_s, \alpha_{cov_S}, \beta_{cov_S}) & \propto \left[\left(\prod_{j=1}^2 S_j^{\sum_{i=1}^n t_{ji} d_i} (1 - S_j)^{\sum_{i=1}^n (1-t_{ji}) d_i} \right) + cov_S (-1)^{2-(t_{1i}+t_{2i})} \right] \\
& \quad \times \prod_{j=1}^2 S_j^{\alpha_{s_j}-1} (1 - S_j)^{\beta_{s_j}-1} cov_S^{\alpha_{cov_S}-1} (l_s - cov_S)^{\beta_{cov_S}-1} \\
p(C_j | C_{3-j}, \mathbf{T}, \mathbf{D}, \alpha_{c_j}, \beta_{c_j}, cov_C, l_c, \alpha_{cov_C}, \beta_{cov_C}) & \propto \left[\left(\prod_{j=1}^2 C_j^{\sum_{i=1}^n (1-t_{ji})(1-d_i)} (1 - C_j)^{\sum_{i=1}^n t_{ji}(1-d_i)} \right) + cov_C (-1)^{2-(t_{1i}+t_{2i})} \right] \\
& \quad \times \prod_{j=1}^2 C_j^{\alpha_{c_j}-1} (1 - C_j)^{\beta_{c_j}-1} cov_C^{\alpha_{cov_C}-1} (l_c - cov_C)^{\beta_{cov_C}-1} \\
p(cov_S | \mathbf{T}, \mathbf{D}, l_s, \alpha_{cov_S}, \beta_{cov_S}) & \propto \left[\left(\prod_{j=1}^2 S_j^{\sum_{i=1}^n t_{ji} d_i} (1 - S_j)^{\sum_{i=1}^n (1-t_{ji}) d_i} \right) + cov_S (-1)^{2-(t_{1i}+t_{2i})} \right] \\
& \quad \times cov_S^{\alpha_{cov_S}-1} (l_s - cov_S)^{\beta_{cov_S}-1} \\
p(cov_C | \mathbf{T}, \mathbf{D}, l_c, \alpha_{cov_C}, \beta_{cov_C}) & \propto \left[\left(\prod_{j=1}^2 C_j^{\sum_{i=1}^n (1-t_{ji})(1-d_i)} (1 - C_j)^{\sum_{i=1}^n t_{ji}(1-d_i)} \right) + cov_C (-1)^{2-(t_{1i}+t_{2i})} \right] \\
& \quad \times cov_C^{\alpha_{cov_C}-1} (l_c - cov_C)^{\beta_{cov_C}-1} \\
d_i | \psi, \mathbf{T} & \sim \text{Bernoulli} \left(\frac{A}{A+B} \right), \text{ donde} \\
A & = \pi \left(\prod_{j=1}^2 S_j^{t_{ji}} (1 - S_j)^{1-t_{ji}} + cov_S (-1)^{2-(t_{1i}+t_{2i})} \right) y \\
B & = (1 - \pi) \left(\prod_{j=1}^2 C_j^{1-t_{ji}} (1 - C_j)^{t_{ji}} + cov_C (-1)^{2-(t_{1i}+t_{2i})} \right)
\end{aligned}$$

1.3.3.5 Modelo dependiente con efectos aleatorios individuo

Para tener en cuenta la dependencia condicional entre las medidas del individuo realizadas con diferentes tests, [Qu et al. \[1996\]](#) propusieron el modelo de clases latentes con efectos aleatorios estimado via máxima verosimilitud y [Dendukuri & Joseph \[2001\]](#) propusieron la aproximación bayesiana de este modelo.

Para el caso de dos tests de diagnóstico con resultado binarios, se define por t_{jk} como el resultado del j th test T_{jk} para el k th individuo (positivo $t_{jk} = 1$ y negativo $t_{jk} = 0$), donde $k = 1, \dots, N$ y $j = 1, 2$. El modelo de efectos aleatorios incorpora un efecto aleatorio individuo. Este efecto individuo I_k modela la dependencia entre las medidas de los individuos realizadas con diversos tests, por que las observaciones dentro del mismo nivel del efecto están correlacionadas. Se asume que $I_k = i_k$ tiene distribución normal estándar. La variable aleatoria latente D tiene distribución Bernoulli; para cada individuo k , $D = d_k$, donde $d_k = 1$ en el caso de individuo enfermo y $d_k = 0$ para individuo no enfermo. Luego, la prevalencia de la enfermedad es representada por $\pi = P(D = 1)$ y para los tests, las sensibilidades son denotadas por S_1 y S_2 y las especificidades por C_1 y C_2 . La probabilidad del resultado positivo para el k th individuo en el j th test es

$$P(T_{jk} = 1 | D_k = d_k, I_k = i_k) = \Phi(a_{jd_k} + b_{d_k} i_k),$$

donde Φ representa la función de distribución acumulada $\mathcal{N}(0, 1)$ y a_{jd_k} , b_{d_k} son parámetros desconocidos y cuando $b_{d_k} = 0$ los tests son condicionalmente independientes. Con esta parametrización, la esperanza del efecto aleatorio es $E(b_{d_k} I_k) = 0$ y la varianza es $Var(b_{d_k} I_k) = b_{d_k}^2$. La sensibilidad de cada test es dada por la integración de todos los posibles valores de I_k , tal que

$$S_j = \int_{-\infty}^{\infty} P(T_{jk} = 1 | D_k = 1, I_k = i_k) d\Phi(I_k) = \Phi\left(\frac{a_{j1}}{\sqrt{1 + b_1^2}}\right) \quad (1.36)$$

donde $j = 1, 2$, y para la especificidad de cada test es

$$C_j = \int_{-\infty}^{\infty} P(T_{jk} = 0 | D_k = 0, I_k = i_k) d\Phi(I_k) = \Phi\left(\frac{-a_{j0}}{\sqrt{1 + b_0^2}}\right) \quad (1.37)$$

para $j = 1, 2$.

1.3.3.6 Verosimilitud para el modelo con efectos aleatorios individuo

Dado el estado de la enfermedad d_k y el efecto aleatorio individuo I_k , la función de verosimilitud es:

$$\begin{aligned} \mathcal{L} &\propto \prod_{k=1}^N P(T_{jk} = t_{jk} | \varpi, D_k = d_k, I_k = i_k) \\ &= \prod_{k=1}^N \left[\pi \prod_{j=1}^2 \Phi(a_{j1} + b_1 i_k)^{t_{jk}} [1 - \Phi(a_{j1} + b_1 i_k)]^{(1-t_{jk})} \right]^{d_k} \\ &\quad \times \left[(1 - \pi) \prod_{j=1}^2 \Phi(a_{j0} + b_0 i_k)^{(1-t_{jk})} [1 - \Phi(a_{j0} + b_0 i_k)]^{t_{jk}} \right]^{(1-d_k)} \end{aligned} \quad (1.38)$$

donde $\varpi = (\pi, a_{11}, a_{21}, b_1, a_{10}, a_{20}, b_0)$ es el vector de parámetros a estimar.

1.3.3.7 Prior para los parámetros del modelo con efectos aleatorios

Al igual que el caso de efectos fijos, las informaciones previas de las opiniones de los expertos, o el resultado de estudios previos, constituyen las creencias prior usadas para la prevalencia y las sensibilidades y especificidades de cada test. Las medias prior para estos parámetros se fijan al centro del intervalo de creencias.

Se asume que la prevalencia tiene una distribución prior Beta, $\pi \sim \text{Beta}(\alpha, \beta)$, y se utilizan los límites del intervalo de creencias para fijar los valores de los hiperparámetros α and β . Las expresiones α y β en función de los límites inferiores y superiores del intervalo de creencias fueron deducidas en la sección 1.3.1.2 con las ecuaciones (1.14).

En el modelo de efecto aleatorio individuo, las distribuciones prior para los parámetros desconocidos $(a_{11}, a_{21}, b_1, a_{10}, a_{20}, b_0)$ se asume que se distribuyen según una Normal con

media μ y desviación estándar σ , luego

$$a_{jd_k} \sim \mathcal{N}(\mu_{a_{jd_k}}, \sigma_{a_{jd_k}})$$

$$b_{d_k} \sim \mathcal{N}(\mu_{b_{d_k}}, \sigma_{b_{d_k}})$$

para $j = 1, 2; d_k = 0, 1; k = 1, \dots, N$.

La media μ_{b_1} de la distribución prior del modelo de efecto aleatorio individuo y la media de los coeficientes a_{11} y a_{21} ($\mu_{a_{11}}$ y $\mu_{a_{21}}$) son obtenidas de la ecuación de la covarianza entre tests entre los individuos enfermos:

$$cov_S = \int_{-\infty}^{\infty} \Phi(\mu_{a_{11}} + \mu_{b_1} i_k) \Phi(\mu_{a_{21}} + \mu_{b_1} i_k) d\Phi(i_k) - S_1 S_2 \quad (1.39)$$

y de las ecuaciones de las sensibilidades de los tests 1 e 2, deducidas de la ecuación (1.36), de forma que

$$\mu_{a_{j1}} = \Phi^{-1}(S_j) \sqrt{1 + \mu_{b_1}^2}, j = 1, 2 \quad (1.40)$$

Debido a que son tres ecuaciones involucrando integrales, (S_1, S_2, cov_S) , y tres incógnitas a calcular, $(\mu_{a_{11}}, \mu_{a_{21}}, \mu_{b_1})$, el procedimiento usado para encontrar simultáneamente estos valores es conocido como *bisectional search algorithm* [Thisted, 1988]. Un procedimiento similar fue desarrollado en código escrito en R [R Core Team, 2012] para resolver estas ecuaciones (Apéndice F). El mismo algoritmo fue usado para calcular $(\mu_{a_{10}}, \mu_{a_{20}}, \mu_{b_0})$. En este caso, la ecuación para la covarianza entre los dos tests entre los individuos no enfermos es:

$$cov_C = \int_{-\infty}^{\infty} \Phi(\mu_{a_{10}} + \mu_{b_0} i_k) \Phi(\mu_{a_{20}} + \mu_{b_0} i_k) d\Phi(i_k) - C_1 C_2 \quad (1.41)$$

De la ecuación de la especificidad para los tests 1 y 2 en la ecuación (1.37), se deduce que

$$\mu_{a_{j0}} = \Phi^{-1}(C_j) \sqrt{1 + \mu_{b_0}^2}, j = 1, 2 \quad (1.42)$$

Para poder solucionar estas ecuaciones es necesario fijar un valor para cov_S y cov_C . Se puede realizar fijando “arbitrariamente” el valor de estas covarianzas a la mitad del rango que se obtiene a partir de las ecuaciones (1.33), o estimarlas ajustando el modelo de efectos fijos definido en la sección 1.3.3.1.

La desviación estándar $\sigma_{a_{jd_k}}$ puede ser aproximada usando $1/4$ del rango de la información prior para S_1, S_2, C_1 y C_2 delimitado por los límites de las opiniones de los expertos para cada parámetro. De las ecuaciones (1.40) y (1.42), y sustituyendo las sensibilidades y las especificidades por sus límites, el valor mínimo y máximo para σ es obtenido. Luego, la información prior fijada para la desviación estándar es:

$$\sigma_{a_{jd_k}} = \frac{\Phi^{-1}(u)\sqrt{1 + \mu_{b_{d_k}}^2} - \Phi^{-1}(l)\sqrt{1 + \mu_{b_{d_k}}^2}}{4}$$

No es fácil determinar un valor aproximado para $\sigma_{b_{d_k}}$ dado que, por ejemplo, para el mismo valor de b_1 puede corresponder al rango completo de los valores de su respectiva sensibilidad (como ilustrado en la figura 1.9) ya que b_{d_k} tiene más peso en la covarianza entre tests y a_{jd_k} en la media.

Dada esta dificultad, la $\sigma_{b_{d_k}}$ puede ser fijada a 0,5, es decir, usando una distribución prior amplia (*wide prior*).

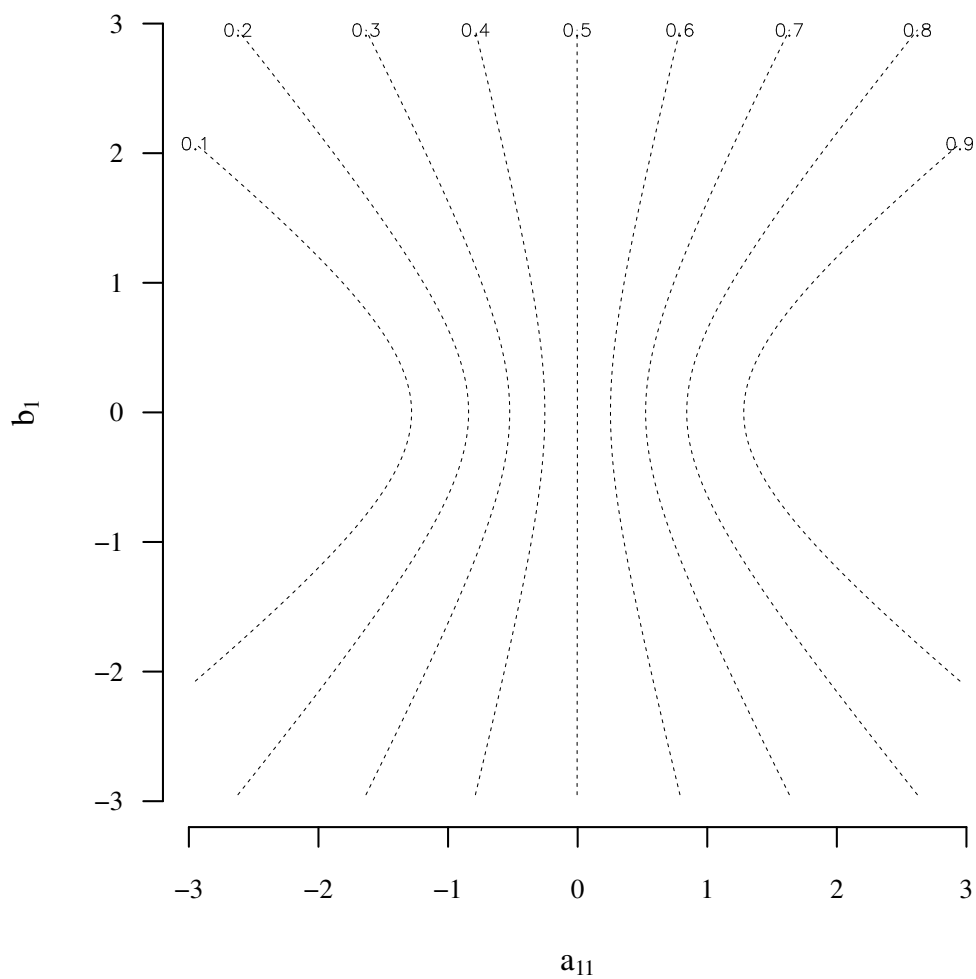


Figura 1.9: Gráfico de contorno para S_1 en el plano dado por sus respectivas coordenadas (a_{11}, b_1)

1.3.3.8 Distribución posterior para el modelo con efectos aleatorios

Para obtener la distribución posterior de los parámetros es necesario utilizar el algoritmo muestreo de Gibbs (o Metropolis-Hastings) y en consecuencia se requiere la especificación de las distribuciones condicionales completas de los parámetros. Estas distribuciones condicionales completas se resumen en:

$$\begin{aligned}
p(\pi \mid \mathbf{D}, \alpha_\pi, \beta_\pi) &\propto \pi^{\alpha_\pi + \sum_{k=1}^N d_k - 1} (1 - \pi)^{\beta_\pi + \sum_{k=1}^N (1 - d_k) - 1} \\
&\Rightarrow \pi \mid \mathbf{D}, \alpha_\pi, \beta_\pi \sim \text{Beta} \left(\alpha_\pi + \sum_{k=1}^N d_k, \beta_\pi + \sum_{k=1}^N (1 - d_k) \right), \\
p(d_k \mid \mathbf{T}, \varpi, i_k) &\propto \left[\pi \prod_{j=1}^2 \Phi(a_{j1} + b_1 i_k)^{t_{jk}} [1 - \Phi(a_{j1} + b_1 i_k)]^{(1 - t_{jk})} \right]^{d_k} \\
&\quad \times \left[(1 - \pi) \prod_{j=1}^2 \Phi(a_{j0} + b_0 i_k)^{(1 - t_{jk})} [1 - \Phi(a_{j0} + b_0 i_k)]^{t_{jk}} \right]^{(1 - d_k)} \\
&\Rightarrow d_k \mid \mathbf{T}, \varpi, i_k \sim \text{Bernoulli} \left(\frac{A}{A + B} \right), \\
&\quad \text{donde } A = \pi \prod_{j=1}^2 \Phi(a_{j1} + b_1 i_k)^{t_{jk}} [1 - \Phi(a_{j1} + b_1 i_k)]^{(1 - t_{jk})} \\
&\quad \text{y } B = (1 - \pi) \prod_{j=1}^2 \Phi(a_{j0} + b_0 i_k)^{(1 - t_{jk})} [1 - \Phi(a_{j0} + b_0 i_k)]^{t_{jk}}; \\
p(a_{j1}, b_1 \mid \mathbf{T}, \mathbf{D}, i_k) &\propto \prod_{k=1}^N \Phi(a_{j1} + b_1 i_k)^{t_{jk} d_k} [1 - \Phi(a_{j1} + b_1 i_k)]^{(1 - t_{jk}) d_k} d\Phi(a_{j1}, b_1), \\
p(a_{j0}, b_0 \mid \mathbf{T}, \mathbf{D}, i_k) &\propto \prod_{k=1}^N \Phi(a_{j0} + b_0 i_k)^{(1 - t_{jk})(1 - d_k)} [1 - \Phi(a_{j0} + b_0 i_k)]^{t_{jk}(1 - d_k)} d\Phi(a_{j0}, b_0), \\
p(i_k \mid \mathbf{T}, \mathbf{D}, \varpi) &\propto \prod_{j=1}^2 \Phi(a_{j1} + b_1 i_k)^{t_{jk} d_k} [1 - \Phi(a_{j1} + b_1 i_k)]^{(1 - t_{jk}) d_k} \\
&\quad \times \prod_{j=1}^2 \Phi(a_{j0} + b_0 i_k)^{(1 - t_{jk})(1 - d_k)} [1 - \Phi(a_{j0} + b_0 i_k)]^{t_{jk}(1 - d_k)}.
\end{aligned}$$

1.4 Inclusión de covariables en el modelo

El uso de covariable en estudios diagnósticos es una forma de mitigar la variabilidad que puede estar fuera del alcance exclusivo del test diagnóstico, es decir, la prevalencia y los parámetros de validez de los tests pueden ser estimados con más exactitud cuando el modelo contempla otras fuentes de información. La función de verosimilitud en (1.12) es bastante flexible cuando se incluye covariable. A veces es útil considerar para cada individuo los parámetros prevalencia (π_i), sensibilidad (S_i) o especificidad (C_i) en función de una o más covariables. Con esto, es posible identificar factores significativos que el test no es capaz de detectar de forma aislada. Las funciones de enlace para conectar los parámetros de interés a las covariables pueden ser diversas. En la literatura es común el uso del enlace logit por su popularidad y debido a la atractiva facilidad de interpretación directa de la Odds Ratio, ampliamente utilizada en diversos estudios de las ciencias de la salud [Achcar et al., 2005, Martínez et al., 2005, 2009, 2008].

En un ejemplo, Martínez et al. [2005] evaluaron el posible efecto de una covariable sobre los resultados de los tests diagnósticos y de la prevalencia de cáncer de próstata en dos grupos. Los autores propusieron los análisis de los datos con el enlace logit, en donde utilizaron una variable de grupo control (color de la piel con dos categorías, blanca y negra) para estimar en cada grupo la prevalencia y las sensibilidades y especificidades de dos tests diagnósticos (DRE, *digital rectal examination* y PSA, *prostate-specific antigen*). Para estos análisis, algunos sujetos fueron verificados con el test *gold standard* (biopsia) y otros no, lo que son denominados de análisis de datos “semi-latentes”, además los autores compararon los resultados de modelos de dependencia con los resultados de modelos de independencia. En otro ejemplo los autores propusieron modelos que se ajustan a la historia clínica o a las características demográficas de los pacientes cuando hay más de dos tests. Martínez et al. [2008] utilizaron la aproximación bayesiana para estimar la prevalencia, la sensibilidades y la especificidades de tres tests diagnósticos usados en el cribado de cáncer cervical, dependiendo de la edad y del embarazo de las mujeres participantes del estudio. Un estudio de simulación consideró la estimación de los parámetros para el caso general de modelos bayesianos con tests diagnósticos en ausencia de

gold standard, con aplicación en la enfermedad de chagas en pacientes de un hospital en Brasil, con inclusión de tres covariables, cada una codificada con dos categorías (edad, género y presencia de factor de riesgo) [de~Araujo~Pereira et~al., 2012]. El uso de covariables categóricas estratifica la población en dos o más categorías y así se permiten estimaciones menos sesgadas de los parámetros que se pretende estimar [Hui & Walter, 1980].

En el caso general, el parámetro puede depender de covariables, l . Usando el enlace probit, el modelo es $(\pi_i, S_i, C_i) = \Phi(\mathbf{X}\boldsymbol{\beta})$, donde $\mathbf{X} = (X_{i0}, X_{i1}, \dots, X_{il})^T$ es la matriz de covariables, X_{i0} es el vector constante $\mathbf{1}$ y $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_l)$ es el vector de parámetros desconocidos a estimar. Los parámetros $\boldsymbol{\beta}$ en este caso representa la “cantidad” o el “peso” que cada individuo tiene para aportar en la media de la prevalencia y en los parámetros de validez del test diagnóstico a partir de las fuentes de información que no son exclusivas del test.

Objetivos

1. Modelización de la prevalencia, sensibilidad y especificidad de dos o más tests diagnósticos en ausencia de *gold standard* para el caso que existan medidas repetidas de los tests por individuo.
 - (a) Proponer una extensión del modelo bayesiano con efectos aleatorios para estimar la prevalencia, la sensibilidad y la especificidad de dos o más tests diagnósticos en ausencia de *gold standard* para el caso que existan medidas repetidas de los tests por individuo.
 - (b) Aplicación de la extensión para evaluar el funcionamiento de dos tests para diagnosticar la presencia de parásitos intestinales, concretamente el parásito *Hookworm*.
 - (c) Comparación del modelo propuesto con el modelo que asume independencia entre tests y replicas y con los dos modelos de dependencia los cuales solo tienen en cuenta la dependencia condicional entre los tests.
2. Estimar la prevalencia de la infección por tuberculosis latente en niños y determinar la validez y la seguridad de dos tests (sin gold standard): Tuberculin Skin Test (TST) y Quantiferon-*Gold in tube* (QFT), teniendo en cuenta el tiempo y la intensidad de exposición.

Objetivo 1. Modelo individuo-test

3.1 Material y métodos

3.1.1 Objetivo 1.(a) Extensión del modelo bayesiano con efectos aleatorios para estimar la prevalencia, la sensibilidad y la especificidad de dos o más tests diagnósticos en ausencia de *gold standard* para el caso que existan medidas repetidas de los tests por individuo

En algunos estudios clínicos, varias lecturas de un mismo individuo son realizadas con el mismo test en las mismas condiciones (mediciones replicadas), y por lo tanto, estas mediciones para cada individuo no son condicionalmente independientes. Hasta el momento, la metodología estadística no tiene en cuenta esta dependencia entre las mediciones replicadas. Nuestra propuesta es una extensión del modelo de efectos aleatorios propuesto por Dendukuri y Joseph [Dendukuri & Joseph, 2001]. Esta extensión posibilita medir varias fuentes de variabilidad *within* individuos. Para eso, el modelo incluye dos efectos aleatorios independientes: el efecto aleatorio individuo, que tiene en cuenta la correlación entre las mediciones de los mismos individuos realizadas por múltiples tests; y el efecto aleatorio de interacción individuo-test que tiene en cuenta la correlación entre las mediciones replicadas de los individuos con el mismo test. Mediante la inclusión de estos efectos aleatorios se asume que las propiedades del test (sensibilidad y especificidad) no son constantes. El efecto aleatorio individuo induce variaciones en

la validez del test entre los individuos, pero no cambia con el test. Por lo tanto, la intensidad o la gravedad de la enfermedad en cada individuo hace que la validez del test varíe entre los individuos debido a que la gravedad individual interfiere en la detección de la enfermedad. Sin embargo, con el efecto aleatorio de interacción individuo-test la facilidad de detección de la enfermedad depende de las características o propiedades de cada test, por lo tanto, la detección de cada test varía en función de la gravedad de la enfermedad en cada individuo.

En este capítulo se presenta el modelo y la aproximación Bayesiana para estimar la prevalencia, la sensibilidad y especificidad cuando los tests son replicados y el resultado del diagnóstico es de respuesta binaria. La metodología propuesta se aplica para evaluar la validez de dos tests que diagnostican la presencia de parásitos intestinales, específicamente el parásito *Hookworm*. Por otra parte, se llevó a cabo un estudio de simulación para evaluar el funcionamiento de la metodología propuesta y el modelo propuesto se comparó con un modelo que asume independencia entre tests replicados, y también con los dos modelos de dependencia propuestos por [Dendukuri & Joseph, 2001] que llevan en cuenta la dependencia condicional sólo entre los tests.

3.1.1.1 Modelo con efecto aleatorio Individuo-test

Se supone que los individuos del estudio son diagnosticados por el mismo test varias veces bajo las mismas condiciones de laboratorio. T_{kjr} indica los resultados del k th individuo para el j th test en la r th repetición con $k = 1, \dots, N$, $j = 1, \dots, J$ y $r = 1, \dots, R$. Utilizando un efecto aleatorio, las observaciones dentro del mismo nivel del efecto aleatorio están correlacionadas. Por tanto, se consideran dos vectores de efectos aleatorios: el efecto aleatorio individuo I_k para modelar la dependencia entre las medidas del individuo realizadas con varios tests, y el vector de efecto aleatorio interacción individuo-test, I_{kj} que tiene en cuenta la dependencia entre las repeticiones del individuo. Las variables aleatorias $I_k = i_k$ and $I_{kj} = i_{kj}$ tienen distribución normal estándar. En ausencia de un *gold standard*, el verdadero estado de la enfermedad del sujeto es desconocido, y puede ser representado por una variable aleatoria latente D con distribución Bernoulli; para cada individuo k , $D = d_k$, donde $d_k = 1$ en el caso de la presencia de la enfermedad y $d_k = 0$ para la ausencia de la enfermedad. La prevalencia

de la enfermedad $P(D = 1)$ es representada por π . Para j tests, la sensibilidad se denota por S_j y la especificidad por C_j .

Considerando que el resultado del test es binario para cada individuo k (positivo $T_{kjr} = 1$ y negativo $T_{kjr} = 0$), la probabilidad de un resultado positivo en el k th individuo para el j th test en la r th replica es:

$$P(T_{kjr} = 1 | D_k = d_k, I_k = i_k, I_{kj} = i_{kj}) = \Phi(a_{jd_k} + b_{1d_k}i_k + b_{2d_k}i_{kj}),$$

donde Φ representa la función de distribución acumulada $\mathcal{N}(0, 1)$ y a_{jd_k} , b_{1d_k} y b_{2d_k} son los parámetros desconocidos. Con esta parametrización, las esperanzas de los efectos aleatorios son $E(b_{1d_k}I_k) = E(b_{2d_k}I_{kj}) = 0$ y las varianzas son $Var(b_{1d_k}I_k) = b_{1d_k}^2$ y $Var(b_{2d_k}I_{kj}) = b_{2d_k}^2$.

La sensibilidad de cada test se obtiene mediante la integración de todos los posibles valores de I_k y I_{kj} ,

$$\begin{aligned} S_j &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(T_{kjr} = 1 | D_k = 1, I_k = i_k, I_{kj} = i_{kj}) d\Phi(I_k) d\Phi(I_{kj}) \\ &= \Phi\left(\frac{a_{j1}}{\sqrt{1 + b_{11}^2 + b_{21}^2}}\right) \end{aligned} \tag{3.1}$$

de forma análoga, la especificidad es:

$$\begin{aligned} C_j &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(T_{kjr} = 0 | D_k = 0, I_k = i_k, I_{kj} = i_{kj}) d\Phi(I_k) d\Phi(I_{kj}) \\ &= \Phi\left(\frac{-a_{j0}}{\sqrt{1 + b_{10}^2 + b_{20}^2}}\right) \end{aligned} \tag{3.2}$$

3.1.1.2 Estimación de los parámetros del modelo

La función de verosimilitud, condicionada al estado de la enfermedad d_k y los efectos aleatorios I_k y I_{kj} es:

$$\begin{aligned} \mathcal{L} &\propto \prod_{k=1}^N \prod_{r=1}^R P(T_{k1r} = t_{k1r}, \dots, T_{kjr} = t_{kjr} | \varpi, D_k = d_k, I_k = i_k, I_{k1} = i_{k1}, \dots, I_{kj} = i_{kj}) \\ &= \prod_{k=1}^N \left[\pi \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})^{t_{kjr}} [1 - \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})]^{(1-t_{kjr})} \right]^{d_k} \\ &\quad \times \left[(1 - \pi) \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})^{(1-t_{kjr})} [1 - \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})]^{t_{kjr}} \right]^{(1-d_k)} \end{aligned} \quad (3.3)$$

donde $\varpi = (\pi, a_{j1}, b_{11}, b_{21}, a_{j0}, b_{10}, b_{20})$ es el vector de parámetros a estimar. Los parámetros ϖ son estimados mediante el Teorema de Bayes (sección 1.2.1), combinando la función de verosimilitud de los datos (3.3) y la información prior dada en la sección 3.1.1.3 para obtener la distribución posterior de estos parámetros.

3.1.1.3 Distribución prior

Se asume que la prevalencia tiene una distribución prior $\text{Beta}(\alpha, \beta)$. La distribución prior para los parámetros desconocidos $(a_{j1}, b_{11}, b_{21}, a_{j0}, b_{10}, b_{20})$ relacionados con la media de los parámetros de validez de los tests (S_j and C_j) se supone Normal con media μ y desviación típica σ , luego:

$$a_{jd_k} \sim \mathcal{N}(\mu_{a_{jd_k}}, \sigma_{a_{jd_k}})$$

$$b_{1d_k} \sim \mathcal{N}(\mu_{b_{1d_k}}, \sigma_{b_{1d_k}})$$

$$b_{2d_k} \sim \mathcal{N}(\mu_{b_{2d_k}}, \sigma_{b_{2d_k}})$$

La modelización bayesiana se implementa mediante el programa WinBUGS [Lunn et al., 2000] con el paquete estadístico R [R Core Team, 2012] a partir de la librería R2WinBUGS [Sturtz et al., 2005]. El código WinBUGS para la ejecución del modelo propuesto se muestra en el Apéndice C.

3.1.1.4 Distribución posterior

La implementación del algoritmo muestreo de Gibbs (o Metropolis-Hastings) requiere la especificación de las distribuciones condicionales completas de los parámetros, es decir, las distribuciones condicionadas de cada parámetro dado los valores de los otros parámetros.

$$p(\pi \mid \mathbf{D}, \alpha_\pi, \beta_\pi) \propto \pi^{\alpha_\pi + \sum_{k=1}^N d_k - 1} (1 - \pi)^{\beta_\pi + \sum_{k=1}^N (1 - d_k) - 1}$$

$$\Rightarrow \pi \mid \mathbf{D}, \alpha_\pi, \beta_\pi \sim \text{Beta} \left(\alpha_\pi + \sum_{k=1}^N d_k, \beta_\pi + \sum_{k=1}^N (1 - d_k) \right),$$

$$p(d_k \mid \mathbf{T}, \varpi, i_k, i_{kj})$$

$$\propto \left[\pi \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})^{t_{kjr}} [1 - \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})]^{(1-t_{kjr})} \right]^{d_k}$$

$$\times \left[(1 - \pi) \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})^{(1-t_{kjr})} [1 - \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})]^{t_{kjr}} \right]^{(1-d_k)}$$

$$\Rightarrow d_k \mid \mathbf{T}, \varpi, i_k, i_{kj} \sim \text{Bernoulli} \left(\frac{A}{A + B} \right),$$

$$\text{donde } A = \pi \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})^{t_{kjr}} [1 - \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})]^{(1-t_{kjr})}$$

$$\text{y } B = (1 - \pi) \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})^{(1-t_{kjr})} [1 - \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})]^{t_{kjr}};$$

$$p(a_{j1}, b_{11}, b_{21} \mid \mathbf{T}, \mathbf{D}, i_k, i_{kj}) \propto \prod_{k=1}^N \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})^{t_{kjr}d_k}$$

$$\times [1 - \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})]^{(1-t_{kjr})d_k} d\Phi(a_{j1}, b_{11}, b_{21}),$$

$$p(a_{j0}, b_{10}, b_{20} \mid \mathbf{T}, \mathbf{D}, i_k, i_{kj}) \propto \prod_{k=1}^N \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})^{(1-t_{kjr})(1-d_k)}$$

$$\times [1 - \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})]^{t_{kjr}(1-d_k)} d\Phi(a_{j0}, b_{10}, b_{20}),$$

$$p(i_k, i_{kj} \mid \mathbf{T}, \mathbf{D}, \varpi) \propto \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})^{t_{kjr}d_k} [1 - \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})]^{(1-t_{kjr})d_k}$$

$$\times \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})^{(1-t_{kjr})(1-d_k)} [1 - \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})]^{t_{kjr}(1-d_k)}.$$

3.1.2 Objetivo 1.(b) Aplicación de la extensión para evaluar el funcionamiento de dos tests para diagnosticar la presencia de parásitos intestinales

El estudio se llevó a cabo en 2009 en la comunidad agrícola de Río Pardo del municipio de Presidente Figueiredo y fue aprobado previamente por el Comité de Ética en Investigación de la Fundación Oswaldo Cruz (Protocolo 384/07 del 20/08/2007). Esta comunidad se encuentra aproximadamente a 160 km al norte de la ciudad de Manaus, Estado de Amazonas, Brasil. Una muestra de heces frescas se obtuvo de cada niño y adulto que participaron en un proyecto de transmisión de patógenos intestinales. Dos técnicas de diagnóstico de sedimentación espontánea se utilizaron para detectar la infección por parásitos intestinales: Lutz (T_1) and Paratest (T_2). Las muestras fueron inicialmente procesadas en una unidad de laboratorio local ubicada en la comunidad. Primero cada espécimen se homogeneizó mediante la agitación con una espátula de madera durante al menos 1 minuto. Después de la homogeneización, cada muestra fue dividida en tres submuestras. Un total de 143 individuos fueron analizados [Gonçalves et al., 2014]. Se encontraron varios parásitos intestinales comunes, pero en esta aplicación se considera el más prevalente, el parásito *Hookworm*. La sensibilidad y especificidad del test Lutz se denotan por S_1 y C_1 y del Paratest S_2 y C_2 , respectivamente.

El objetivo es estimar la prevalencia de *Hookworm* y evaluar la sensibilidad y especificidad de las técnicas de Lutz and Paratest. Para los datos de parásitos *Hookworm*, las opiniones científicas de los expertos acerca de los dos tests variaron desde 0,3 a 0,6 para las sensibilidades, desde 0,9 a 1,0 para las especificidades y desde 0,0 a 0,5 para la prevalencia de la infección. Esta información se utilizó para todos los cuatro modelos. Los parámetros de la distribución beta para la prevalencia se calcularon utilizando la ecuación (1.14). Además, se utilizaron la misma distribución prior para prevalencia en los cuatro modelos.

3.1.2.1 Información prior

Para definir los parámetros de las distribuciones prior, las opiniones de expertos sobre la prevalencia, sensibilidades y especificidades de los tests son resumidas en los intervalos

delimitados por l (límite inferior) y u (superior límite).

Se supone que la prevalencia tiene una distribución prior $\text{Beta}(\alpha, \beta)$ y los límites se utilizan para fijar los valores de los hiperparámetros α and β . Por tanto, α y β son calculados mediante la fijación de la media de la distribución Beta con el punto medio del intervalo de las opiniones de los expertos, y la fijación de la desviación típica de la distribución Beta a $1/4$ del rango del mismo intervalo. Las expresiones, que dependen de los límites l y u fueron deducidas en 1.14.

Para simplificar la información prior necesaria, hemos tenido en cuenta las mismas opiniones de los expertos para la media de la distribución del efecto aleatorio de interacción individuo-test ($\mu_{b_{21}}$ y $\mu_{b_{20}}$). La información se obtuvo a través de modelo lineal generalizado mixto, con el que se modeló la probabilidad de un resultado positivo en función del efecto fijo test, y del efecto individuo y la interacción individuo-test como efectos aleatorios. De esta forma, la estimación de la desviación típica de la interacción individuo-test se usó como información de $\mu_{b_{21}}$ y $\mu_{b_{20}}$. Mediante la fijación de la media de la distribución del efecto aleatorio de interacción individuo-test, la media de la distribución prior del efecto aleatorio individuo, $\mu_{b_{11}}$ y la media de los coeficientes a_{11} y a_{21} ($\mu_{a_{11}}$ y $\mu_{a_{21}}$) se obtienen a partir de las ecuaciones de la covarianza entre los dos tests para los individuos con la enfermedad

$$\begin{aligned} \text{cov}S_{12} = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\mu_{a_{11}} + \mu_{b_{11}}i_k + \mu_{b_{21}}i_{k1}) \\ & \times \Phi(\mu_{a_{21}} + \mu_{b_{11}}i_k + \mu_{b_{21}}i_{k2}) d\Phi(I_k) d\Phi(I_{kj}) - S_1 S_2 \end{aligned} \quad (3.4)$$

y de la ecuación de la sensibilidad para los tests 1 y 2, deducida de la ecuación (3.1),

$$\mu_{a_{j1}} = \Phi^{-1}(S_j) \sqrt{1 + \mu_{b_{11}}^2 + \mu_{b_{21}}^2}, j = 1, 2 \quad (3.5)$$

Los cálculos de los valores de $\mu_{a_{11}}$, $\mu_{a_{21}}$ y $\mu_{b_{11}}$ pueden ser hallados mediante *bisectional search algorithm* [Thisted, 1988] o con el *script* R dado en el Apéndice F. Un procedimiento similar fue usado para calcular la $\mu_{a_{10}}$, $\mu_{a_{20}}$ and $\mu_{b_{10}}$. En este caso, la ecuación de la covarianza

entre los dos tests para los individuos sin la enfermedad es

$$\begin{aligned} covC_{12} = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\mu_{a_{10}} + \mu_{b_{10}}i_k + \mu_{b_{20}}i_{k1}) \\ & \times \Phi(\mu_{a_{20}} + \mu_{b_{10}}i_k + \mu_{b_{20}}i_{k2}) d\Phi(I_k) d\Phi(I_{kj}) - C_1 C_2 \end{aligned} \quad (3.6)$$

A partir de la ecuación de la especificidad para los tests 1 y 2 en la ecuación (3.2), se deduce que

$$\mu_{a_{j0}} = \Phi^{-1}(C_j) \sqrt{1 + \mu_{b_{10}}^2 + \mu_{b_{20}}^2}, j = 1, 2 \quad (3.7)$$

La desviación típica $\sigma_{a_{jd_k}}$ puede ser aproximada usando 1/4 del rango de la información prior de S_1, S_2, C_1 y C_2 denotado por l y u . De las ecuaciones (3.5) y (3.7), y sustituyendo las sensibilidades y especificidades por sus límites, es obtenido el máximo y el mínimo límite para σ . luego, la información prior es fijada en:

$$\frac{\Phi^{-1}(u) \sqrt{1 + \mu_{b_{1d_k}}^2 + \mu_{b_{2d_k}}^2} - \Phi^{-1}(l) \sqrt{1 + \mu_{b_{1d_k}}^2 + \mu_{b_{2d_k}}^2}}{4} \quad (3.8)$$

3.1.3 Objetivo 1.(c) Simulación de datos para evaluar la calidad de los modelos que estiman la validez de los tests diagnósticos

Se implementó un estudio de simulación, considerando cuatro escenarios con el fin de examinar el sesgo relativo (sesgo dividido por el valor real del parámetro), la desviación típica de las estimaciones, el error cuadrático medio y la cobertura media del intervalo de credibilidad de las estimaciones Bayesianas. Para fijar diferentes escenarios, los valores de los efectos aleatorios, b_{1d_k} and b_{2d_k} , fueron elegidos de acuerdo con el peso de estos componentes en la variabilidad total. Se utilizaron las componentes de partición de la varianza (VPC) definidos por Goldstein [2011] como instrumento de medición del “tamaño” o “peso” del efecto aleatorio introducido en cada escenario. Para calcular el VPC con respuesta binaria se usó el método de simulación propuesto por Browne et~al. [2005]. La estrategia de calcular el coeficiente de partición de varianzas surgió de la necesidad de medir la cantidad de varianza explicada a partir

de dos fuentes: individuo y individuo-test. Los escenarios fueron planteados considerando los pesos de ambos efectos aleatorios y los valores fijados para los parámetros son presentados en la Tabla 3.1. De esta manera, el estudio de simulación se planificó con diferentes escenarios de interés que permitió el mecanismo de procesamiento para evaluar el desempeño de cada modelo. El funcionamiento del modelo de efectos aleatorios individuo y de interacción individuo-test ($SUB \times TST$) propuesto se comparó con los resultados obtenidos con el modelo de efecto aleatorio individuo (SUB), con el modelo de dependencia condicional de efectos fijos (DEP) y con el modelo de independencia condicional (IND).

Cuatro escenarios fueron planteados considerando los pesos de ambos efectos aleatorios. Los valores fijados para los parámetros son presentados en la Tabla 3.1.

Sc	Par	Media	a_{jd_k}	b_{1d_k}	b_{2d_k}	VPC_{SUB}	$VPC_{SUB \times TST}$
I	S_1	0,52	0,08	0,80	0,20	26,0 %	1,6 %
	S_2	0,35	-0,51			25,3 %	1,6 %
	C_1	0,91	2,21	1,30	0,10	36,3 %	0,3 %
	C_2	0,95	2,70			33,0 %	0,3 %
	Prev	0,30	—	—	—	—	—
II	S_1	0,52	0,09	0,80	0,60	21,8 %	12,4 %
	S_2	0,35	-0,56			21,1 %	12,3 %
	C_1	0,91	2,58	1,30	1,00	23,4 %	21,0 %
	C_2	0,95	3,16			20,1 %	20,7 %
	Prev	0,30	—	—	—	—	—
III	S_1	0,52	0,09	0,80	1,00	16,5 %	26,7 %
	S_2	0,37	-0,54			16,0 %	26,6 %
	C_1	0,89	2,58	1,30	1,30	19,8 %	30,0 %
	C_2	0,93	3,16			17,0 %	29,6 %
	Prev	0,25	—	—	—	—	—
IV	S_1	0,80	1,77	0,67	1,73	5,5 %	47,6 %
	S_2	0,35	-0,81			6,5 %	49,3 %
	C_1	0,70	1,69	0,86	2,94	4,5 %	66,4 %
	C_2	0,75	2,17			4,2 %	66,0 %
	Prev	0,75	—	—	—	—	—

Sc, escenario; Par, parámetro: S_1 , sensibilidad del test 1; S_2 , sensibilidad del test 2; C_1 , especificidad del test 1; C_2 , especificidad del test 2; Prev, prevalencia. VPC_{SUB} , componente de partición de la varianza del modelo SUB; $VPC_{SUB \times TST}$, componente de partición de la varianza del modelo $SUB \times TST$.

Tabla 3.1: Verdaderos valores de los parámetros de acuerdo con los escenarios (I-IV)

Escenario I - Simulación de datos con muy débil efecto aleatorio de interacción individuo-test y con moderado efecto aleatorio individuo. Para el escenario I, los valores de los parámetros fueron con valores parecidos a las estimaciones obtenidos en los datos de parásitos intestinales *Hookworm* descritos en la Tabla 3.3 de la sección 3.1.2. Las sensibilidades medias fueron ligeramente modificadas de modo que una podría ser mayor que 0,50 y otra inferior a este valor. La VPC del efecto aleatorio individuo-test fue cercana a cero (Tabla 3.1).

Escenario II - Simulación de datos con bajo efecto aleatorio de interacción individuo-test y con moderado efecto aleatorio individuo. Se incrementó la variabilidad del individuo-test en el escenario II. Este escenario enfatiza las diferencias entre el desempeño de los modelos debido a la mayor presencia del efecto individuo-test. La VPC para el efecto individuo fue de alrededor de 20 %, siendo el 15 % para el individuo-test (Tabla 3.1)

Escenario III - Simulación de datos con moderado efecto aleatorio de interacción individuo-test y con débil efecto aleatorio individuo. En este escenario, la variabilidad del efecto individuo-test fue mayor que la variabilidad del efecto individuo. La VPC para el efecto individuo fue de aproximadamente 15 %, siendo superior al 25 % para el efecto individuo-test (Tabla 3.1).

Escenario IV - Simulación de datos con muy elevado efecto aleatorio de interacción individuo-tests y con muy débil efecto aleatorio individuo. El escenario IV es una exageración. Mucho mayor peso fue dado para la variabilidad individuo-test (VPC casi 49 % para los individuos con la enfermedad y 66 % para los individuos sin la enfermedad) que para la variabilidad individuo (VPC entre 4 % y 7 %), y por tanto, la prevalencia fue alrededor de 0,75 (Tabla 3.1).

Para los cuatro escenarios descritos anteriormente, la distribución prior de los parámetros fue informativa para todos los modelos, y se utilizó la misma opinión de los expertos para todos los modelos en cada escenario. Esta información se muestra en la Tabla 3.2. La definición de distribución prior de acuerdo con el modelo es:

- Modelo IND. Se considera que S_1, S_2, C_1, C_2 tienen una distribución prior Beta con parámetros determinados por la ecuación (1.14).
- Modelo DEP. La covarianza entre los tests entre los individuos con y sin enfermedad se considera que tienen distribución Beta, con hiperparámetros definidos iguales a los de [Dendukuri & Joseph \[2001\]](#).
- Modelo SUB. Si requiere la distribución prior de los parámetros $a_{11}, a_{21}, a_{10}, a_{20}, b_{11}, b_{10}$. La media de la distribución prior de estos parámetros se calcula mediante la fijación de la sensibilidad y la especificidad en el punto medio del intervalo de la opinión de los expertos, y la covarianza entre los tests se fijó en el punto medio del intervalo de credibilidad (mediana) de la estimación *a posteriori* dada por el modelo DEP. Utilizando las ecuaciones (3.4), (3.5), (3.6), (3.7) y estableciendo igual a cero los parámetros en función del SUB×TST, se obtuvieron las medias de la prior. La desviación típica de estas distribuciones prior se fijó utilizando la ecuación (3.8). Sin embargo, la desviación típica de b_{11} y b_{10} fueron fijadas en 0,5.
- Modelo SUB×TST. El mismo proceso se realiza para obtener los hiperparámetros de las distribuciones prior de $a_{11}, a_{21}, a_{10}, a_{20}, b_{11}, b_{10}$, pero la media de los b_{21} y b_{20} se calculó como se mencionó anteriormente en la sección 3.1.2.1, y sus desviaciones típicas se fijó conforme la Tabla 3.2.

El número de individuos se fijó en 60, y 500 simulaciones fueron generadas para cada escenario. Para garantizar la convergencia de las cadenas en el estudio de simulación y para ahorrar tiempo de cálculo, se estudiaron previamente el número aproximadamente óptimo de iteraciones para la fase *burning*. Por otra parte, se fijó una condición de convergencia en la que el valor del error de Monte Carlo para cada parámetro fue inferior al 5 % de la desviación típica de la muestra [[Spiegelhalter et al., 2003](#)] de los valores simulados y al final se confirmaron las convergencias usando los métodos de diagnóstico de convergencia descritos en la sección 1.2.4.

Sc	Par	Media	Inf	Sup	Modelo IND o DEP		Modelo SUB				Modelo SUB×TST								
					Alpha	Beta	Opinión (Expertos)		Distr. Normal		Distr. Normal		Distr. Normal						
							Distr. Beta	Distr. Normal	Par	Media	SD	Par	Media	SD	Par	Media	SD		
I	S ₁	0,52	0,40	0,62	42,00	40,14	a ₁₁	0,09	0,19	b ₁₁	0,81	0,10	a ₁₁	0,08	0,18	b ₁₁	0,80	0,10	
	S ₂	0,35	0,24	0,45	29,20	55,25	a ₂₁	-0,53	0,17	b ₂₁	-	-	a ₂₁	-0,51	0,18	b ₂₁	0,20	0,10	
	C ₁	0,91	0,86	0,99	55,40	4,26	a ₁₀	2,22	0,62	b ₁₀	1,32	0,10	a ₁₀	2,21	0,60	b ₁₀	1,30	0,10	
	C ₂	0,95	0,90	1,00	69,80	3,88	a ₂₀	2,69	0,74	b ₂₀	-	-	a ₂₀	2,70	0,75	b ₂₀	0,10	0,10	
	Prev	0,30	0,19	0,40	22,60	53,42													
	CovS ₁₂	0,08	0,00	0,17	3,58	39,04													
CovC ₁₂	0,02	0,00	0,05	3,88	166,94														
II	S ₁	0,52	0,40	0,62	42,00	40,14	a ₁₁	0,10	0,22	b ₁₁	0,82	0,50	a ₁₁	0,09	0,20	b ₁₁	0,80	0,50	
	S ₂	0,35	0,24	0,45	29,20	55,25	a ₂₁	-0,58	0,19	b ₂₁	-	-	a ₂₁	-0,56	0,20	b ₂₁	0,60	0,50	
	C ₁	0,91	0,86	0,99	55,40	4,26	a ₁₀	2,59	0,71	b ₁₀	1,31	0,50	a ₁₀	2,58	0,70	b ₁₀	1,30	0,50	
	C ₂	0,95	0,90	1,00	69,80	3,88	a ₂₀	3,15	0,87	b ₂₀	-	-	a ₂₀	3,16	0,88	b ₂₀	1,00	0,50	
	Prev	0,30	0,19	0,40	22,60	53,42													
	CovS ₁₂	0,08	0,00	0,17	3,58	39,04													
CovC ₁₂	0,02	0,00	0,05	3,88	166,94														
III	S ₁	0,52	0,37	0,67	22,55	20,82	a ₁₁	0,12	0,33	b ₁₁	0,82	0,50	a ₁₁	0,09	0,31	b ₁₁	0,80	0,50	
	S ₂	0,37	0,25	0,45	31,50	58,50	a ₂₁	-0,56	0,21	b ₂₁	-	-	a ₂₁	-0,54	0,22	b ₂₁	1,00	0,50	
	C ₁	0,89	0,87	0,96	139,66	12,97	a ₁₀	2,60	0,34	b ₁₀	1,31	0,50	a ₁₀	2,58	0,33	b ₁₀	1,30	0,50	
	C ₂	0,93	0,92	0,98	199,61	10,51	a ₂₀	3,12	0,32	b ₂₀	-	-	a ₂₀	3,16	0,34	b ₂₀	1,30	0,50	
	Prev	0,25	0,00	0,50	2,75	8,25													
	CovS ₁₂	0,09	0,00	0,18	3,56	36,49													
CovC ₁₂	0,03	0,00	0,06	3,84	119,57														
IV	S ₁	0,80	0,70	0,98	22,20	4,23	a ₁₁	1,75	0,79	b ₁₁	0,65	0,50	a ₁₁	1,77	0,81	b ₁₁	0,67	0,30	
	S ₂	0,35	0,20	0,40	24,90	58,10	a ₂₁	-0,80	0,32	b ₂₁	-	-	a ₂₁	-0,81	0,31	b ₂₁	1,73	0,30	
	C ₁	0,70	0,65	0,95	21,96	5,49	a ₁₀	1,68	1,00	b ₁₀	0,85	0,50	a ₁₀	1,69	0,75	b ₁₀	0,86	0,30	
	C ₂	0,75	0,70	0,98	22,20	4,23	a ₂₀	2,21	1,26	b ₂₀	-	-	a ₂₀	2,17	0,75	b ₂₀	2,94	0,30	
	Prev	0,75	0,50	1,00	8,25	2,75													
	CovS ₁₂	0,04	0,00	0,07	3,83	105,46													
CovC ₁₂	0,09	0,00	0,18	3,56	37,15														

Sc, escenario; Par, parámetro; Prev, prevalencia; S₁, sensibilidad del test 1; S₂, sensibilidad del test 2; C₁, especificidad del test 1; C₂, especificidad del test 2; covS₁₂, covarianza entre los tests para los individuos enfermos; covC₁₂, covarianza entre los tests para los individuos no enfermos. SUB×TST, considera dos efectos aleatorios: individuo y interacción individuo-test; SUB, considera el efecto aleatorio individuo; DEP, modelo dependiente (fixed effect); IND, modelo independiente.

Tabla 3.2: Distribución prior y opinión de los expertos considerados en los escenarios

3.2 Resultados

3.2.1 Datos *Hookworm*

En la tabla 3.3 se resumen las estimaciones de los parámetros para los cuatro modelos. La prevalencia de la infección por *Hookworm* fue de 0,03 con una baja sensibilidad de 0,45 para T₁ y 0,43 para T₂ y altas especificidades de 0,91 para T₁ y 0,96 para T₂.

Los resultados de los modelos SUB y SUB×TST fueron muy similares, ya que el peso del efecto aleatorio SUB×TST fue muy bajo, con una VPC para los individuos con enfermedad de 1,7 % y para los individuos sin enfermedad de 0,2 %. En comparación con los modelos de efectos aleatorios, los modelos DEP e IND presentan pocos cambios en las propiedades de los tests (sensibilidad y especificidad). La mayor diferencia se observó con la prevalencia que aumentó más del doble, y su intervalo de credibilidad fue más largo.

Parámetro	Modelo			
	SUB×TST	SUB	DEP	IND
S ₁	0,45 (0,29;0,59)	0,44 (0,27;0,60)	0,43 (0,28;0,58)	0,47 (0,33;0,60)
S ₂	0,43 (0,26;0,58)	0,42 (0,25;0,58)	0,41 (0,26;0,58)	0,40 (0,27;0,56)
C ₁	0,91 (0,88;0,94)	0,92 (0,88;0,94)	0,92 (0,89;0,96)	0,94 (0,91;0,97)
C ₂	0,96 (0,93;0,98)	0,96 (0,93;0,98)	0,97 (0,95;0,99)	0,98 (0,96;0,99)
Prev	0,03 (0,01;0,08)	0,03 (0,01;0,08)	0,07 (0,02;0,15)	0,09 (0,04;0,17)
covS ₁₂	—	—	0,07 (0,02;0,16)	—
covC ₁₂	—	—	0,01 (0,00;0,02)	—
a ₁₁	-0,19 (-0,69;0,32)	-0,19 (-0,70;0,31)	—	—
a ₂₁	-0,26 (-0,76;0,26)	-0,26 (-0,77;0,25)	—	—
a ₁₀	2,11 (1,78;2,49)	2,07 (1,75;2,43)	—	—
a ₂₀	2,63 (2,25;3,03)	2,59 (2,23;2,98)	—	—
b ₁₁	0,44 (0,00;2,95)	0,46 (0,00;2,98)	—	—
b ₂₁	0,13 (0,00;1,45)	—	—	—
b ₁₀	1,27 (0,66;2,18)	1,25 (0,72;2,03)	—	—
b ₂₀	0,05 (0,00;0,46)	—	—	—
VPC _{SUB} con enfermedad T ₁	18,8 %	20,7 %	—	—
VPC _{SUB} con enfermedad T ₂	18,7 %	20,6 %	—	—
VPC _{SUB} sin enfermedad T ₁	28,7 %	29,6 %	—	—
VPC _{SUB} sin enfermedad T ₂	24,4 %	25,4 %	—	—
VPC _{SUB×TST} con enfermedad T ₁	1,7 %	—	—	—
VPC _{SUB×TST} con enfermedad T ₂	1,7 %	—	—	—
VPC _{SUB×TST} sin enfermedad T ₁	0,2 %	—	—	—
VPC _{SUB×TST} sin enfermedad T ₂	0,2 %	—	—	—

Parámetro: Prev, prevalencia; S₁, sensibilidad del test 1; S₂, sensibilidad del test 2; C₁, especificidad del test 1; C₂, especificidad del test 2; covS₁₂, covarianza entre los dos tests para los individuos con la enfermedad; covC₁₂, covarianza entre los dos tests para los individuos sin la enfermedad. ICr, 95 % Intervalo de credibilidad bayesiano; VPC_{SUB}, componente de partición de la varianza del efecto aleatorio individuo; VPC_{SUB×TST}, componente de partición de la varianza del efecto aleatorio de interacción individuo-test. Modelo: SUB×TST, efectos aleatorios individuo y interacción individuo-test; SUB, efecto aleatorio individuo; DEP, dependiente y IND, Independiente.

Tabla 3.3: Resultados de las estimaciones de los parámetros para los datos Hookworm.

3.2.2 Simulación

En relación con la prevalencia, los resultados de los cuatro modelos en los escenarios I y II fueron muy similares y con muy poco sesgo. En el Escenario III, el modelo SUB×TST mostró los mejores resultados mientras que los modelos SUB y DEP obtuvieron los peores resultados. El modelo SUB sobrestimó fuertemente la prevalencia, calculado valores 24,0 % superiores al valor real y el modelo DEP alrededor de 10,6 %. En el escenario IV, en la que la prevalencia real (0,75) fue mayor que en los otros escenarios, el SUB×TST presentó el sesgo relativo más pequeño. El peor modelo fue el IND, que subestima la prevalencia por un valor de 9,46 % (Tabla 3.4 y Figura 3.1.a).

Con respecto a la sensibilidad de los tests, el SUB×TST presentó el sesgo más pequeño en casi todos los escenarios. El sesgo aumentó gradualmente en los modelos SUB, IND y DEP en línea con el aumento de la variabilidad del efecto SUB×TST. De estos tres modelos, la tendencia de aumento de sesgo fue más pronunciado con el modelo IND. Además, el modelo IND siempre sobrestima fuertemente la sensibilidad hasta en un 15 % (Tabla 3.4, Figura 3.1.b y Figura 3.1.c).

En relación a las especificidades en los escenarios I y II, los sesgos de los cuatro modelos fueron pequeños y similares. En el escenario III y IV, el modelo SUB×TST mostró bajo sesgo. Por consiguiente, las diferencias más pronunciadas se observaron en el escenario IV, en el que el modelo SUB×TST presentó un sesgo menor que 1 %, mientras en los otros modelos fueron como mínimo 14 % mayor que el valor real del parámetro (Tabla 3.4, Figura 3.1.d y Figura 3.1.e).

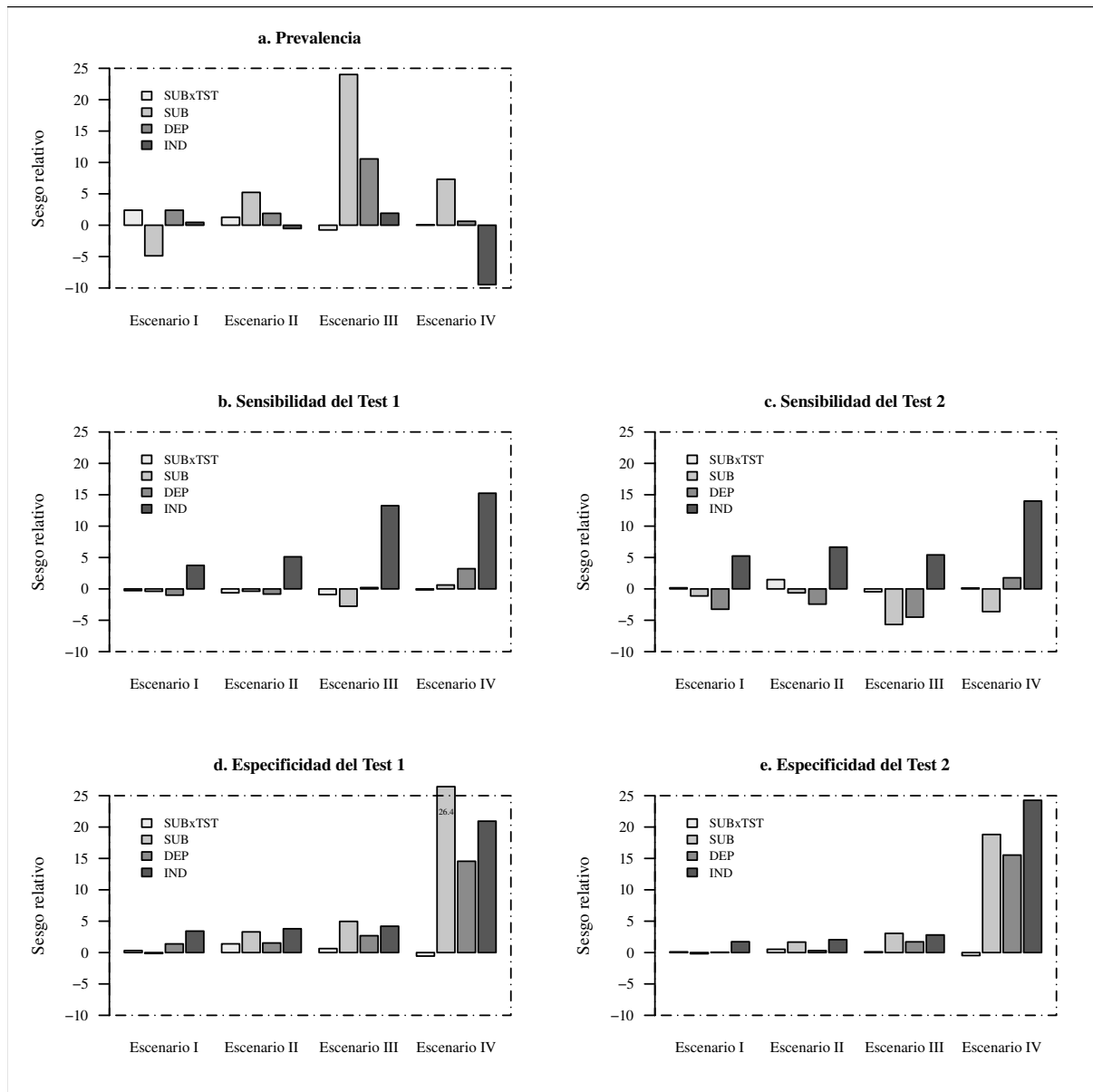


Figura 3.1: *Seago relativo calculado en los escenarios I, II, III y IV de acuerdo con los cuatro modelos; SUB×TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio individuo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente.*

En los escenarios I y II todos los parámetros (prevalencia, sensibilidad y especificidad) mostraron un MSE similar. En los escenarios III y IV, la prevalencia del modelo SUB×TST tenía un MSE más bajo que el del modelo SUB. Sin embargo, el modelo DEP presentó un MSE ligeramente inferior al SUB×TST debido a menor variabilidad de las estimaciones (Tabla 3.4 y Figura 3.2).

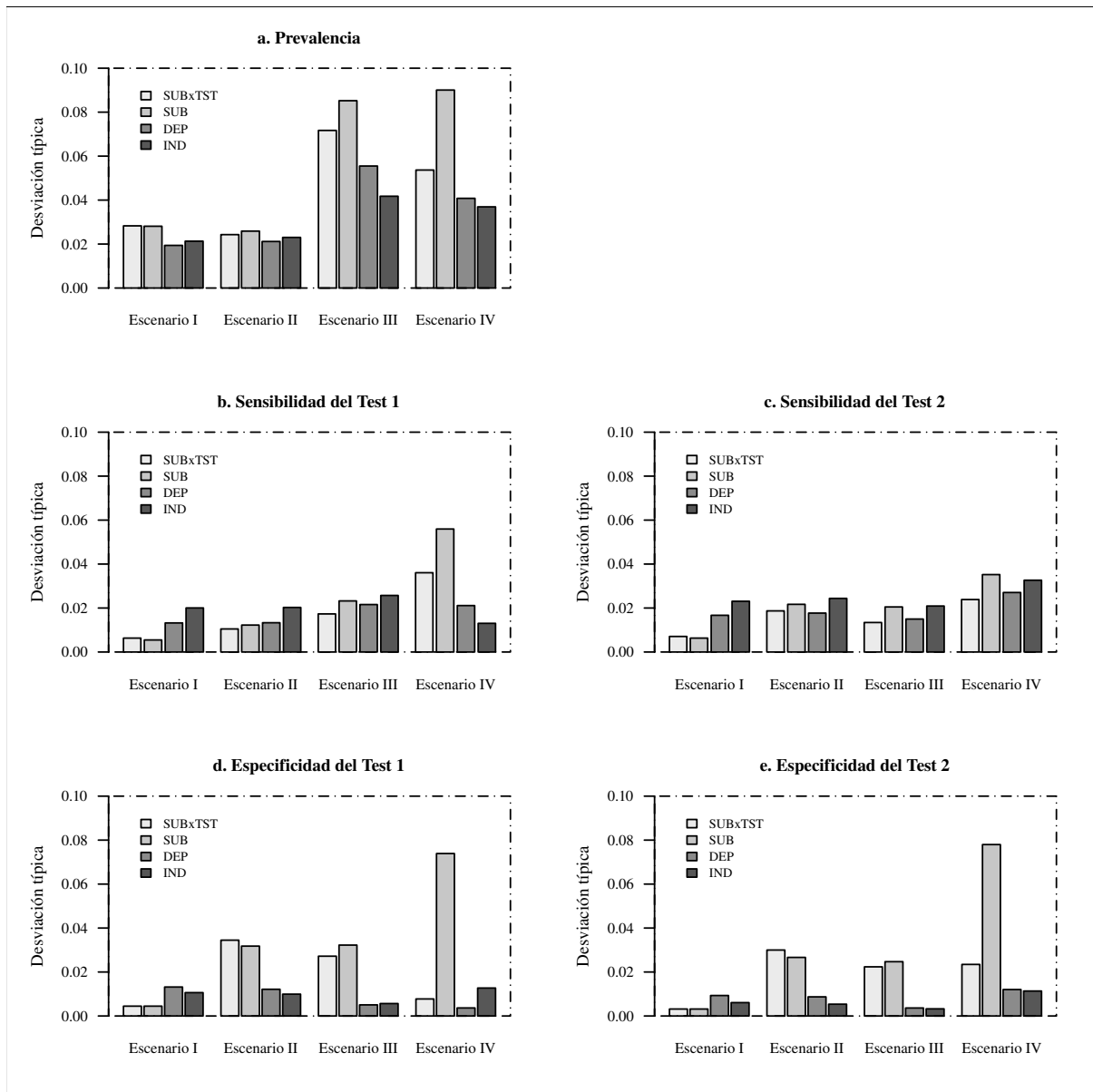


Figura 3.2: Desviación típica calculada en los escenarios I, II, III y IV de acuerdo con los cuatro modelos; SUB×TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio individuo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente.

Sc	Par	Real	Sesgo Relativo (x 100)						Desviación Típica						Error Cuadrático Medio (x 100)					
			SUB×TST	SUB	DEP	IND	SUB×TST	SUB	DEP	IND	SUB×TST	SUB	DEP	IND						
I	S ₁	0,52	-0,28	-0,38	-1,00	3,75	0,006	0,005	0,013	0,020	0,004	0,003	0,020	0,079						
	S ₂	0,35	0,17	-1,13	-3,24	5,24	0,007	0,006	0,017	0,023	0,005	0,006	0,041	0,086						
	C ₁	0,91	0,32	-0,14	1,38	3,42	0,004	0,004	0,013	0,011	0,003	0,002	0,033	0,108						
	C ₂	0,95	0,11	-0,19	0,06	1,73	0,003	0,003	0,009	0,006	0,001	0,001	0,009	0,031						
	Prev	0,30	2,40	-4,88	2,39	0,45	0,028	0,028	0,019	0,021	0,085	0,100	0,043	0,046						
II	S ₁	0,52	-0,62	-0,37	-0,83	5,11	0,010	0,012	0,013	0,020	0,012	0,016	0,020	0,113						
	S ₂	0,35	1,48	-0,63	-2,43	6,66	0,019	0,022	0,018	0,024	0,038	0,048	0,038	0,113						
	C ₁	0,91	1,39	3,30	1,53	3,80	0,034	0,032	0,012	0,010	0,135	0,191	0,034	0,129						
	C ₂	0,95	0,53	1,67	0,31	2,05	0,030	0,027	0,009	0,005	0,093	0,096	0,008	0,041						
	Prev	0,30	1,27	5,23	1,89	-0,53	0,024	0,026	0,021	0,023	0,060	0,091	0,048	0,053						
III	S ₁	0,52	-0,90	-2,77	0,21	13,26	0,017	0,023	0,022	0,026	0,032	0,075	0,047	0,544						
	S ₂	0,37	-0,46	-5,68	-4,50	5,42	0,013	0,020	0,015	0,021	0,018	0,086	0,050	0,084						
	C ₁	0,89	0,64	4,97	2,69	4,20	0,027	0,032	0,005	0,006	0,077	0,300	0,060	0,143						
	C ₂	0,93	0,12	3,06	1,71	2,81	0,022	0,025	0,004	0,003	0,050	0,143	0,027	0,070						
	Prev	0,25	-0,76	24,01	10,56	1,91	0,072	0,085	0,055	0,042	0,514	1,087	0,378	0,177						
IV	S ₁	0,80	-0,16	0,62	3,22	15,25	0,036	0,056	0,021	0,013	0,130	0,316	0,111	1,505						
	S ₂	0,35	0,14	-3,62	1,77	14,00	0,024	0,035	0,027	0,033	0,057	0,141	0,077	0,347						
	C ₁	0,70	-0,56	26,44	14,55	20,94	0,008	0,074	0,004	0,013	0,008	3,972	1,038	2,164						
	C ₂	0,75	-0,48	18,82	15,53	24,26	0,023	0,078	0,012	0,011	0,056	2,600	1,372	3,325						
	Prev	0,75	0,06	7,32	0,63	-9,46	0,054	0,090	0,041	0,037	0,288	1,112	0,169	0,640						

Sc, escenario; Par, parámetro; Prev, prevalencia; S₁, sensibilidad del test 1; S₂, sensibilidad del test 2; C₁, especificidad del test 1; C₂, especificidad del test 2. SUB×TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio individuo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente.

Tabla 3.4: Sesgo relativo, desviación típica y error cuadrático medio de las estimaciones.

En relación a la sensibilidad, el MSE fue muy similar para todos los modelos excepto en el modelo IND, que tuvo los mayores MSE. Con respecto a las especificidades, en el escenario III el MSE de todos los modelos fueron similares y en el escenario IV, los modelos SUB, DEP e IND presentaron un MSE más alto que el modelo SUB \times TST (Tabla 3.4 y Figura 3.3).

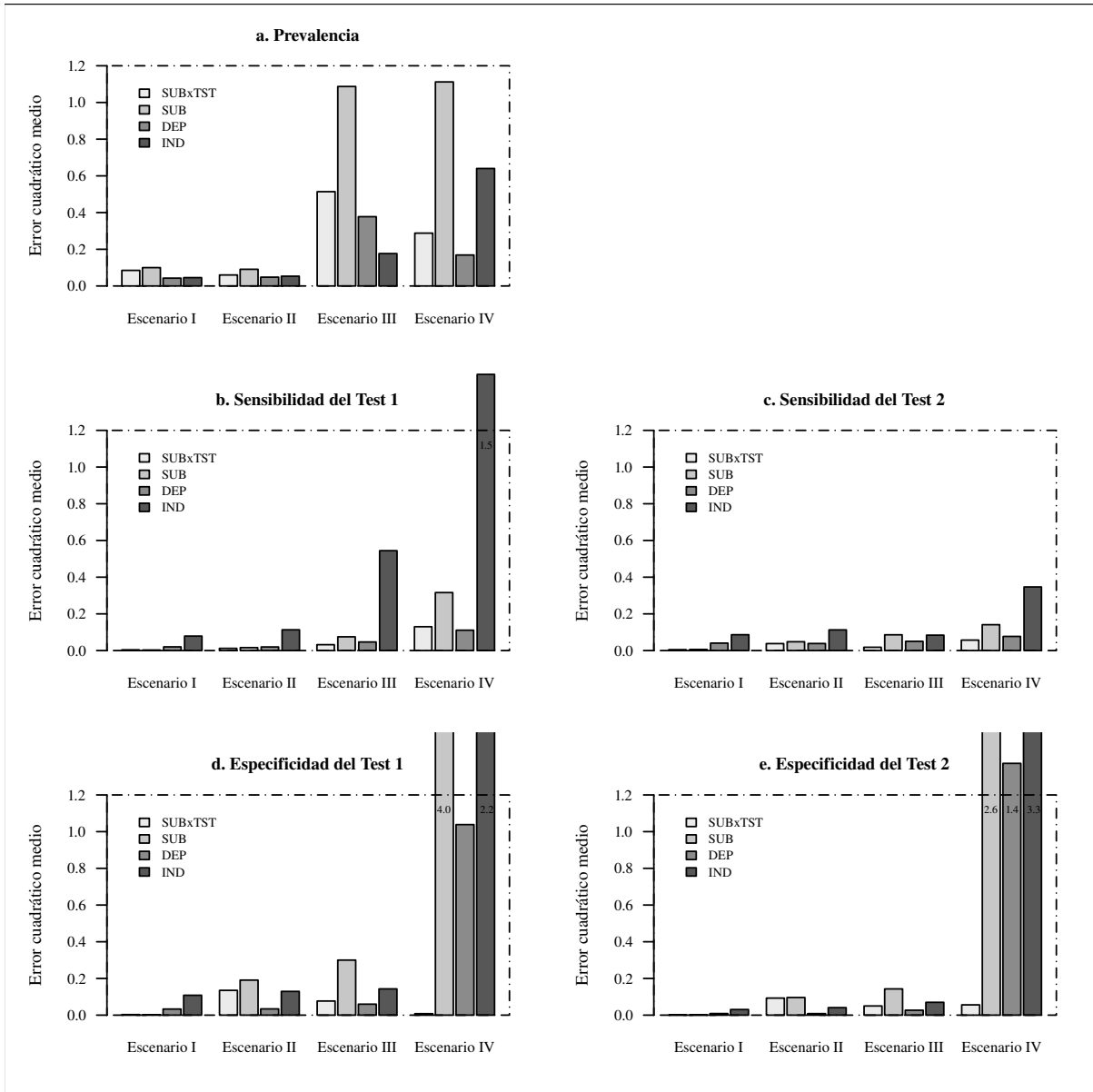


Figura 3.3: Error cuadrático medio calculado en los escenarios I, II, III y IV de acuerdo con los cuatro modelos; SUB \times TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio individuo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente.

En general, la tasa media de cobertura (tabla 3.5) fue muy alta (cerca del 100 %). Esto fue porque los intervalos fueron largos debido a la aplicación de sólo dos tests diagnósticos y no por el comportamiento de los modelos. Se observaron excepciones a esto en los escenarios III y IV para los modelos SUB e IND. Específicamente, el modelo SUB en ambos escenarios, sólo la prevalencia que mostró una cobertura media inferior al valor nominal (de un 90 %). En el escenario III, las especificidades del modelo IND mostraron cobertura media baja, para la especificidad del test 1 60 % y 50 % para la especificidad del test 2. En el escenario IV todos los parámetros del modelo IND presentan una cobertura inferior al 75 %, siendo 74,9 % para la prevalencia, 4,4 % y 74,9 % para la sensibilidad del test 1 y 2 respectivamente, y 28,2 % y 0,0 % para la especificidad de los tests 1 y 2, respectivamente. La figura 3.4 presenta los box-plots de las estimaciones simuladas de todos los parámetros de cada modelo en los cuatro escenarios.

Tasa de cobertura - Escenario I					Tasa de cobertura - Escenario II				
Parámetro	SUB×TST	SUB	DEP	IND	Parámetro	SUB×TST	SUB	DEP	IND
<i>Prev</i>	98,6 %	98,0 %	100,0 %	100,0 %	<i>Prev</i>	100,0 %	99,6 %	100,0 %	100,0 %
<i>S</i> ₁	100,0 %	100,0 %	100,0 %	100,0 %	<i>S</i> ₁	100,0 %	100,0 %	100,0 %	99,6 %
<i>S</i> ₂	100,0 %	100,0 %	100,0 %	99,6 %	<i>S</i> ₂	100,0 %	100,0 %	100,0 %	99,0 %
<i>C</i> ₁	100,0 %	100,0 %	100,0 %	93,8 %	<i>C</i> ₁	98,6 %	97,4 %	100,0 %	90,2 %
<i>C</i> ₂	100,0 %	100,0 %	100,0 %	99,6 %	<i>C</i> ₂	98,2 %	99,0 %	100,0 %	98,4 %
Tasa de cobertura - Escenario III					Tasa de cobertura - Escenario IV				
Parámetro	SUB×TST	SUB	DEP	IND	Parámetro	SUB×TST	SUB	DEP	IND
<i>Prev</i>	97,4 %	90,6 %	97,4 %	97,4 %	<i>Prev</i>	96,6 %	89,2 %	100,0 %	75,0 %
<i>S</i> ₁	100,0 %	100,0 %	100,0 %	97,2 %	<i>S</i> ₁	99,6 %	99,0 %	100,0 %	4,4 %
<i>S</i> ₂	100,0 %	100,0 %	100,0 %	99,6 %	<i>S</i> ₂	100,0 %	100,0 %	98,8 %	75,0 %
<i>C</i> ₁	99,8 %	92,8 %	99,0 %	60,8 %	<i>C</i> ₁	100,0 %	99,6 %	100,0 %	28,2 %
<i>C</i> ₂	99,8 %	93,6 %	99,6 %	49,4 %	<i>C</i> ₂	100,0 %	100,0 %	72,2 %	0,0 %

Tabla 3.5: Tasa de cobertura para los cuatro escenarios

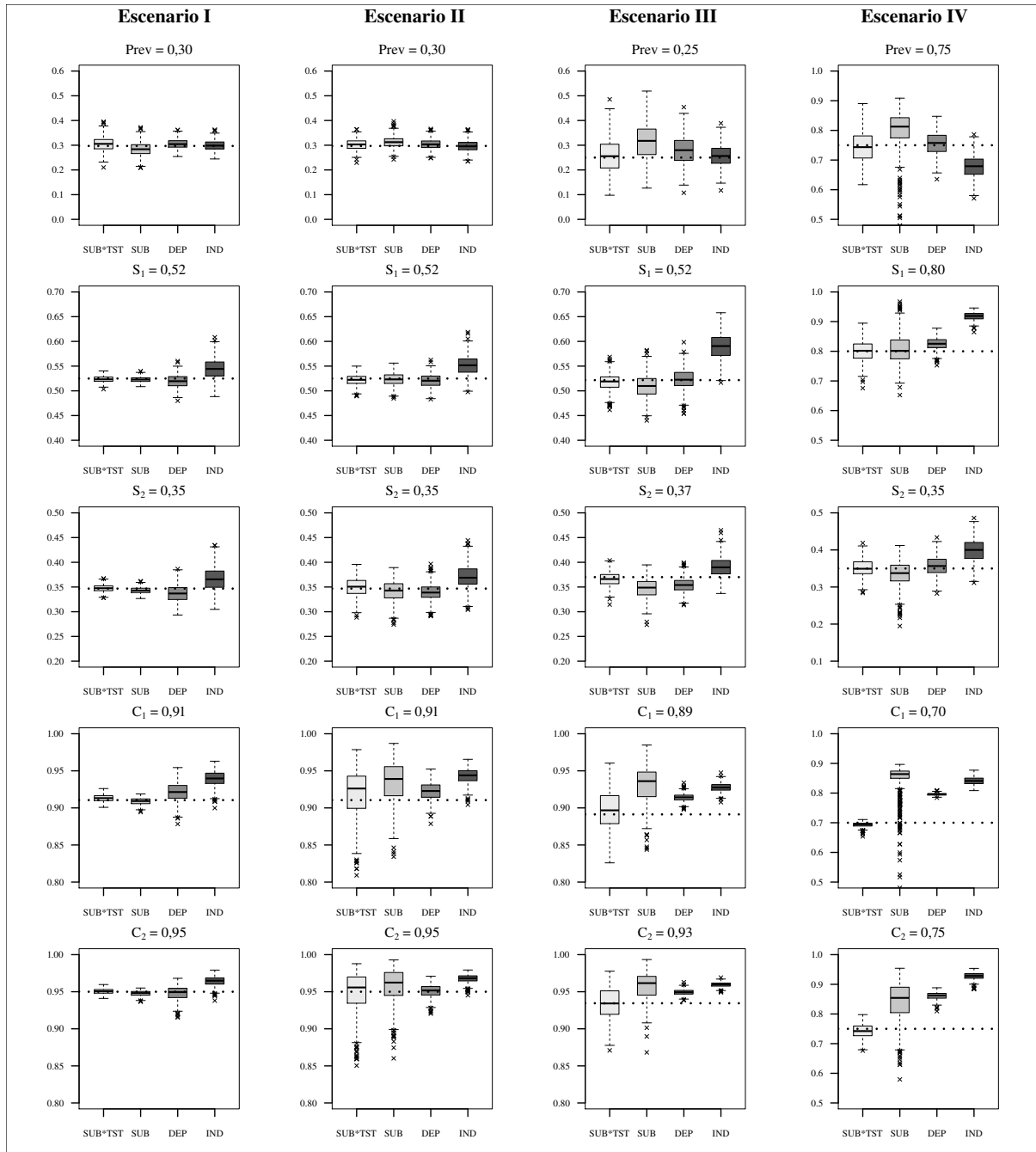


Figura 3.4: Resultados de la simulación para los escenarios I, II, III and IV de acuerdo con los cuatro modelos; SUB×TST, se consideran dos efectos aleatorios: individuo y interacción individuo-test; SUB, tiene en cuenta sólo el efecto aleatorio individuo; DEP, modelo dependiente condicional al verdadero estado del paciente (efecto fijo); IND, modelo independiente condicional al verdadero estado del paciente. Prev, prevalencia; S₁, sensibilidad del test 1; S₂, sensibilidad del test 2; C₁, especificidad del test 1; C₂, especificidad del test 2. El verdadero valor del parámetro está en la línea discontinua.

3.3 Discusión

En este estudio se ha propuesto un modelo de clases latentes para el diagnóstico clínico en ausencia de un *gold standard* que tiene en cuenta la dependencia entre las medidas replicadas de los tests. Esta dependencia se modeló mediante la adición de un efecto aleatorio individuo-test. A través de un estudio de simulación, se comparó el funcionamiento del modelo propuesto con los modelos de dependencia condicional descritos en [Dendukuri & Joseph \[2001\]](#) (modelos de efectos fijos y aleatorios) y con el modelo independiente de [Joseph et-al. \[1995\]](#).

El estudio de simulación mostró que nuestro modelo presenta una buena exactitud y precisión en los diferentes escenarios y proporciona mejores resultados que el modelo de efectos aleatorios individuo y que los modelos de dependencia e independencia condicional. Por otra parte, el estudio de simulación realizado reveló que si la variabilidad de las réplicas es mayor que la variabilidad del individuo, nuestro modelo estima correctamente los parámetros. Incluso cuando la variabilidad de las réplicas es baja no hubo una marcada diferencia entre el comportamiento de nuestro modelo y los otros. Desde este punto de vista, el modelo propuesto podría aplicarse en situaciones en las que son completamente desconocidas las fuentes de variabilidad.

Otra ventaja de la extensión propuesta incluyendo efectos aleatorios es que este modelo permite modelar diferentes estructuras de correlación de una manera simple. Por ejemplo, cuando se replica el test para cada individuo en el tiempo (varias puntos evaluados en el tiempo) [[Espeland et-al., 1988](#), [Hui & Zhou, 1998](#)], se puede considerar un efecto aleatorio con una estructura de correlación temporal. Es importante definir correctamente la estructura de correlación porque las estimaciones de los parámetros de diagnóstico son sesgadas bajo una estructura de dependencia mal especificada [[Albert, 2009](#), [Zhang et-al., 2012](#)]. En presencia de covariables el modelo de efectos aleatorios individuo-test se puede ampliar para incluir esta información como efectos fijos en el predictor lineal como ha señalado [Dendukuri et-al. \[2009\]](#) usando la función de enlace probit o logit [[Martinez et-al., 2008](#)].

Cuando en un estudio no se realizan replicas, el comportamiento del modelo dependiente condicional y del modelo con efecto aleatorio individuo es mejor que del modelo independiente. En los escenarios en que las probabilidades eran alrededor de 0,5 y el efecto réplica tenía

un peso mayor que la variabilidad individuo (escenarios III y IV), el modelo independiente presentó el mayor sesgo con marcadas diferencias en comparación con otros modelos. Cuando el efecto réplica fue pequeño, hubo pequeñas diferencias entre el funcionamiento del modelo con efecto aleatorio individuo-test, y de los modelos con efecto aleatorio individuo y dependiente. Sin embargo el modelo independiente otra vez mostró los peores resultados. A pesar de que el modelo independiente conduce a estimaciones sesgadas de la sensibilidad, especificidad y prevalencia cuando el supuesto de independencia es violado, en una revisión sistemática [van Smeden et al. \[2014\]](#) encontró que más de la mitad de los estudios revisados aplicaron esta fuerte asunción y en la mayoría de estos estudios no la explicaron ni discutieron.

El número de parámetros estimados en los estudios de diagnóstico está limitado a la cantidad de tests diagnósticos disponibles. La no identificabilidad ocurre cuando los grados de libertad de los datos son menores que el número de parámetros a estimar. Pero incluso si los grados de libertad son mayores que el número de parámetros no garantiza la identificación del modelo. Para resolver el problema de identificabilidad es necesario el uso de distribuciones prior informativas para al menos algunos de los parámetros [[Dendukuri et al., 2009](#), [Gustafson, 2005a,b](#)]. La información para las distribuciones prior debe estar estrechamente vinculada a las creencias basadas en la opinión científica de expertos, de los meta-análisis o de estudios previos. Los grados de libertad pueden elevarse si el número de poblaciones estudiadas aumenta, reduciendo así la dependencia de la información prior [[Georgiadis et al., 2003](#), [Toft et al., 2005](#)]. El código WinBUGS para el modelo propuesto se muestra en el apéndice C para facilitar el trabajo de otros autores.

En nuestro estudio de simulación, se decidió utilizar distribución informativa para todos los parámetros. Además, la elección de la prior para cada modelo y escenario se hizo con la misma opinión de los expertos para no influir en los resultados de la simulación. Para el SUB×TST, la media de la distribución prior de efectos aleatorios, b_{20} y b_{21} se fijó de acuerdo con las opiniones de los expertos. Sin embargo, surge un problema con la desviación típica de la distribución porque no hay información disponible. Para resolver este problema hemos probado varios valores: 0,1, 0,3, 0,5 y 2, y en cada escenario se eligió el valor que no generaba problemas de convergencia.

Es difícil resumir la información prior a ser utilizada en las estimaciones de los parámetros y su adaptación a la distribución elegida es un ejercicio que debe realizarse con cuidado para evitar problemas de convergencia y equívocos en la inferencia. Un método conveniente es usar la distribución normal truncada en los hiperparámetros en función de los valores posibles de los parámetros. En nuestros modelos, la distribución normal truncada ayudó a evitar simulaciones con errores inesperados o *bugs* o resultados numéricos tendiendo al \pm infinito. Identificar el “mejor” valor de la desviación típica de los hiperparámetros es otro importante camino que minimiza problemas de convergencia para algunos datos. Después de considerar cada uno de los problemas expuestos anteriormente, además de considerar largas cadenas en las simulaciones, en nuestro estudio los modelos ajustados a los datos reales y simulados tuvieron tasa de convergencia de 100 % de los casos.

El escenario IV fue pensado para tener la alta prevalencia (0,75) que puede aparecer en algunos estudios, como por ejemplo, en epidemiología y particularmente en parasitología como se ve en [Joseph et al. \[1995\]](#). Para el valor de sensibilidad de los tests se decidió uno elevado ($S_1=0,80$) y el otro bajo ($S_2=0,35$) con el fin de evaluar cómo los modelos funcionan en presencia de diferencias marcadas de validez. Las especificidades también fueron elegidas para presentar valores lejos de un *gold standard* ($C_1=0,70$ y $C_2=0,75$). Además, la amplitud de las creencias prior de cada parámetro en el escenario IV no puede ser comparable con los otros tres escenarios, por lo que la longitud media posterior de los intervalos de credibilidad de los parámetros fue mayor. En la mayoría de los escenarios las tasas de cobertura de los parámetros estaban por encima de 95 %. Los parámetros sesgados fueron los que no alcanzaron esta cobertura.

En conclusión, el modelo propuesto con dos efectos aleatorios es muy flexible y permite realizar ajustes del modelo para una gran variedad de estudios como en aquellos en los que los tests se repiten en el tiempo o en aquellos en que los individuos presentan agrupaciones.

Objetivo 2. Estudio de la prevalencia y validez de los tests (sin gold-standard) en el diagnóstico de la Tuberculosis en niños

La tuberculosis (TB) es una de las infecciones contagiosas más prevalentes en todo el mundo [Organization et~al., 2015]. La incidencia a nivel mundial está disminuyendo en aproximadamente un 1,5 % anualmente; sin embargo, esta reducción no es suficiente de acuerdo con la meta de eliminación fijada para 2050 propuesta por la Organización Mundial de la Salud (OMS) [Organization et~al., 2014b]. Esto puede explicarse por la falta de diagnóstico de la enfermedad que se ha estimado en alrededor de un tercio de los casos incidentes [Organization et~al., 2015]. Para un mejor control de las enfermedades la OMS ha propuesto la necesidad de no sólo diagnóstico adecuado y tratamiento de los casos de TB, sino también el diagnóstico y tratamiento de la infección tuberculosa latente (LTBI) en grupos de riesgo de desarrollar la enfermedad, tales como niños [Lewinsohn & Lewinsohn, 2008, Marais et~al., 2004, Organization et~al., 2014b, 2015]. La mayoría de los casos de la edad pediátrica se deben al contacto en el hogar con los adultos con tuberculosis y el riesgo de contagio es proporcional a la intensidad de la exposición, así como a la duración de la exposición a un caso índice [Dodd et~al., 2014, Starke, 2003]. Por lo tanto, las medidas de control pediátricas relacionadas con el control de contacto en el que se detectan y tratan de prevenir la enfermedad en las personas infectadas con TB niños y niños

con infección tuberculosa [Organization et~al., 2014a].

El test cutáneo tuberculina (TST) y Interferon- γ *release assays* (IGRA) son técnicas utilizadas para el diagnóstico de la infección tuberculosa y se basan en desafiar al sistema inmunológico con el fin de detectar si reconoce los antígenos de Mtb [Pai et~al., 2004, Starke et~al., 2014]. Sin embargo, relativamente pocos estudios se han llevado a cabo en los niños y los resultados sugieren que el rendimiento de estos ensayos se difieren de los resultados obtenidos en los adultos, especialmente en países con una alta incidencia de TB y TB en los niños pequeños [Sollai et~al., 2014]. La ausencia de un test de referencia para estas técnicas hace difícil su evaluación y comparación [Pai et~al., 2004]. El modelo bayesiano introducido por Joseph et~al. [1995] y principalmente los modelos definidos por Dendukuri & Joseph [2001] permiten estimar la prevalencia y los parámetros de validez de varios tests de diagnóstico sin gold-standard.

El objetivo del presente estudio fue estimar la prevalencia de la infección por Mtb y determinar la validez y la seguridad del TST y *QuantIFERON-TB Gold In-Tube* (QFT) tests en los niños pequeños, teniendo en cuenta la intensidad de la exposición de los niños a un caso índice.

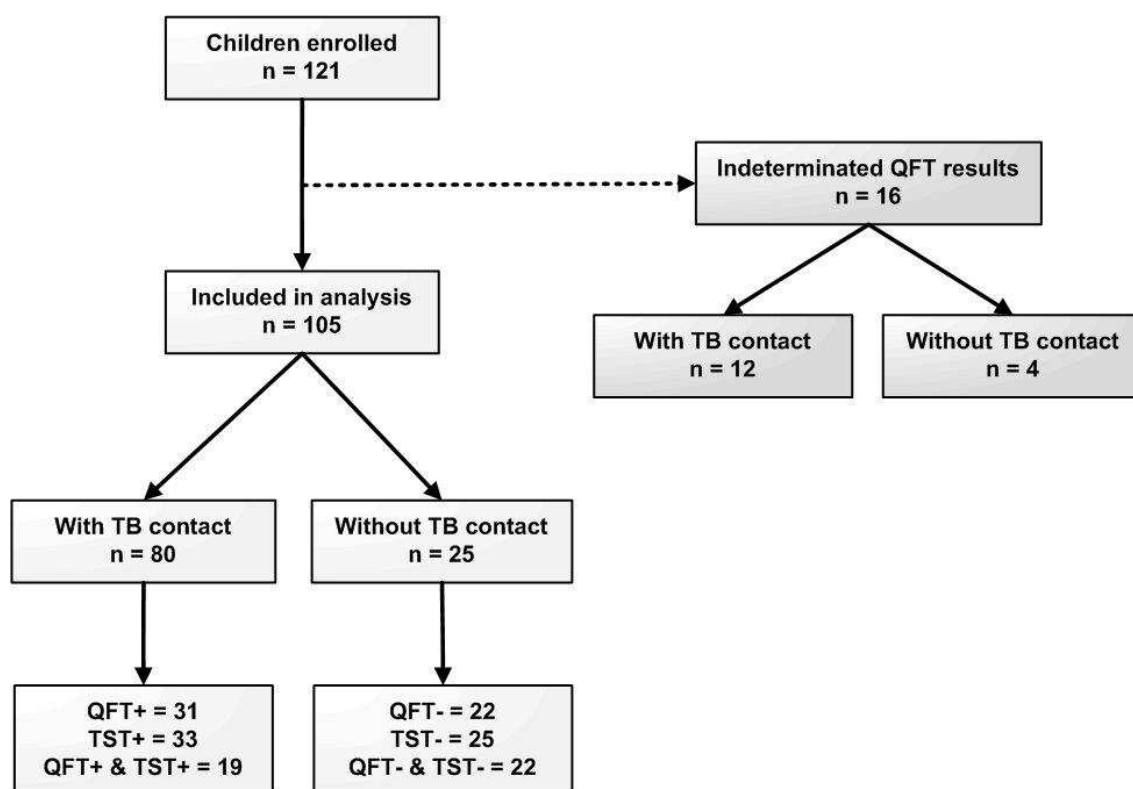
4.1 Material y Métodos

Población y diseño del estudio

Se realizó un estudio transversal para calcular la prevalencia de la infección latente y determinar la exactitud y la validez del test Tuberculin Skin Test (TST) y *Quantiferon-Gold in tube* (QFT). El estudio incluye una muestra de niños menores de 6 años sin ningún contacto reciente conocido con TB y otro grupo de niños con contacto a un caso índice en los 12 últimos meses. El reclutamiento de los datos se realizó en la *Policlínica Cardoso Fontes* (centro de referencia de TB) y en *Fundação de Medicina Tropical Dr. Heitor Vieira Dorado* (FMT-HVD), Manaus, Amazonas, Brazil entre marzo del 2009 a febrero del 2010.

Los adultos (mayores de 12 años) diagnosticados con TB en ambos centros fueron entre-

vistados sobre contacto con niños menores de 6 años. Aquellos que respondieron afirmativamente fueron invitados a llevar a los niños al centro para ser evaluados. Todos los casos índices adultos tenían esputo y cultivo positivo. Los individuos que recibieron tratamiento o profilaxis de TB fueron excluidos (Fig 4.1).



TB: tuberculosis, QFT: QuantiFERON-TB Gold In-Tube, TST: Tuberculin Skin Test.

Figura 4.1: Diagrama de flujo de la inclusión de los niños en el estudio.

Aspectos éticos

El estudio fue aprobado por el Comité Ético de investigación de FMT-HVD, el 26 de octubre de 2007 (Protocolo 2865-07). Todos los representantes legales de los participantes proporcionaron consentimiento informado por escrito para su inclusión en el estudio.

Recolección de datos

Fueron recogidos los datos demográficos, así como la historia epidemiológica de la exposición a un caso índice de TB y la historia clínica y examen médico. Se realizó una radiografía

de tórax, el TST y análisis de sangre (antes del TST). El test para detectar el virus de la inmunodeficiencia humana (VIH) no era obligatorio, pero fue recomendado a todos los participantes. El score de contact *Mycobacterium tuberculosis* (MTC-score) fue utilizado para evaluar la intensidad de la exposición. El MTC-score tiene un rango de 0 a 15 y esta basado en la suposición de que el gradiente de la exposición Mtb es una función compuesta de la infectividad del caso índice (0 – 4), la duración de horas de exposición por día (0 – 4), la relación con el caso índice (0 – 4) y el tipo de exposición (0 – 3) [Hesseling et~al., 2009]. El tiempo (meses) desde el inicio de los síntomas hasta el momento del inicio del tratamiento del caso índice se calculó para medir el tiempo total de exposición (contagio) del niño al caso índice.

Pruebas diagnósticas

El TST se realizó con una inyección intradérmica de 2 unidades de tuberculina (TU) de PPD RT23 (*Statens Serum Institut*, Copenhague, Dinamarca) y fue leído a partir de las 72 horas. Una fuerte reacción del TST (TST+) se definió cuando la induración era $\geq 10mm$ [Organization et~al., 2014a]. El test QFT (Cellestis, Carnegie, Australia) se llevó a cabo y se interpretó de acuerdo con las instrucciones del fabricante por los técnicos de laboratorio experimentados que no tenían conocimiento de los datos de los individuos del estudio.

Análisis estadístico

Las variables categóricas se describieron mediante frecuencias absolutas y porcentajes y las cuantitativas utilizando los percentiles 25, 50 y 75. Se utilizó la prueba de chi-cuadrado para estudiar la asociación entre los resultados de los dos tests y las variables género y grupos de edad (menor o igual a 24 meses y mayores de 24) y la prueba U de Mann Whitney se utilizó para las variables tiempo de exposición y MTC-score. La concordancia entre el QFT y e TST se midió con el índice kappa.

El análisis estadístico se dividió en dos fases. En la primera fase se estimó la validez de los dos tests y se determinó la prevalencia de la infección en función de si los niños habían estado en contacto con un caso índice o no. Para estimar estos parámetros se utilizó la propuesta de

Dendukuri & Joseph [2001], pero se estratificó el modelo para estimar la prevalencia de niños con y sin contacto con un caso índice. La dependencia condicional entre tests fue estimada usando la covarianza entre tests en cada clase de la enfermedad. A continuación, se asumió que las combinaciones de los dos resultados de los tests tienen una distribución multinomial con probabilidades iguales a:

$$P(QFT = +, TST = +) = P_i[S_1S_2 + covS_{12}] + (1 - P_i)[(1 - C_1)(1 - C_2) + covC_{12}]$$

$$P(QFT = +, TST = -) = P_i[S_1(1 - S_2) - covS_{12}] + (1 - P_i)[(1 - C_1)C_2 - covC_{12}]$$

$$P(QFT = -, TST = +) = P_i[(1 - S_1)S_2 - covS_{12}] + (1 - P_i)[C_1(1 - C_2) - covC_{12}]$$

$$P(QFT = -, TST = -) = P_i[(1 - S_1)(1 - S_2) + covS_{12}] + (1 - P_i)[C_1C_2 + covC_{12}]$$

Donde P_i es la prevalencia de LTBI para el i th estrato (1: niños con contacto con un caso índice, y 2: sin contacto con un caso índice), S_1 y S_2 son la sensibilidad de los tests QFT y TST, respectivamente, y C_1 y C_2 son las especificidades. Los parámetros $covS_{12}$ y $covC_{12}$ representan la covarianza entre los dos tests para los individuos con y sin la infección latente, respectivamente.

Para ambos prevalencias se asumió la distribución prior no informativa Beta(1,1). Basado en la literatura de LTBI, razonablemente se consideró que la sensibilidad y la especificidad de los dos tests se encuentran dentro del rango 50% – 100%. Por consiguiente, la distribución prior informativa para estos parámetros fue la Beta(8,25, 2,75). Se asumió que los parámetros de covarianza tienen distribución prior Beta(1,1) y sus rangos factibles fueron dados como $0 \leq covS_{12} \leq \min(S_1, S_2) - S_1S_2$ y $0 \leq covC_{12} \leq \min(C_1, C_2) - C_1C_2$. El Apéndice D contiene el modelo bayesiano (código WinBUGS) para estimar los parámetros.

En la segunda fase, de acuerdo con la propuesta de Martinez et al. [2008], el modelo anterior fue expandido y en vez de estudiar la prevalencia en función de si hubo o no la exposición a un caso índice, se estudió en base a la magnitud de la intensidad y del tiempo de exposición a un caso índice. Se denotó el estado real de la infección (latente) del k th individuo como D_k (0: entre los individuos no infectados y 1 entre los individuos infectados), la prevalencia de la LTBI es la probabilidad de estar verdaderamente infectado $P(D_k = 1)$ y ocurrió en función de

covariables:

$$P(D_k = 1) = \Phi(\beta_0 + \beta_1 \times ExpoTime + \beta_2 \times MTC-score)$$

donde Φ representa la función de distribución acumulada de la distribución $\mathcal{N}(0, 1)$.

La dependencia condicional entre los tests se modeló utilizando el modelo de efectos aleatorios definido por [Dendukuri & Joseph \[2001\]](#). En este caso, las sensibilidades y especificidades de los tests son modeladas como una función del específico efecto aleatorio individuo, i_k . Por tanto, la probabilidad de que el k th individuo tenga un resultado positivo en el j th test (1: QFT y 2: TST) fue dada por:

$$P(T_{jk} = + | D_k = 1, I_k = i_k) = \Phi(\beta_{3j} + b_1 i_k),$$

y de que tenga un resultado negativo:

$$P(T_{jk} = - | D_k = 0, I_k = i_k) = \Phi(\beta_{4j} + b_0 i_k),$$

donde i_k es la intensidad del k th individuo y es una variable aleatoria con distribución $\mathcal{N}(0, 1)$.

De esta manera, la sensibilidad media del j th test en todos los individuos fue:

$$S_j = P(T_j = + | D = 1) = \int_{-\infty}^{\infty} P(T_{jk} = + | D_k = 1, I_k = i_k) d\Phi(I_k) = \Phi\left(\frac{\beta_{3j}}{\sqrt{1 + b_1^2}}\right)$$

donde $j = 1, 2$, y la especificidad media del j th test en todos los individuos fue:

$$C_j = P(T_j = - | D = 0) = \int_{-\infty}^{\infty} P(T_{jk} = - | D_k = 0, I_k = i_k) d\Phi(I_k) = \Phi\left(\frac{-\beta_{4j}}{\sqrt{1 + b_0^2}}\right)$$

para $j = 1, 2$.

Utilizando las estimaciones obtenidas en la primera fase para la prevalencia, sensibilidad y especificidad y covariancias entre los tests, los parámetros de las distribuciones prior de las

constantes del modelo fueron:

$$\beta_0 \sim \mathcal{N}(-1,59, 1,8)$$

$$\beta_{31} \sim \mathcal{N}(0,24, 0,32)$$

$$\beta_{32} \sim \mathcal{N}(0,83, 0,49)$$

$$\beta_{41} \sim \mathcal{N}(1,15, 0,32)$$

$$\beta_{42} \sim \mathcal{N}(2,01, 0,48)$$

y para b_1 y b_0 fueron $\mathcal{N}(0,77, 0,5)$ y $\mathcal{N}(0,99, 0,5)$, respectivamente. La distribución prior para los coeficientes del tiempo e intensidad de exposición MTC-score fue $\mathcal{N}(0, 0,1)$ [Gelman & Rubin, 1996]. El Apéndice E contiene el modelo bayesiano (código WinBUGS) para estimar los parámetros.

4.2 Resultados

Se estudiaron un total de 121 niños, 92 (76,0 %) de los cuales habían tenido contacto con un caso índice y 29 (24,0 %) no tuvieron. Dieciséis niños (13,2 %) fueron excluidos del análisis debido a que presentan un resultado indeterminado en el QFT. De los 105 niños analizados, el 41,9 % eran hombres y el 23,8 % eran de menos de 24 meses de edad. Veinticinco niños (23,8 %) habían tenido contacto con un caso índice, y 3 niños presentaron un resultado positivo QFT y ninguno para el TST. De los 80 niños (76,2 %) que tuvieron contacto con un caso índice, la mediana MTC-score fue de 12,0 puntos (IQR = 3,75) y el tiempo de exposición fue de 2 meses (IQR = 4,0). Treinta y un niños (38,8 %) mostraron resultado positivo en el QFT y 33 (41,2 %) con el TST, pero sólo 19 niños (23,9 %) dieron positivos para ambos tests. Ninguno de los resultados positivos TST eran débiles (entre 5-9 mm). Los dos tests mostraron una baja concordancia de 0,364 ($p < 0,001$).

Los resultados de los dos tests no se asociaron ni con el sexo ni la edad de los niños. El tiempo y la intensidad de la exposición de los niños se asociaron con los resultados de QFT y TST (tabla 4.1). El tiempo y la intensidad de la exposición de los niños con un resultado QFT

positivo (mediana del tiempo de exposición = 2; mediana MTC-score = 12,0) fue mayor que la de los niños con una QFT negativo (mediana del tiempo de exposición = 1,0; $p = 0,026$; mediana MTC-score = 10,0, $p = 0,021$). El tiempo y la intensidad de la exposición de los niños con un resultado positivo TST (mediana del tiempo de exposición = 4; mediana MTC-score = 12,0) fue mayor que en aquellos con un resultado negativo de TST (mediana del tiempo de exposición = 1,0, $p < 0,001$, la mediana MTC-score = 9,5, $p < 0,001$).

	QFT			TST		
	+	-	p-valor	+	-	p-valor
Masculino	14 (31,8 %)	30 (68,2 %)	0,917	14 (31,8 %)	30 (68,2 %)	0,942
Femenino	20 (32,8 %)	41 (67,2 %)		19 (31,1 %)	42 (68,9 %)	
Edad <= 24 meses	7 (28,0 %)	18 (72,0 %)	0,592	8 (32,0 %)	17 (68,0 %)	0,944
Edad >24 meses	27 (33,8 %)	53 (66,2 %)		25 (31,2 %)	55 (68,8 %)	
Edad (meses)*	47,76 (22,5)	43,68 (21,5)	0,384	43,91 (22,7)	45,5 (22,4)	0,737
Tiempo de Exposición**	2 (1-5,25)	1 (0-3)	0,026	4 (1-10)	1 (0-3)	<0,001
MTC-score**	12 (8,75-14,0)	10 (0-12)	0,021	12 (10-14)	9,5 (0-12)	<0,001

* media (desviación típica). Student t

** mediana (cuartil 25-cuartil 75). Mann-Whitney U

TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, MTC-SCORE: Mycobacterium tuberculosis contact score.

Tabla 4.1: Análisis descriptivo y comparaciones entre los resultados de los tests QFT y TST.

La prevalencia de la infección en el grupo de niños que no tuvieron contacto con un caso índice fue de 0,04 (95 % ICr [0,00, 0,20]), siendo de 0,50 en niños con contacto con un caso índice (95 % ICr [0,28, 0,81]). La sensibilidad del QFT fue de 0,58 (95 % ICr [0,41, 0,78]) y 0,75 (95 % ICr [0,49, 0,94]) para el TST. La especificidad del QFT fue de 0,79 (95 % ICr [0,67, 0,91]) y 0,92 (95 % ICr [0,78, 0,98]) para el TST (tabla 4.2).

	Mediana	95 % ICr
P(LTBI) Desconocida exposición de los niños	0,04	0,00-0,20
P(LTBI) Niños que están expuestos	0,50	0,28-0,81
Sensibilidad del QFT	0,58	0,41-0,78
Especificidad del QFT	0,79	0,67-0,91
Sensibilidad del TST	0,75	0,49-0,94
Especificidad del TST	0,92	0,78-0,98

ICr: Intervalo de credibilidad, TST: *Tuberculin Skin Test*, QFT: *QuantiFERON-TB Gold In-Tube*, MTC-SCORE: *Mycobacterium Tuberculosis* escore de contacto, LTBI: Infección Latent por Tuberculosis.

Tabla 4.2: Resultados de la prevalencia de LTBI de acuerdo con la exposición a un caso índice de TB y la sensibilidad y la especificidad de los tests QFT y TST (primera fase).

Los valores predictivos positivos de ambos tests fueron muy bajos en el grupo de niños que no tuvieron contacto con un caso índice y por tanto, la prevalencia de la infección fue muy baja. En el grupo de niños con el contacto con un caso índice, la probabilidad de que un niño diagnosticado como positivo realmente este infectado fue de 0,74 (95 % ICr [0,47, 0,95]) para QFT y 0,91 (95 % CrI [0,61, 0,99]) para TST. Por otra parte, la probabilidad de que un niño diagnosticado como negativo realmente no este infectado fue de 0,65 (95 % CrI [0,27, 0,88]) para QFT y 0,79 (95 % CrI [0,31, 0,96]) para TST. Al evaluar el valor predictivo positivo con los resultados de ambos tests combinados encontramos que la probabilidad de estar realmente infectado cuando los dos tests son positivos fue de 0,93 (95 % ICr [0,69, 0,99]). Cuando sólo el QFT fue positivo este valor fue de 0,35 (95 % ICr [0,04, 0,90]), siendo 0,90 (95 % ICr [0,30, 0,99]) cuando sólo el TST fue positivo. En relación con los valores predictivos negativos, la probabilidad de no estar infectado cuando ambos tests fueron negativos fue 0,83 (95 % ICr [0,37, 0,97]), y cuando sólo el QFT fue negativo este valor fue de 0,1 (95 % ICr [0,01, 0,70]), siendo 0,65 (95 % CrI [0,10, 0,96]) (tabla 4.3) cuando sólo el TST fue negativo.

Teniendo en cuenta la magnitud de la intensidad (MTC-score) y el tiempo de exposición,

	Desconocida exposición de los niños		Niños que están expuestos	
	Mediana	95 % ICr	Mediana	95 % ICr
P(LTBI QFT+)	0,01	0,00-0,51	0,74	0,47-0,95
P(LTBI TST+)	0,27	0,01-0,79	0,91	0,61-0,99
P(LTBI QFT+ TST+)	0,31	0,01-0,83	0,93	0,69-0,99
P(LTBI QFT+ TST-)	0,02	0,00-0,34	0,35	0,04-0,90
P(LTBI QFT- TST+)	0,23	0,01-0,82	0,90	0,31-0,99

ICr: Intervalo de credibilidad, TST: *Tuberculin Skin Test*, QFT: *QuantiFERON-TB Gold In-Tube*, MTC-SCORE: *Mycobacterium tuberculosis* escore de contacto, LTBI: Infección por tuberculosis latente.

Tabla 4.3: Probabilidad de LTBI de acuerdo con los resultados de los dos tests en el modelo de clases latentes (primera fase).

de acuerdo con el modelo utilizado, la prevalencia aumenta con la intensidad ($\beta = 0,14$, 95 % CrI [0,06, 0,23]) y el tiempo de exposición ($\beta = 0,19$, 95 % ICr [0,05, 0,49]) (tabla 4.4).

	Desconocida exposición de los niños		Niños que están expuestos	
	Mediana	95 % ICr	Mediana	95 % ICr
P(LTBI ^c QFT-)	0,98	0,88-1,00	0,65	0,27-0,88
P(LTBI ^c TST-)	0,99	0,90-1,00	0,79	0,31-0,96
P(LTBI ^c QFT+ TST-)	0,98	0,66-1,00	0,64	0,10-0,96
P(LTBI ^c QFT- TST+)	0,76	0,18-0,99	0,1	0,01-0,69
P(LTBI ^c QFT- TST-)	0,99	0,93-1,00	0,83	0,37-0,97

ICr: Intervalo de credibilidad, TST: *Tuberculin Skin Test*, QFT: *QuantiFERON-TB Gold In-Tube*, MTC-SCORE: *Mycobacterium tuberculosis* escore de contacto, LTBI: Infección por tuberculosis latente.

Tabla 4.4: Probabilidad de no LTBI (LTBI^c) de acuerdo con los resultados de los dos tests en el modelo de clases latentes (primera fase).

En la figura 4.2 muestra los valores de la prevalencia para los tiempos de exposición de 0, 1, 3, 6 y 12 meses y los valores MTC-score de 4, 8, 13 y 15. Los niños con un tiempo de exposición de 12 meses y MTC-score superior a 8 presentan prevalencias mayores al 75 %, mientras que los niños con tiempos de exposición más bajos (menor o igual a 1 mes) y un MTC-score inferior a 5 presentan prevalencias menores del 25 %. En este modelo, la sensibilidad y especificidad de los tests se modificó ligeramente en tomar en cuenta la intensidad y tiempo de exposición. La figura 4.3 A y B muestran los valores predictivos positivos en base a la intensidad y el tiempo de exposición. Se puede ver que con tiempos de exposición elevados

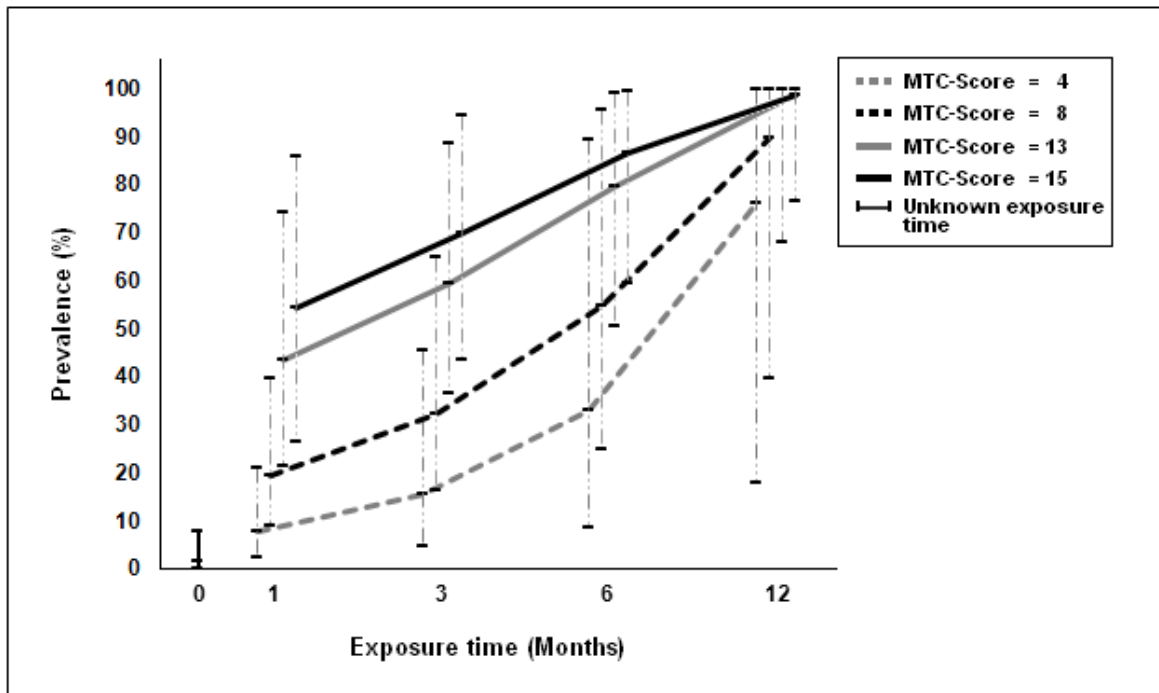


Figura 4.2: Prevalencia en los tiempos de exposición de 0, 1, 3, 6 y 12 meses y MTC-score con intensidades 4, 8, 13 y 15 de exposición.

(mayores que o igual a 6 meses) y exposición mayor que o igual a 13, el valor predictivo fue mayor que 0,70 para QFT y mayor que 0,90 para TST. Los valores predictivos negativos de ambos tests mostraron un comportamiento similar, siendo elevados con un bajo tiempo e intensidad de exposición y bajo cuando el tiempo y la intensidad fueron altos (figura 4.3 C y 4.3 D).

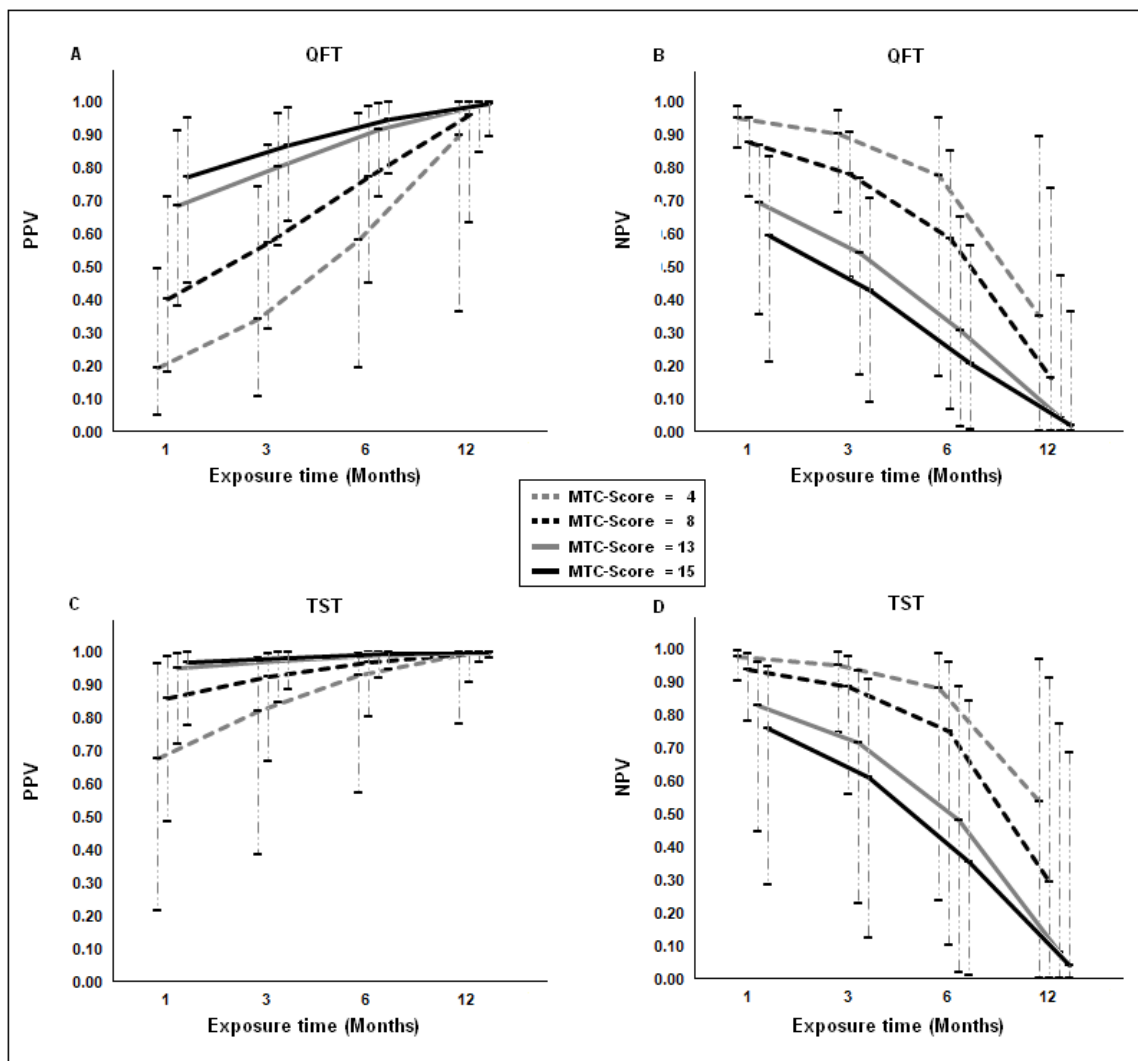


Figura 4.3: Valores predictivos positivos y negativos para QFT en los tiempos de exposición de 0, 1, 3, 6 y 12 meses y MTC-score con intensidades 4, 8, 13 y 15 de exposición.

4.3 Discusión

La estimación de la prevalencia de una enfermedad y los parámetros (sensibilidad, especificidad y valores predictivos) de los tests utilizados para estudiar esta enfermedad en ausencia de un *gold standard* es un importante avance para la evaluación de tests de diagnóstico. Esto es de especial relevancia en el actual contexto epidemiológico y diagnóstico del estudio de la infección por Mtb en la que es imposible determinar si un individuo está realmente infectado o si esta persona presenta respuesta inmune a la infección previa.

La prevalencia de la infección tuberculosa en el grupo expuesto a un caso índice fue de 50,4 % (27,1 % -83,3 %), y esta prevalencia o probabilidad de llegar a ser infectado aumenta con la intensidad (MTC-score) y el tiempo de exposición [Perez-Porcuna et al., 2014]. La prevalencia en el grupo control fue de 3,9 % (0,10 % -20,6 %). La prevalencia obtenida sugiere que los niños pequeños son muy susceptibles a la exposición a un caso índice, con esta prevalencia posiblemente es mayor que en otros grupos de la población [Fox et al., 2013]. Por otro lado, la mayoría de los casos de infección y de TB en los niños pequeños son el resultado de exposición a un conocido caso índice cercano [Dodd et al., 2014]. Estas dos razones refuerzan la importancia crucial de las estrategias de diagnóstico de la infección y la enfermedad en estudio de contactos pediátricos intradomiciliar con el fin de controlar la TB [Cruz et al., 2014].

El conocimiento de la prevalencia de la infección por Mtb en función del tiempo y la intensidad de la exposición permite la estimación de la prevalencia y, por tanto, del riesgo de infección en un niño pequeño antes de obtener el resultado de los tests diagnósticos. De hecho, encontramos que una alta intensidad y tiempo de exposición resulta en muy alta prevalencia (superior a 70 %).

Dado que es difícil saber con precisión en qué momento una persona es contagiada, se estimó el período de contagio como el tiempo desde que el caso índice muestra sintomatología compatible con TB hasta el momento del tratamiento. Del mismo modo, para evaluar la intensidad de la exposición con el caso índice se utilizó el MTC-score que evalúa la capacidad de contagio del caso índice y el grado de relación con el mismo. Los resultados sugieren que el uso del tiempo de exposición y MTC-score son factibles para los modelos de estudio de la

infección tuberculosa en niños.

La concordancia entre el TST y QFT fue menor de lo esperada. Esto ha sido reportado previamente en niños pequeños, sobre todo en regiones con una alta incidencia de TB [Machingaidze et al., 2011] y puede estar relacionado con la coinfección por helmintos, desnutrición, tiempos de exposición cortos, inmunosupresión y los fenómenos de reversión de IGRAs [Pai et al., 2009, Perez-Porcuna et al., 2014, Thomas et al., 2010].

A pesar de la baja concordancia, los tests presentaron sensibilidades y especificidades similares, sin diferencias significativas. Es de señalar que la especificidad del TST no fue inferior al QFT. Este resultado es sorprendente en vista de la generalizada opinión de que la vacuna BCG produce falsos positivos en el TST. Algunos autores han sugerido que este efecto es menor en las regiones tropicales [Hill & Ota, 2010], y que los diferentes tipos de vacunas BCG también pueden tener diferentes efectos en el TST [Ritz et al., 2012].

Los valores predictivos que dependen de la prevalencia y de la exactitud de los tests no mostraron diferencias significativas según el test aplicado, a pesar de que el TST es superior en todos los valores. Del mismo modo, los resultados obtenidos sugieren que ambos tests tienen una muy alta capacidad predictiva cuando la intensidad y tiempo de exposición son altos y son por tanto útiles para el diagnóstico en el estudio de contactos. En el caso de los controles se observaron valores predictivos positivos muy bajos poniendo así de manifiesto su escasa utilidad de estos tests en los *screenings* poblacionales [Organization et al., 2014b].

En la evaluación de los resultados de los valores predictivos en los dos tests combinados se encontró que los valores de los dos tests aislados mejoraron, aunque no significativamente. Según el modelo de variables latentes, cuando los resultados entre los tests son discordantes, el TST presentó mayor probabilidad de clasificar el individuo correctamente. Por tanto, en los niños pequeños resultados discordantes de QFT con los del TST deben interpretarse con cautela, ya que los valores predictivos caen notablemente. Esto puede explicar, en parte, la discordancia observada entre los dos tests en tiempos cortos de exposición, similar a lo que se ha descrito anteriormente en este grupo [Perez-Porcuna et al., 2014].

El número de resultados indeterminados del QFT fueron elevados en nuestro estudio. La proporción de resultados indeterminados varía en los estudios en niños [Bergamini et al., 2009,

Connell et al., 2008, Haustein et al., 2009] y puede estar asociada con problemas técnicos y probablemente también con los factores que alteran la respuesta inmune Th1 [Grare et al., 2010] tales como la edad, la presencia de helmintos y la inmunosupresión [Thomas et al., 2010].

Conclusiones

1. El modelo propuesto (con efecto aleatorio individuo y individuo-test) presenta una buena exactitud y precisión en los diferentes escenarios y proporciona mejores resultados que el modelo de efectos aleatorios individuo y los modelos de (in)dependencia condicional.
2. Cuando la variabilidad de las réplicas es mayor que la variabilidad del individuo, nuestro modelo estima correctamente los parámetros y cuando la variabilidad de las réplicas es baja no hubo una marcada diferencia entre el comportamiento de nuestro modelo y los otros. Por tanto el modelo propuesto podría aplicarse en situaciones en las que son completamente desconocidas las fuentes de variabilidad.
3. El modelo propuesto con dos efectos aleatorios es muy flexible y permite realizar ajustes del modelo para una gran variedad de estudios como en aquellos en los que los tests se repiten en el tiempo o en aquellos en que los individuos presentan agrupaciones.
4. Es uno de los primeros estudios en donde se estima la prevalencia de la infección por M. tuberculosis en los niños y los parámetros de sus principales test de diagnóstico mediante el modelo de clases latentes. Este estudio de la infección por Mtb en los niños pequeños es de gran importancia debido a la alta posibilidad de desarrollar la enfermedad y el reservorio de casos que esto puede representar en el futuro si no reciben tratamiento. Los resultados sugieren que los niños en contacto con un caso índice tienen un alto riesgo de infección en función de la intensidad y duración de la exposición al caso índice y esto debe tenerse en

cuenta en el diseño de estrategias de control de la tuberculosis.

5. El TST y QFT mostraron similares características y comportamiento, pero la concordancia fue menor de lo esperado. El uso combinado de las dos pruebas en nuestro estudio mostró poca mejoría en el diagnóstico de la infección latente. Sabemos que las pruebas de diagnóstico de la infección por Mtb no son perfectas, sin embargo, el modelo de variables latentes permite una mejor evaluación de estas pruebas y la estimación de la prevalencia de la infección por Mtb, permitiendo así el desarrollo de enfoques más adecuados para este problema crucial de control de la tuberculosis. El diagnóstico de la infección por Mtb y la enfermedad es la clave para el control de la TB pediátrica y requiere más y mayores esfuerzos en este sentido.

Apéndice

A Paper I

Research Article


Received 16 December 2014,

Accepted 16 March 2017

Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7339

A Bayesian approach to model the conditional correlation between several diagnostic tests and various replicated subjects measurements

Hélio Doyle Pereira da Silva,^a Carlos Ascaso,^{a,b}
Alessandra Queiroga Gonçalves,^{c,d,e} Patricia Puccinelli Orlandi^f
and Rosa Abellana^{a,g,*†} 

Two key aims of diagnostic research are to accurately and precisely estimate disease prevalence and test sensitivity and specificity. Latent class models have been proposed that consider the correlation between subject measures determined by different tests in order to diagnose diseases for which gold standard tests are not available. In some clinical studies, several measures of the same subject are made with the same test under the same conditions (replicated measurements), and thus, replicated measurements for each subject are not independent. In the present study, we propose an extension of the Bayesian latent class Gaussian random effects model to fit the data with binary outcomes for tests with replicated subject measures. We describe an application using data collected on hookworm infection carried out in the municipality of Presidente Figueiredo, Amazonas State, Brazil. In addition, the performance of the proposed model was compared with that of current models (the subject random effects model and the conditional (in)dependent model) through a simulation study. As expected, the proposed model presented better accuracy and precision in the estimations of prevalence, sensitivity and specificity. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: Bayesian approach; diagnostic test; latent class model; replicated measurement; sensitivity; specificity

1. Introduction

In clinical practice and public health, it is important to estimate disease prevalence and the accuracy (sensitivity and specificity) of the diagnostic tests used in order to provide adequate patient care, monitor the effectiveness of control programmes and obtain a better understanding of disease epidemiology. Nevertheless, with some diseases, diagnoses are obtained from the results of several tests, none of which can be considered a perfect reference standard (gold standard).

Latent class models have been extensively applied to estimate target disease prevalence and the sensitivity and specificity of diagnostic tests in the absence of a gold standard test. Latent class models originated in the field of the social sciences [1, 2] and since then have been applied in several health science disciplines [3–5]. Without a gold standard test, a patient's true disease status cannot be assessed

^aBiostatistics Unit, Public Health Department, University of Barcelona Barcelona, Spain

^bIDIBAPS, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain

^cABS Tortosa-Oest, Institut Català de la Salut (ICS), Tortosa, Tarragona, Spain

^dUnitat de Suport a la Recerca Terres de l'Ebre, Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Tortosa, Tarragona, Spain

^eUniversitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès, Spain

^fInstituto Leônidas e Maria Deane – Fiocruz Amazônia, Manaus, Brazil

^gUnitat de Suport de Barcelona, Institut Universitari d'Investigació en Atenció Primària Jordi Gol, Barcelona, Spain

*Correspondence to: Rosa Abellana, Biostatistics Unit, Public Health Department, University of Barcelona, Casanova 143, 08036, Barcelona, Spain.

†E-mail: rabellana@ub.edu

(latent). With current estimation procedures, data analysis is based on the assumption that the results of the different tests are statistically dependent conditional on the true disease status of the subject. It is important to take this conditional dependence into account because violation of the independent assumption causes biased estimations of prevalence and test accuracy [6]. The magnitude of the bias is related to the values for prevalence, sensitivity and specificity, as well as the size of the covariances ignored in model fitting [6–8].

Recently, some authors reviewed the statistical methods to estimate the diagnostic accuracy of tests and prevalence in the absence of a gold standard test, covering the frequentist framework and the Bayesian approach [9, 10]. This latter, proposed by Joseph *et al.* [11], was developed on the assumption of statistical independence between tests conditioned by the patient's true status. Moreover, under the assumption of independence, several authors have extended this model to include covariates of the subjects under study [12–15]. Qu *et al.* [16] proposed a random effects latent class model to take conditional dependence among multiple diagnostic tests into account using maximum likelihood estimation. The random effects model determines the correlation between the tests by measuring the 'intensity' or 'severity' of illness in each subject. Later, Dendukuri and Joseph [17] proposed a Bayesian approach for this model.

In some clinical studies, several same subject measurements can be performed using the same test under the same conditions (replicated measurements), and thus, these replicated measurements for each subject are not conditional independent. The current statistical methodology does not allow this dependence between the replicated measurements to be taken into account. Our proposal is an extension of the random effects model proposed by Dendukuri and Joseph [17], which allows several sources of variability within subjects. Our model included two independent random effects: the subject random effect, which takes into account the correlation between same subject measurements using several tests, and subject–test interaction random effects, which take into account the correlation between replicated subject measurements using the same test. By including these random effects, we are also assuming that the test properties (sensitivity and specificity) are not constant. The subject random effect induces variations in test performance between subjects but does not change with the test. Thus, test performance will vary between subjects according to the severity or intensity of each subject's disease, owing to the individual severity, which affects the ease of disease detection. However, with the subject–test interaction random effects, the ease of disease detection depends on the characteristics or performance of each test; thus, the detection of each test will vary according to each subject's disease severity.

In the present study, we developed a Bayesian approach to estimate prevalence and test sensitivity and specificity when tests are replicated and the diagnostic result is a binary outcome. The proposed methodology was then applied to evaluate the performance of two tests that diagnose the presence of intestinal parasites, in this case, the hookworm parasite. In addition, a simulation study was conducted to evaluate the performance of the proposed methodology, and the proposed model was compared with a model that assumes independence between tests and replicates and with the two dependence models proposed by Dendukuri and Joseph [17], which only take the conditional dependence between tests into account.

In Section 2, we present the subject–test random effects model, which has two random effects and defines the likelihood function and the mechanism to calculate the parameters of the informative prior distributions. In Section 3, we report application to a hookworm dataset using two tests and three replications with the same subject. The simulation study is presented in Section 4, and the results are discussed in Section 5.

2. Materials and methods

2.1. Subject–test random effects model

The study subjects were measured by the same test repeated several times under the same laboratory conditions. T_{kjr} denotes the results of the k th subject by the j th test in the r th time with $k = 1, \dots, N$, $j = 1, \dots, J$ and $r = 1, \dots, R$. Using a random factor, the observations within the same level of the random effects were correlated. Then two vectors of random effects were considered: the subject random effect, I_k , to model the dependence between subject measures performed with several tests, and the subject–test interaction random effect, I_{kj} , to take into account the dependence between subject replicates. The distribution of the two random variables $I_k = i_k$ and $I_{kj} = i_{kj}$ was standard normal. In the absence of a gold standard test, the subject's true disease status is unknown, and it can be represented by a latent random variable D with Bernoulli distribution; for each subject k , $D = d_k$, where $d_k = 1$ when disease is present

and $d_k = 0$ when disease is absent. Thus, disease prevalence is $P(D = 1)$, represented by θ . For j tests, sensitivity was denoted by S_j and specificity by C_j .

Considering a binary outcome for each subject k (positive $T_{kjr} = 1$ and negative $T_{kjr} = 0$), the probability of a positive outcome in the k th subject by the j th test in the r th time replicate is

$$P(T_{kjr} = 1 | D_k = d_k, I_i = i_k, I_{kj} = i_{kj}) = \Phi(a_{jd_k} + b_{1jd_k}i_k + b_{2rd_k}i_{kj})$$

where Φ represents the cumulative distribution function of $\mathcal{N}(0, 1)$ and a_{jd_k} , b_{1jd_k} and b_{2rd_k} are real, unknown parameters. With this parametrization, the random effects expectations were $E(b_{1jd_k}I_k) = E(b_{2jd_k}I_{kj}) = 0$, and the variances were $Var(b_{1jd_k}I_k) = b_{1jd_k}^2$ and $Var(b_{2rd_k}I_{kj}) = b_{2rd_k}^2$, where $d_k = 1$ for a subject with disease and $d_k = 0$ for a subject without disease. Then b_{1jd_k} is the standard deviation of the random effect associated to test j in disease class d_k , and b_{2rd_k} is the standard deviation of the random effect associated to replicate r in disease class d_k . For simplicity, the same standard deviation for test ($b_{1jd_k} = b_{1d_k}$) and replica ($b_{2rd_k} = b_{2d_k}$) was considered. This implies a constant correlation. Therefore, the probability of a positive outcome in the k th subject by the j th test in the r th time replicate is

$$P(T_{kjr} = 1 | D_k = d_k, I_i = i_k, I_{kj} = i_{kj}) = \Phi(a_{jd_k} + b_{1d_k}i_k + b_{2d_k}i_{kj})$$

The sensitivity of each test was obtained by integrating all the possible values of I_k and I_{kj} , so that

$$S_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(T_{kjr} = 1 | D_k = 1, I_k = i_k, I_{kj} = i_{kj}) d\Phi(I_k) d\Phi(I_{kj}) = \Phi\left(\frac{a_{j1}}{\sqrt{1 + b_{11}^2 + b_{21}^2}}\right) \tag{1}$$

and the specificity of each test was

$$C_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(T_{kjr} = 0 | D_k = 0, I_k = i_k, I_{kj} = i_{kj}) d\Phi(I_k) d\Phi(I_{kj}) = \Phi\left(\frac{-a_{j0}}{\sqrt{1 + b_{10}^2 + b_{20}^2}}\right) \tag{2}$$

2.2. Estimation

The likelihood function, conditioned by disease status d_k and random effects I_k and I_{kj} , was

$$\begin{aligned} L &\propto \prod_{k=1}^N \prod_{r=1}^R P(T_{k1r} = t_{k1r}, \dots, T_{kjr} = t_{kjr} | \varpi, D_k = d_k, I_k = i_k, I_{k1} = i_{k1}, \dots, I_{kj} = i_{kj}) \\ &= \prod_{k=1}^N \left[\theta \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})^{t_{kjr}} [1 - \Phi(a_{j1} + b_{11}i_k + b_{21}i_{kj})]^{(1-t_{kjr})} \right]^{d_k} \\ &\quad \times \left[(1 - \theta) \prod_{j=1}^J \prod_{r=1}^R \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})^{(1-t_{kjr})} [1 - \Phi(a_{j0} + b_{10}i_k + b_{20}i_{kj})]^{t_{kjr}} \right]^{(1-d_k)} \end{aligned} \tag{3}$$

where $\varpi = (\theta, a_{j1}, b_{11}, b_{21}, a_{j0}, b_{10}, b_{20})$ are the vector of parameters to be estimated. The parameters ϖ were estimated using Bayes theorem, combining the likelihood function of the data (3) and the prior information explained in Section 2.3 to obtain the posterior distribution of these parameters.

2.3. Prior distribution

Prevalence was assumed to have a beta(α, β) prior distribution. The prior distribution for the unknown parameters ($a_{j1}, b_{11}, b_{21}, a_{j0}, b_{10}, b_{20}$) concerning the mean test performance (S_j and C_j) was assumed to be normal with a mean of μ and a standard deviation of σ ; thus

$$\begin{aligned} a_{jd_k} &\sim \mathcal{N}(\mu_{a_{jd_k}}, \sigma_{a_{jd_k}}) \\ b_{1d_k} &\sim \mathcal{N}(\mu_{b_{1d_k}}, \sigma_{b_{1d_k}}) \\ b_{2d_k} &\sim \mathcal{N}(\mu_{b_{2d_k}}, \sigma_{b_{2d_k}}) \end{aligned} \tag{4}$$

Bayesian modelling [18] was implemented by calling WinBUGS [19] with the R statistical package [20] using the R2WinBUGS [21] package. The WinBUGS code for the proposed model is shown in Appendix A.

3. Application to hookworm data

The study was carried out in 2009 in the agricultural community of Rio Pardo in the municipality of Presidente Figueiredo, located approximately 160 km to the north of the city of Manaus, Amazonas State (Brazil), and was previously approved by the Research Ethics Committee of the Oswaldo Cruz Foundation (Protocol 384/07 of 20/08/2007) [22]. One fresh stool sample each was obtained from a total of 143 children and adults participating in a study on intestinal pathogen transmission, and two spontaneous sedimentation diagnostic techniques were used to detect intestinal parasite infection: Lutz (T_1) and Paratest (T_2). For analysis, samples were initially processed in a local laboratory unit located in the community. First, each entire specimen was thoroughly homogenized by stirring with a wooden spatula for at least 1 min, and then each sample was separated into three equal replicates of faeces for each test. Several common intestinal parasites were found, but in this application, we considered the most prevalent one, the hookworm parasite. The sensitivity and specificity of the Lutz test were denoted by S_1 and C_1 , and those of the Paratest by S_2 and C_2 , respectively.

Our main objective was to estimate the prevalence of hookworms and to evaluate the sensitivities and specificities of the Lutz and Paratest techniques. These parameters were estimated using the subject–test random effects model (SUB \times TST), the conditional dependent models described by Dendukuri and Joseph [17] (the fixed (DEP) and random effects models (SUB)) and the independent model (IND) developed by Joseph *et al.* [11].

3.1. Prior information

To define the prior distribution parameters, the expert opinions on prevalence, S_1 , S_2 , C_1 and C_2 , were summarized by the intervals delimited by l (lower limit) and u (upper limit).

Prevalence was assumed to have a beta(α , β) prior distribution, and the limits were used to fix the values for hyperparameters α and β . Thus, α and β were calculated by fixing the beta distribution mean to the midpoint of the interval of the expert opinions and fixing the standard deviation of beta distribution to 1/4 of the range of the same interval. The expressions, which depend on the l and u limits, were

$$\alpha = -\frac{1}{2} \frac{(5l^3 + 11l^2u - 8l^2 + 11lu^2 - 16lu + 5u^3 - 8u^2)}{(l - u)^2} \quad (5)$$

$$\beta = \alpha \frac{(2 - l - u)}{l + u}$$

For two tests, the unknown parameters were (a_{11} , a_{21} , b_{11} , b_{21} , a_{10} , a_{20} , b_{10} , b_{20}) and were considered to have a normal prior distribution defined in (4). To simplify the prior information required, we took the same expert opinions into account for the mean of the distribution of the subject–test interaction random effect ($\mu_{b_{21}}$ and $\mu_{b_{20}}$). The information was obtained by modelling the probability of a positive result according to the test as a fixed effect and the subject and subject–test interaction as random effects via generalized linear mixed models. Then the estimation of the standard deviation of the subject–test interaction was used as the information of $\mu_{b_{21}}$ and $\mu_{b_{20}}$.

By fixing the mean of the distribution of the subject–test interaction random effect, the mean of the prior distribution of the subject random effect, we can obtain $\mu_{b_{11}}$ and the mean of the coefficients a_{11} and a_{21} ($\mu_{a_{11}}$ and $\mu_{a_{21}}$) from the equations of covariance between the two tests among subjects with disease:

$$CovS_{12} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\mu_{a_{11}} + \mu_{b_{11}} i_k + \mu_{b_{21}} i_{k1}) \Phi(\mu_{a_{21}} + \mu_{b_{11}} i_k + \mu_{b_{21}} i_{k2}) d\Phi(I_k) d\Phi(I_{kj}) - S_1 S_2 \quad (6)$$

and from the sensitivity equation for tests 1 and 2, deduced from equation (1),

$$\mu_{a_{j1}} = \Phi^{-1}(S_j) \sqrt{1 + \mu_{b_{11}}^2 + \mu_{b_{21}}^2}, j = 1, 2 \quad (7)$$

The values for $\mu_{a_{11}}$, $\mu_{a_{21}}$ and $\mu_{b_{11}}$ can be solved using a bisectional search algorithm [23]. A similar procedure was used to calculate the $\mu_{a_{10}}$, $\mu_{a_{20}}$ and $\mu_{b_{10}}$. In this case, the equation of covariance between the two tests among subjects without disease was

$$CovC_{12} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\mu_{a_{10}} + \mu_{b_{10}} i_k + \mu_{b_{20}} i_{k1}) \Phi(\mu_{a_{20}} + \mu_{b_{10}} i_k + \mu_{b_{20}} i_{k2}) d\Phi(I_k) d\Phi(I_{kj}) - C_1 C_2 \quad (8)$$

Table I. Posterior summary of the parameters for hookworm data.

Par	Model			
	SUB × TST	SUB	DEP	IND
S_1	0.45 (0.29; 0.59)	0.44 (0.27; 0.60)	0.43 (0.28; 0.58)	0.47 (0.33; 0.60)
S_2	0.43 (0.26; 0.58)	0.42 (0.25; 0.58)	0.41 (0.26; 0.58)	0.40 (0.27; 0.56)
C_1	0.91 (0.88; 0.94)	0.92 (0.88; 0.94)	0.92 (0.89; 0.96)	0.94 (0.91; 0.97)
C_2	0.96 (0.93; 0.98)	0.96 (0.93; 0.98)	0.97 (0.95; 0.99)	0.98 (0.96; 0.99)
Prev	0.03 (0.01; 0.08)	0.03 (0.01; 0.08)	0.07 (0.02; 0.15)	0.09 (0.04; 0.17)
$CovS_{12}$	—	—	0.07 (0.02; 0.16)	—
$CovC_{12}$	—	—	0.01 (0.00; 0.02)	—
a_{11}	-0.19 (-0.69; 0.32)	-0.19 (-0.70; 0.31)	—	—
a_{21}	-0.26 (-0.76; 0.26)	-0.26 (-0.77; 0.25)	—	—
a_{10}	2.11 (1.78; 2.49)	2.07 (1.75; 2.43)	—	—
a_{20}	2.63 (2.25; 3.03)	2.59 (2.23; 2.98)	—	—
b_{11}	0.44 (0.00; 2.95)	0.46 (0.00; 2.98)	—	—
b_{21}	0.13 (0.00; 1.45)	—	—	—
b_{10}	1.27 (0.66; 2.18)	1.25 (0.72; 2.03)	—	—
b_{20}	0.05 (0.00; 0.46)	—	—	—
VPC_{SUB} with disease T_1	18.8%	20.7%	—	—
VPC_{SUB} with disease T_2	18.7%	20.6%	—	—
VPC_{SUB} without disease T_1	28.7%	29.6%	—	—
VPC_{SUB} without disease T_2	24.4%	25.4%	—	—
$VPC_{SUB \times TST}$ with disease T_1	1.7%	—	—	—
$VPC_{SUB \times TST}$ with disease T_2	1.7%	—	—	—
$VPC_{SUB \times TST}$ without disease T_1	0.2%	—	—	—
$VPC_{SUB \times TST}$ without disease T_2	0.2%	—	—	—

Par, parameters; S_1 , sensitivity of the test 1; S_2 , sensitivity of the test 2; C_1 , specificity of test 1; C_2 , specificity of the test 2; Prev, prevalence; $CovS_{12}$, covariance between the two tests among the subjects with disease; $CovC_{12}$, covariance between the two tests among the subjects without disease; CI, 95% Bayesian credible interval; VPC_{SUB} , variance partitioning component of the subject random effect; $VPC_{SUB \times TST}$, variance partitioning component of the subject interaction test random effect; SUB × TST, subject–test random effects model; SUB, subject random effects model; DEP, conditional dependent fixed effect model; and IND, conditional independent model.

From the equation of the specificity for tests 1 and 2 in equation (2), it is deduced that

$$\mu_{a_{j0}} = \Phi^{-1}(C_j) \sqrt{1 + \mu_{b_{10}}^2 + \mu_{b_{20}}^2}, j = 1, 2 \quad (9)$$

The standard deviation $\sigma_{a_{j_{dk}}}$ can be approximated using 1/4 of the range of the prior information of S_1, S_2, C_1 and C_2 denoted by l and u . From equations (7) and (9), and replacing the sensitivities and specificities by their limits, the minimum and maximum limits for the σ are obtained. Then the prior information is fixed to

$$\frac{\Phi^{-1}(u) \sqrt{1 + \mu_{b_{1dk}}^2 + \mu_{b_{2dk}}^2} - \Phi^{-1}(l) \sqrt{1 + \mu_{b_{1dk}}^2 + \mu_{b_{2dk}}^2}}{4} \quad (10)$$

For hookworm data, the expert opinions of the two tests ranged from 0.3 to 0.6 for the sensitivities, from 0.9 to 1.0 for the specificities and from 0.0 to 0.5 for infection prevalence. This information was used for all the models. The parameters of beta distribution for prevalence were established using equation (5). In addition, the same prior distribution for prevalence was used in all the models.

Definition of prior distributions according to the model:

- IND model: S_1, S_2, C_1 and C_2 are considered to have a beta prior distribution with parameters determined by equation (5).
- DEP model: The covariance between tests among subjects with and without disease is considered to have a beta distribution with hyperparameters defined as equal to those of Dendukuri and Joseph [17].
- SUB model: The prior distribution of the parameters $a_{11}, a_{21}, a_{10}, a_{20}, b_{11}$ and b_{10} is required. The mean of the prior distribution of these parameters was computed by fixing sensitivity and specificity to the midpoint of the expert opinion interval, and the covariance between tests was set at the midpoint of the posterior credibility interval estimated by the DEP model. Using equations (6), (7), (8) and (9) and setting the parameters depending on the SUB \times TST equal to zero, the means of the prior distribution were obtained. The standard deviation of these prior distributions was fixed using equation (10). However, the standard deviation of b_{11} and b_{10} was fixed at 0.5.
- SUB \times TST model: The same process was performed to obtain the hyperparameters of the prior distributions of $a_{11}, a_{21}, a_{10}, a_{20}, b_{11}$ and b_{10} , but the mean of b_{21} and b_{20} was computed as mentioned earlier, and their standard deviations were fixed at 0.2.

Table I summarizes estimation of the parameters. The prevalence of hookworm infection was 0.03 with low sensitivities of 0.45 for T_1 and 0.43 for T_2 and high specificities of 0.91 for T_1 and 0.96 for T_2 .

The results of the SUB \times TST and SUB models were very similar because the weight of the SUB \times TST random effect was very low, with a variance partitioning component (VPC) for subjects with disease of 1.7% and for subjects without disease of 0.2%. Compared with the random effects models, the DEP and IND models presented few changes in the test's properties (sensitivity and specificity). The main difference was observed with prevalence, which increased more than twofold, and its credibility interval was larger.

4. Simulation

A simulation study was implemented in four scenarios in order to examine the bias or relative bias (bias divided by the real value of the parameter), the mean square error (MSE) and the average coverage of the 95% credibility intervals of the Bayesian estimations. To establish different scenarios, we chose the value of the random effects, b_{1dk} and b_{2dk} , according to the weight of these components in the total variability. The VPCs defined by Goldstein [24] were used. To calculate the VPC for binary response, we computed the simulation method proposed by Browne *et al.* [25]. The strategy of calculating the variance partition coefficient arose from the need to measure the amount of variance explained from two sources: subject and subject–test. Thus, the simulation study was designed with different scenarios of interest, which allowed a processing mechanism to evaluate the performance of each model.

The performance of the propose Bayesian approach for the subject–test random effects model (SUB \times TST) was compared with the results of the subject random effects model (SUB), the conditional dependence fixed effect model (DEP) and the conditional independence model (IND) [11].

Four scenarios were considered, mainly modifying the values of the weights of both random effects. The values fixed to the parameters are shown in Table II.

Table II. True values of the parameters according to the scenarios (I–IV).

Sc	Par	Mean	a_{jd_k}	b_{1d_k}	b_{2d_k}	VPC _{SUB} (%)	VPC _{SUB×TST} (%)
I	S_1	0.52	0.08	0.80	0.20	26.0	1.6
	S_2	0.35	−0.51			25.3	1.6
	C_1	0.91	2.21	1.30	0.10	36.3	0.3
	C_2	0.95	2.70			33.0	0.3
	Prev	0.30	—	—	—	—	—
II	S_1	0.52	0.09	0.80	0.60	21.8	12.4
	S_2	0.35	−0.56			21.1	12.3
	C_1	0.91	2.58	1.30	1.00	23.4	21.0
	C_2	0.95	3.16			20.1	20.7
	Prev	0.30	—	—	—	—	—
III	S_1	0.52	0.09	0.80	1.00	16.5	26.7
	S_2	0.37	−0.54			16.0	26.6
	C_1	0.89	2.58	1.30	1.30	19.8	30.0
	C_2	0.93	3.16			17.0	29.6
	Prev	0.25	—	—	—	—	—
IV	S_1	0.80	1.77	0.67	1.73	5.5	47.6
	S_2	0.35	−0.81			6.5	49.3
	C_1	0.70	1.69	0.86	2.94	4.5	66.4
	C_2	0.75	2.17			4.2	66.0
	Prev	0.75	—	—	—	—	—

Sc, scenario; Par, parameter: S_1 , sensitivity of the test 1; S_2 , sensitivity of the test 2; C_1 , specificity of the test 1; C_2 , specificity of the test 2; Prev, prevalence; VPC_{SUB}, variance partitioning component of SUB model; VPC_{SUB×TST}, variance partitioning component of SUB × TST model.

4.1. Scenario I

For scenario I, the values of all the parameters were set very close to the median posterior results of the hookworm data shown in Table I. The mean sensitivities were slightly modified so that one would be greater and another lower than 0.50. The VPC of the subject–test was close to zero (Table II).

4.2. Scenario II

In scenario II, subject–test variability was increased. This scenario emphasized model performance differences owing to the greater presence of the subject–test effect. The VPC for the subject effect was around 20%, and 15% for the subject–test (Table II).

4.3. Scenario III

In this scenario, subject–test variability was greater than subject variability. The VPC for the subject effect was approximately 15%, and greater than 25% for the subject–test (Table II).

4.4. Scenario IV

Scenario IV represented an exaggeration. Much greater weight was given to subject–test variability (VPC almost 49% for subjects with disease and 66% for subjects without disease) than to subject variability (VPC between 4% and 7%), and thus, prevalence was about 0.75 (Table II).

The prior distribution of the parameters was informative for all models and scenarios, and the same expert opinion was used. This information is shown in table (Table III). These values were obtained in the same way as those of the hookworm data.

The number of subjects was fixed at 60, and 500 datasets were generated under each scenario. To guarantee the convergence of the chains in the simulation study and to save computational time, we previously studied the optimum thin and number of burn-in iterations. Moreover, a condition of convergence was set at which the value of Monte Carlo error for each parameter was lower than 5% of the sample standard deviation [26] of the simulated values. If the convergence criterion was not achieved, more iterations were performed. None of the data were discarded.

Table III. Prior distribution and expert opinion considered for all the scenarios.

Sc	Par	IND or DEP model				SUB model				SUB × TST model									
		Expert opinion		Beta distribution		Normal distribution		Normal distribution		Normal distribution		Normal distribution							
		Mean	Low	Up	α	β	Par	Mean	SD	Par	Mean	SD	Par	Mean	SD				
I	S_1	0.52	0.40	0.62	42.00	40.14	a_{11}	0.09	0.19	b_{11}	0.81	0.10	a_{11}	0.08	0.18	b_{11}	0.80	0.10	
	S_2	0.35	0.24	0.45	29.20	55.25	a_{21}	-0.53	0.17	b_{21}	—	—	a_{21}	-0.51	0.18	b_{21}	0.20	0.10	
	C_1	0.91	0.86	0.99	55.40	4.26	a_{10}	2.22	0.62	b_{10}	1.32	0.10	a_{10}	2.21	0.60	b_{10}	1.30	0.10	
	C_2	0.95	0.90	1.00	69.80	3.88	a_{20}	2.69	0.74	b_{20}	—	—	a_{20}	2.70	0.75	b_{20}	0.10	0.10	
	Prev	0.30	0.19	0.40	22.60	53.42													
	$CovS_{12}$	0.08	0.00	0.17	3.58	39.04													
	$CovC_{12}$	0.02	0.00	0.05	3.88	166.94													
II	S_1	0.52	0.40	0.62	42.00	40.14	a_{11}	0.10	0.22	b_{11}	0.82	0.50	a_{11}	0.09	0.20	b_{11}	0.80	0.50	
	S_2	0.35	0.24	0.45	29.20	55.25	a_{21}	-0.58	0.19	b_{21}	—	—	a_{21}	-0.56	0.20	b_{21}	0.60	0.50	
	C_1	0.91	0.86	0.99	55.40	4.26	a_{10}	2.59	0.71	b_{10}	1.31	0.50	a_{10}	2.58	0.70	b_{10}	1.30	0.50	
	C_2	0.95	0.90	1.00	69.80	3.88	a_{20}	3.15	0.87	b_{20}	—	—	a_{20}	3.16	0.88	b_{20}	1.00	0.50	
	Prev	0.30	0.19	0.40	22.60	53.42													
	$CovS_{12}$	0.08	0.00	0.17	3.58	39.04													
	$CovC_{12}$	0.02	0.00	0.05	3.88	166.94													
III	S_1	0.52	0.37	0.67	22.55	20.82	a_{11}	0.12	0.33	b_{11}	0.82	0.50	a_{11}	0.09	0.31	b_{11}	0.80	0.50	
	S_2	0.37	0.25	0.45	31.50	58.50	a_{21}	-0.56	0.21	b_{21}	—	—	a_{21}	-0.54	0.22	b_{21}	1.00	0.50	
	C_1	0.89	0.87	0.96	139.66	12.97	a_{10}	2.60	0.34	b_{10}	1.31	0.50	a_{10}	2.58	0.33	b_{10}	1.30	0.50	
	C_2	0.93	0.92	0.98	199.61	10.51	a_{20}	3.12	0.32	b_{20}	—	—	a_{20}	3.16	0.34	b_{20}	1.30	0.50	
	Prev	0.25	0.00	0.50	2.75	8.25													
	$CovS_{12}$	0.09	0.00	0.18	3.56	36.49													
	$CovC_{12}$	0.03	0.00	0.06	3.84	119.57													
IV	S_1	0.80	0.70	0.98	22.20	4.23	a_{11}	1.75	0.79	b_{11}	0.65	0.50	a_{11}	1.77	0.81	b_{11}	0.67	0.30	
	S_2	0.35	0.20	0.40	24.90	58.10	a_{21}	-0.80	0.32	b_{21}	—	—	a_{21}	-0.81	0.31	b_{21}	1.73	0.30	
	C_1	0.70	0.65	0.95	21.96	5.49	a_{10}	1.68	1.00	b_{10}	0.85	0.50	a_{10}	1.69	0.75	b_{10}	0.86	0.30	
	C_2	0.75	0.70	0.98	22.20	4.23	a_{20}	2.21	1.26	b_{20}	—	—	a_{20}	2.17	0.75	b_{20}	2.94	0.30	
	Prev	0.75	0.50	1.00	8.25	2.75													
	$CovS_{12}$	0.04	0.00	0.07	3.83	105.46													
	$CovC_{12}$	0.09	0.00	0.18	3.56	37.15													

Sc, scenario; Par, parameter: Prev, prevalence; S_1 , sensitivity of the test 1; S_2 , sensitivity of the test 2; C_1 , specificity of test 1; C_2 , specificity of the test 2; $CovS_{12}$, covariance between the two tests among subjects with disease; $CovC_{12}$, covariance between the two tests among the subjects without disease; SUB × TST, subject–test random effects model; SUB, subject random effects model; DEP, conditional dependent fixed effect model; and IND, conditional independent model.

Table IV. Relative bias, standard deviation and mean square error of the estimations.

Sc	Par	Real	Relative bias (×100)				Standard deviation				Mean square error (×100)			
			SUB × TST	SUB	DEP	IND	SUB × TST	SUB	DEP	IND	SUB × TST	SUB	DEP	IND
I	S ₁	0.52	-0.28	-0.38	-1.00	3.75	0.006	0.005	0.013	0.020	0.004	0.003	0.020	0.079
	S ₂	0.35	0.17	-1.13	-3.24	5.24	0.007	0.006	0.017	0.023	0.005	0.006	0.041	0.086
	C ₁	0.91	0.32	-0.14	1.38	3.42	0.004	0.004	0.013	0.011	0.003	0.002	0.033	0.108
	C ₂	0.95	0.11	-0.19	0.06	1.73	0.003	0.003	0.009	0.006	0.001	0.001	0.009	0.031
II	Prev	0.30	2.40	-4.88	2.39	0.45	0.028	0.028	0.019	0.021	0.085	0.100	0.043	0.046
	S ₁	0.52	-0.62	-0.37	-0.83	5.11	0.010	0.012	0.013	0.020	0.012	0.016	0.020	0.113
	S ₂	0.35	1.48	-0.63	-2.43	6.66	0.019	0.022	0.018	0.024	0.038	0.048	0.038	0.113
	C ₁	0.91	1.39	3.30	1.53	3.80	0.034	0.032	0.012	0.010	0.135	0.191	0.034	0.129
III	Prev	0.95	0.53	1.67	0.31	2.05	0.030	0.027	0.009	0.005	0.093	0.096	0.008	0.041
	S ₁	0.30	1.27	5.23	1.89	-0.53	0.024	0.026	0.021	0.023	0.060	0.091	0.048	0.053
	S ₂	0.52	-0.90	-2.77	0.21	13.26	0.017	0.023	0.022	0.026	0.032	0.075	0.047	0.544
	C ₁	0.37	-0.46	-5.68	-4.50	5.42	0.013	0.020	0.015	0.021	0.018	0.086	0.050	0.084
IV	Prev	0.89	0.64	4.97	2.69	4.20	0.027	0.032	0.005	0.006	0.077	0.300	0.060	0.143
	S ₁	0.93	0.12	3.06	1.71	2.81	0.022	0.025	0.004	0.003	0.050	0.143	0.027	0.070
	S ₂	0.25	-0.76	24.01	10.56	1.91	0.072	0.085	0.055	0.042	0.514	1.087	0.378	0.177
	C ₁	0.80	-0.16	0.62	3.22	15.25	0.036	0.056	0.021	0.013	0.130	0.316	0.111	1.505
V	Prev	0.35	0.14	-3.62	1.77	14.00	0.024	0.035	0.027	0.033	0.057	0.141	0.077	0.347
	S ₁	0.70	-0.56	26.44	14.55	20.94	0.008	0.074	0.004	0.013	0.008	3.972	1.038	2.164
	S ₂	0.75	-0.48	18.82	15.53	24.26	0.023	0.078	0.012	0.011	0.056	2.600	1.372	3.325
	C ₁	0.75	0.06	7.32	0.63	-9.46	0.054	0.090	0.041	0.037	0.288	1.112	0.169	0.640

Sc, scenario; Par, parameter: Prev, prevalence; S₁, sensitivity of the test 1; S₂, sensitivity of the test 2; C₁, specificity of test 1; C₂, specificity of the test 2; SUB × TST, subject–test random effects model; SUB, subject random effects model; DEP, conditional dependent fixed effect model; and IND, conditional independent model.

4.5. Simulation results

In relation to prevalence, the results of the four models in scenarios I and II were very similar with very little bias. In scenario III, the SUB × TST showed the best results, and SUB and DEP were the worst models. The SUB model strongly overestimated prevalence, being around 24% of the real value, and the DEP model overestimated by around 10.6%. In scenario IV, in which the real prevalence (0.75) was greater than that in the other scenarios, the SUB × TST presented the smallest relative bias. The worst model was the IND, which underestimated prevalence by 9.46% (Table IV and Figure 1a).

With respect to test sensitivities, the SUB × TST model showed the lowest bias in almost all the scenarios. Bias gradually increased in the SUB, IND and DEP models in line with the increase in the variability of the SUB × TST effect. Of these three models, the trend of increasing bias was more pronounced with the IND model. Moreover, the IND model always strongly overestimated sensitivities by up to 15% (Table IV, Figure 1b and c). In relation to specificities, bias was low and similar for four models in scenarios I and II. In scenarios III and IV, the SUB × TST model showed the lowest bias. Therefore, the most pronounced differences were observed in scenario IV, in which the SUB × TST model presented a bias of less than 1%, whereas the bias of the other models was greater than 14% of the actual parameter value (Table IV, Figure 1d and e).

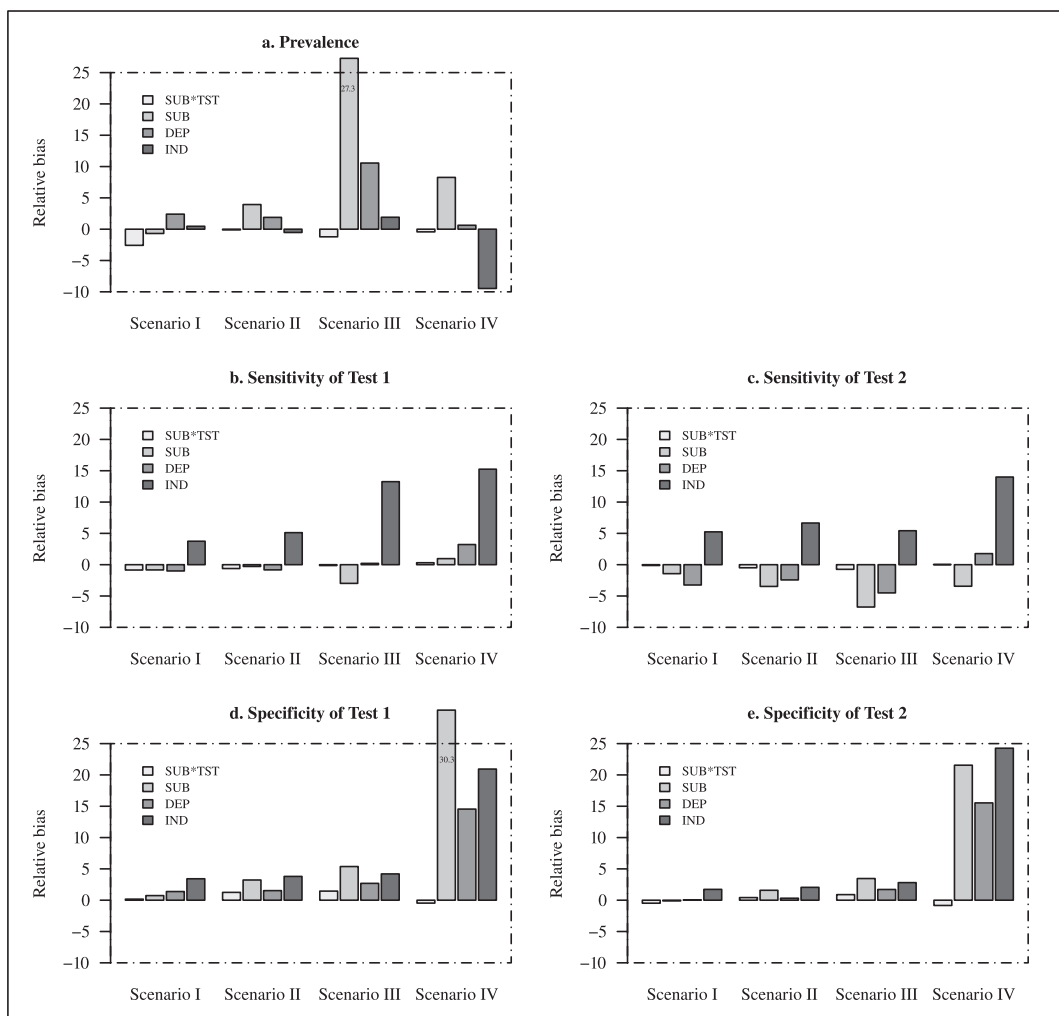


Figure 1. Relative bias calculated in scenarios I, II, III and IV according to the four models: SUB × TST, considers two random effects, the subject and the subject interaction test; SUB, considers only the subject random effect; DEP, dependent model conditional to the true state of the patient (fixed effect); and IND, independent model conditional to the true state of the patient.

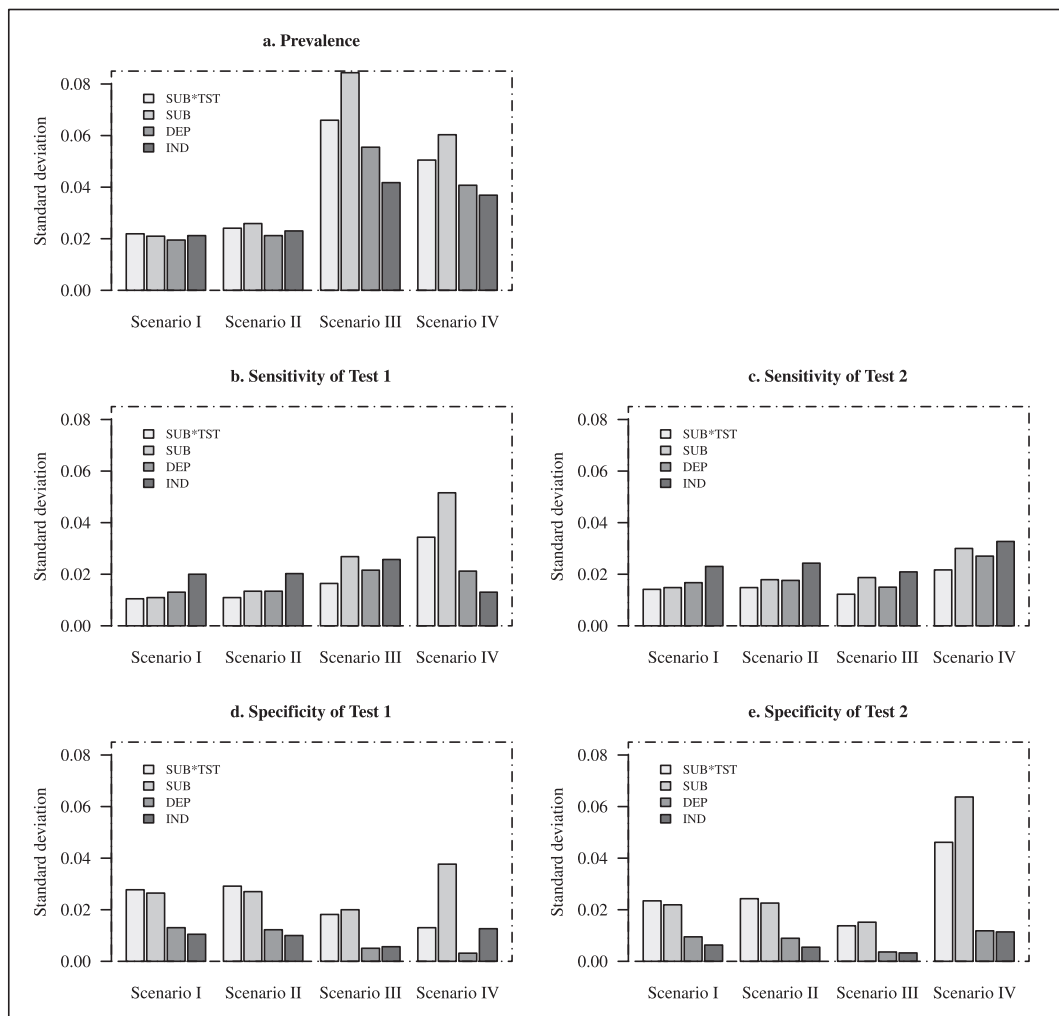


Figure 2. Mean square error calculated for scenarios I, II, III and IV according to the four models: SUB \times TST, subject–test random effects model; SUB, subject random effects model; DEP, conditional dependent fixed effect model; and IND, conditional independent model.

In scenarios I and II, all the parameters (prevalence, sensitivities and specificities) showed a similar MSE (Table IV and Figure 2). In scenarios III and IV, prevalence in the SUB \times TST model had a lower MSE than in the SUB model. However, the DEP model presented a slightly lower MSE than the SUB \times TST model owing to lower variability of the estimates (Table IV and Figure 3). In relation to sensitivities, the MSE was very similar for all the models except in the IND model, which had the highest MSE. With regard to specificities, in scenarios I, II and III, the MSE was similar in all models, while in scenario IV, the SUB, DEP and IND models presented a higher MSE than the SUB \times TST model because of bias (Table IV and Figure 2).

In general, the average coverage rate was very high (close to 100%). This was because the intervals were large owing to the application of only two tests and was not the result of model performance. Exceptions to this were observed in scenarios III and IV for the SUB and IND models. For instance, in the SUB model in both scenarios, only prevalence showed an average coverage lower than the nominal of 90%. In the IND model in scenario III, specificities showed low average coverage, with tests 1 and 2 presenting specificities of 60% and 50%, respectively; while in scenario IV, all the parameters presented coverage lower than 75%, being 74.9% for prevalence, 4.4% and 74.9% for the sensitivity of tests 1 and 2, respectively, and 28.2% and 0.0% for the specificity of tests 1 and 2, respectively. Appendix A presents the boxplot of the estimations of the simulation.

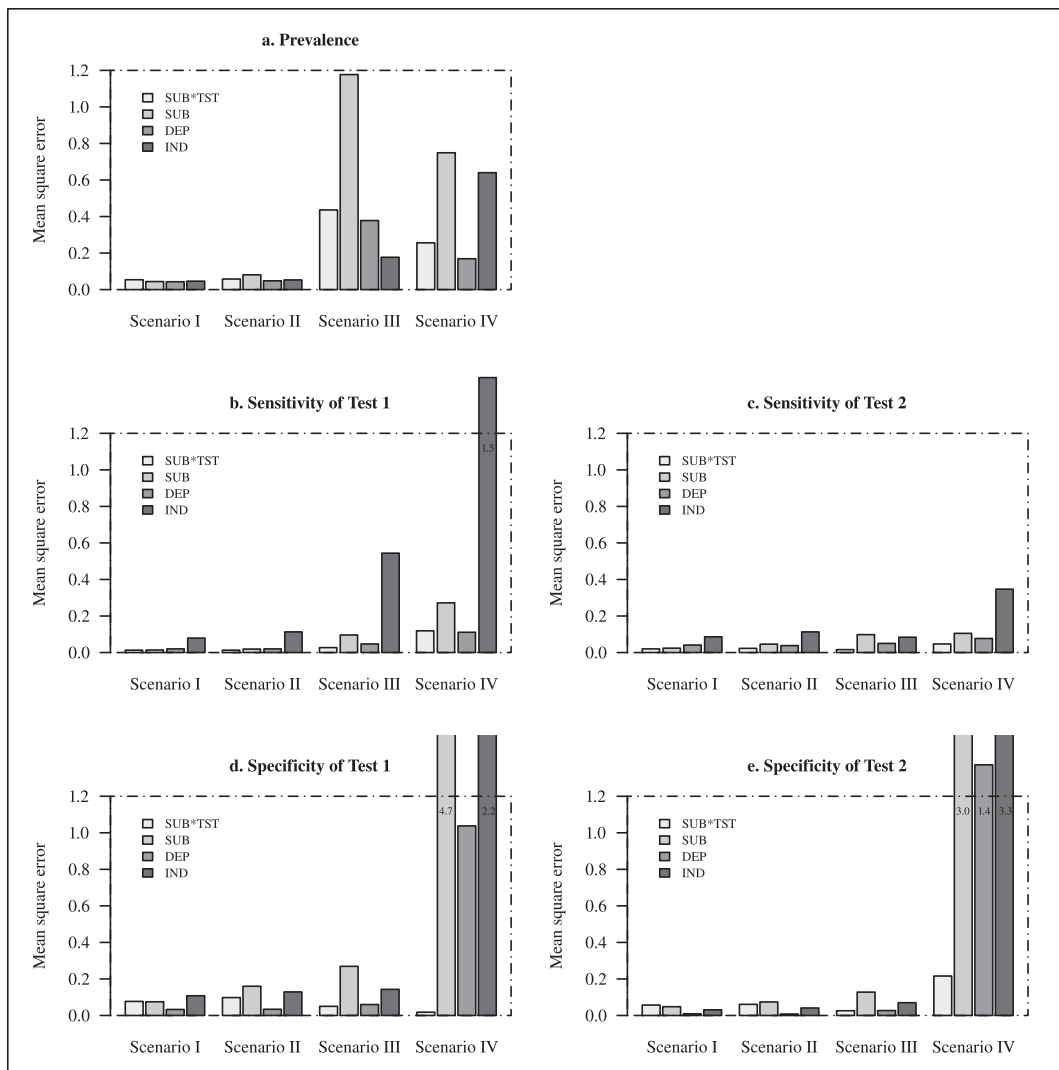


Figure 3. Standard deviation calculated for scenarios I, II, III and IV according to the four models: SUB × TST, subject–test random effects model; SUB, subject random effects model; DEP, conditional dependent fixed effect model; and IND, conditional independent model.

5. Discussion

Here, we propose a latent class model for clinical diagnostic or screening test studies in the absence of a gold standard, which accounts for dependence between replicated same subject measures. This dependence was modelled via the addition of a subject–test random effect. Through a simulation study, the performance of the proposed model was compared with that of the conditional dependent models described by Dendukuri and Joseph [17] (the fixed and random effects models) and the independent model developed by Joseph *et al.* [11].

The simulation study showed that our model presented good accuracy and precision in the different scenarios and provided better results than the subject random effects model and the conditional (in)dependent models. Moreover, the simulation study revealed that when replica variability was higher than subject variability, our model correctly estimated the parameters. Even when replica variability was low, there was no marked difference between the performance of our model and that of the others. Thus, the proposed model could be applied in situations in which the sources of variability are completely unknown.

Another advantage of the proposed extension including random effects is that this model makes it possible to model different correlation structures simply. For example, when the test for each subject is

replicated over time (several time point evaluations) [27, 28], a random effect can then be included in the model with a temporal correlation structure. It is important to define the correlation structure correctly because estimates of diagnostic parameters are biased under a misspecified dependence structure [29, 30]. In the presence of covariates, the subject–test random effects model can be extended to include this information as fixed effects in the linear predictor, as indicated by Dendukuri *et al.* [31] using a probit or a logit link function [14].

When the nature of the study does not have replicated measurements of the same test for each subject, the performance of the conditional dependent and the subject random effects models is better than that of the independent model. In the scenarios in which the probabilities were around 0.5 and the replica effect had greater weight than subject variability (scenarios III and IV), the independent model presented the highest bias with marked differences compared with the other models. When the replica effect was small, there were small differences in the performance of subject-replica random effect, subject random effects and dependent models. However, the independent model again showed the worst properties. Although independence leads to biased estimates of accuracy and prevalence when the independent assumption is violated, a systematic review by van Smeden *et al.* [10] found that more than half of the studies reviewed worked with this assumption, and that most of these studies did not explain or discuss this strong assumption.

The number of parameters estimated in diagnostic studies is limited to the number of tests available. Non-identifiability exists when the degrees of freedom of data are less than the number of parameters to be estimated. However, even if the degree of freedom in the data is as large as the number of parameters, this does not guarantee a identifiable model [32, 33]. To solve this problem, it is necessary to use informative prior distributions for the parameters [31, 34, 35]. The information for prior distributions should be closely linked to beliefs based on scientific opinion, meta-analyses or previous studies. The degrees of freedom can rise if the number of populations studied increases owing to the increase in the number of the parameters available, thereby reducing dependence on prior information [36, 37]. The WinBUGS model code is shown in the appendices to facilitate the work of other authors.

In our simulation study, we decided to use informative prior distribution for all the parameters. Nevertheless, the prior for each scenario was selected by the same expert opinion to avoid influencing the simulation results. For the SUB \times TST model, the mean prior distribution of random effects, b_{20} and b_{21} , was fixed according to the expert opinions: however, a problem arises with the standard deviation of the distribution because no information is available. To solve this problem, we tested several values—0.1, 0.3, 0.5 and 2—and chose the best value in each scenario to avoid problems of convergence.

Summarizing the prior information to be used in the parameter estimations is difficult, and its adaptation to the chosen distribution is an exercise that must be performed carefully to avoid problems of convergence and mistaken inference. An interesting and convenient method is the use of truncated normal distribution, according to the plausible values of the parameters. In our cases, this truncated normal distribution helped to avoid simulations with unexpected errors or bugs, or calculations that were computationally impossible or tended to \pm infinite.

Scenario IV was thought to have a high prevalence (0.75), which may occur in some fields, such as epidemiology and particularly in parasitology, as can be seen in Joseph *et al.* [11] and the references therein. For test sensitivities, we decided that one should be high ($S_1=0.80$) and the other low ($S_2=0.35$), in order to assess how the models operated in the presence of marked differences in test performance. The specificities were also chosen to be small values and far from the perfect test ($C_1=0.70$ and $C_2=0.75$). Furthermore, the amplitude of the prior beliefs of each parameter in scenario IV was not comparable with that of the other scenarios, so that the posterior average length of the credible intervals for the parameters was greater. In the majority of the scenarios, the coverage rates of the parameters were above 95%. The biased parameters did not achieve this coverage.

In conclusion, the proposed model with two random effects is very flexible and allows adjustment with a variety of studies with several random effects, such as those in which the tests are repeated over time or in which the individuals are clustered.

Appendix A: WinBUGS code for subject–test model

```
# Model: Two random effects; subject and subject-test random effects
model{
#=====
# Prior distributions of prevalence, sensitivities and specificities
#=====
prev~dbeta(alpha.prev,beta.prev)
a11~dnorm(A11,tau.A11)
a21~dnorm(A21,tau.A21)
a10~dnorm(A10,tau.A10)
a20~dnorm(A20,tau.A20)
b21~djl.dnorm.trunc(B21,tau.B21,-10,10)
b20~djl.dnorm.trunc(B20,tau.B20,-10,10)
b11~djl.dnorm.trunc(B11,tau.B11,-10,10)
b10~djl.dnorm.trunc(B10,tau.B10,-10,10)
#=====
# Defining the sensitivity and specificity of each subject
#=====
# k = number of individuals; j = number of tests;
r = number of replicas
for (i in 1:k*j*r){
sen1[i]<-phi(a11+b11*t.i[ind[i]]+b21*t.r[tr[i]])
one.minus.spec1[i]<-phi(-a10-b10*t.i[ind[i]]-b20*t.r[tr[i]])
spec1[i]<-1-one.minus.spec1[i]
sen2[i]<-phi(a21+b11*t.i[ind[i]]+b21*t.r[tr[i]])
one.minus.spec2[i]<-phi(-a20-b10*t.i[ind[i]]-b20*t.r[tr[i]])
spec2[i]<-1-one.minus.spec2[i]
}
# k individuals
for (i in 1:k){t.i[i]~dnorm(0,1)}
# j tests
for (i in 1:k*j){t.r[i]~dnorm(0,1)}
#=====
# Conditional probability of a positive observation
#=====
for (i in 1:k*j*r){
pv[i,2,1]<-sen1[i] # sensibility of the test 1
pv[i,1,1]<-1-spec1[i] # 1 - specificity of the test 1
pv[i,2,2]<-sen2[i] # sensibility of the test 2
pv[i,1,2]<-1-spec2[i] # 1 - specificity of the test 2
}
#=====
# Relation between observed data and cond. probability of a positive obs.
#=====
for (i in 1:k*j*r){
D1[i]~dbern(prev)
D[i]<-D1[i]+1
resultat[i]~dbern(pv[i,D[i],Tst[i]])
}
S[1]<-phi(a11/sqrt(1+b11*b11+b21*b21))
S[2]<-phi(a21/sqrt(1+b11*b11+b21*b21))
C[1]<-phi(a10/sqrt(1+b10*b10+b20*b20))
C[2]<-phi(a20/sqrt(1+b10*b10+b20*b20))
}
}
```


Appendix B: Boxplot of the estimations of the simulated data

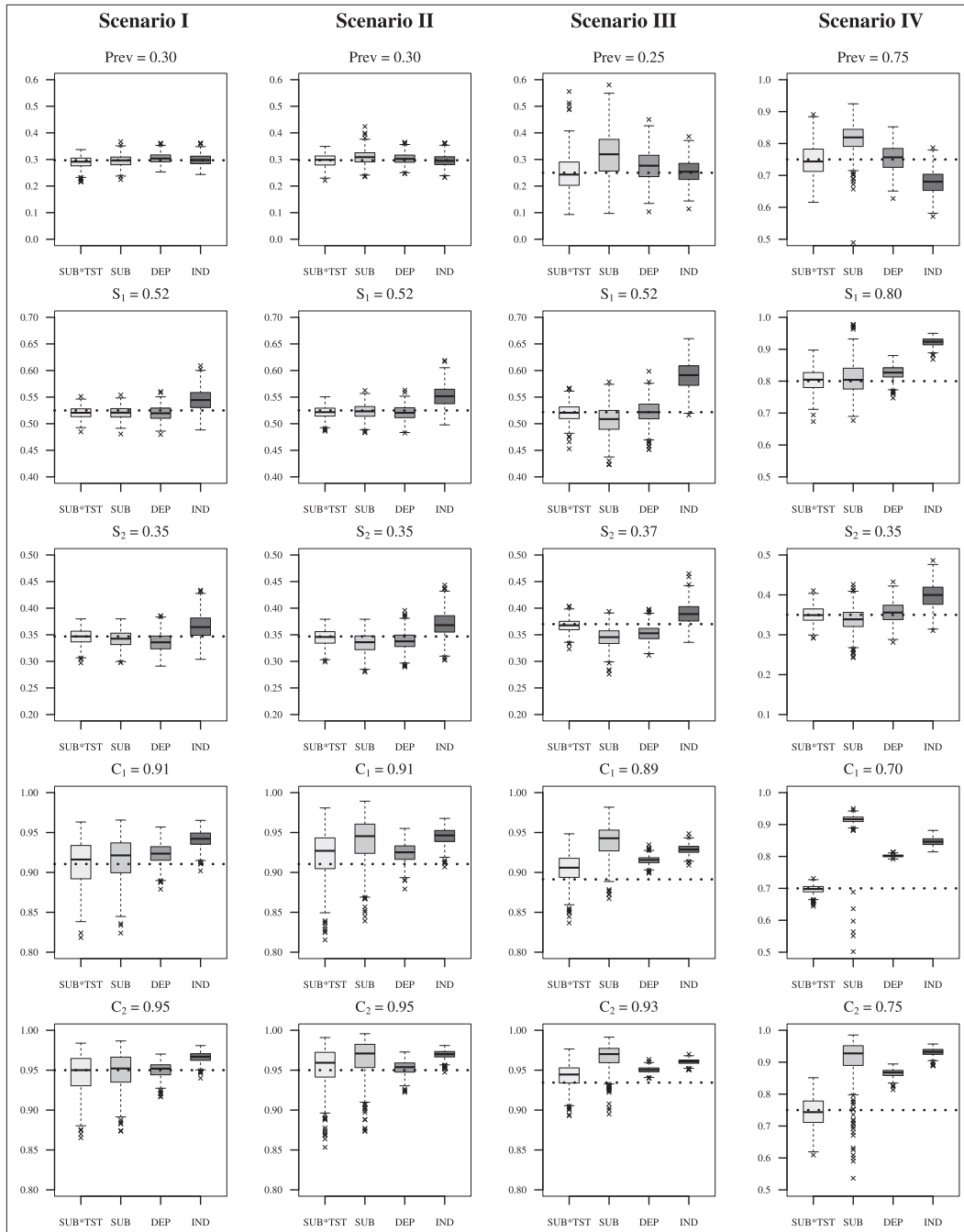


Figure B1. Results of the simulation for scenarios I, II, III and IV according to the four models: SUB \times TST, subject–test random effects model; SUB, subject random effects model; DEP, conditional dependent fixed effect model; and IND, conditional independent model. Prev, prevalence; S_1 , sensitivity of test 1; S_2 , sensitivity of test 2; C_1 , specificity of the test 1; C_2 , specificity of test 2. The true parameter value is in parentheses.

Acknowledgements

This paper is contribution number 22 of the Research Program on Infectious Disease Ecology in the Amazon (RP-IDEA) of the Instituto Leônidas e Maria Deane – Fiocruz Amazônia.

References

1. Lazarsfeld PF. *The Logical and Mathematical Foundations of Latent Structure Analysis*. Princeton University Press: New Jersey, US, 1950.
2. Lazarsfeld PF, Henry NW. *Latent Structure Analysis*. Houghton, Mifflin: Boston, US, 1968.
3. Young MA, Abrams R, Taylor MA, Meltzer HY. Establishing diagnostic criteria for mania. *Journal of Nervous and Mental Disease* 1983; **171**(11):676–682.
4. Young MA, Scheftner WA, Klerman GL, Andreasen NC, Hirschfeld RM. The endogenous sub-type of depression: a study of its internal construct validity. *The British Journal of Psychiatry: the Journal of Mental Science* 1986; **148**(3):257–267.
5. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. *Statistics in Medicine* 1986; **5**(1): 21–27.
6. Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* 1985; **41**(4):959–968.
7. Thibodeau LA. Evaluating diagnostic tests. *Biometrics* 1981; **37**(4):801–804.
8. Walter SD, Macaskill P, Lord SJ, Irwig L. Effect of dependent errors in the assessment of diagnostic or screening test accuracy when the reference standard is imperfect. *Statistics in Medicine* 2012; **31**(11–12):1129–1138.
9. Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Statistics in Medicine* 2014; **33**(24):4141–4169.
10. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot AJ. Latent class models in diagnostic studies when there is no reference standard—a systematic review. *American Journal of Epidemiology* 2014; **179**(4):423–431.
11. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 1995; **141**(3):263–272.
12. Achcar JA, Martinez EZ, Louzada-Neto F. Binary data in the presence of covariates and misclassifications: a Bayesian approach. *Brazilian Journal of Probability and Statistics* 2005; **19**:65–84.
13. Martinez EZ, Achcar JA, Louzada-Neto F. Bayesian estimation of diagnostic tests accuracy for semi-latent data with covariates. *Journal of Biopharmaceutical Statistics* 2005; **15**(5):809–821.
14. Martinez EZ, Louzada-Neto F, Derchain SFM, Achcar JA, Gontijo RC, Sarian LOZ, Syrjänen KJ. Bayesian estimation of performance measures of cervical cancer screening tests in the presence of covariates and absence of a gold standard. *Cancer Informatics* 2008; **6**:33–46.
15. de Araujo Pereira G, Louzada-Neto F, de Fátima Barbosa V, Ferreira-Silva MM, Moraes-Souza H. A general latent class model for performance evaluation of diagnostic tests in the absence of a gold standard: an application to Chagas disease. *Computational and Mathematical Methods in Medicine* 2012; **2012**:1–12.
16. Qu YS, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996; **52**(3):797–810.
17. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; **57**(1):158–167.
18. Ntzoufras Ioannis. *Bayesian Modeling Using Winbugs*, Vol. 698. John Wiley & Sons: New Jersey, US, 2009.
19. Lunn DJ, Thomas A, Best N, Spiegelhalter D. Winbugs—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**(4):325–337.
20. Team R Core. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2013.
21. Sturtz S, Ligges U, Gelman AE. R2winbugs: a package for running WinBUGS from R. *Journal of Statistical Software* 2005; **12**(3):1–16.
22. Gonçalves AQ, Abellana R, da Silva HDP, Santos I, Serra PT, ao GRJ, Orlandi PP, Ascaso C. Comparison of the performance of two spontaneous sedimentation techniques for the diagnosis of human intestinal parasites in the absence of a gold standard. *Acta Tropica* 2014; **131**(0):63–70.
23. Thisted RA. *Elements of Statistical Computing: Numerical Computation*, Vol. 1. CRC Press: New York, US, 1988.
24. Goldstein H. *Multilevel Statistical Models*, Vol. 922. John Wiley & Sons: New Jersey, US, 2011.
25. Browne WJ, Subramanian SV, Jones K, Goldstein H. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2005; **168**(3):599–613.
26. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual, 2003.
27. Espeland MA, Murphy WC, Leverett DH. Assessing diagnostic reliability and estimating incidence rates associated with a strictly progressive disease: dental caries. *Statistics in Medicine* 1988; **7**(3):403–416.
28. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 1998; **7**(4):354–370.
29. Albert PS. Estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard. *Statistics in Medicine* 2009; **28**(5):780–797.
30. Zhang B, Chen Z, Albert PS. Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. *Biometrics* 2012; **68**(4):1294–1302.
31. Dendukuri N, Hadgu A, Wang L. Modeling conditional dependence between diagnostic tests: a multiple latent variable model. *Statistics in Medicine* 2009; **28**(3):441–461.
32. Swartz TB, Haitovsky Y, Vexler A, Yang TY. Bayesian identifiability and misclassification in multinomial data. *Canadian Journal of Statistics* 2004; **32**(3):285–302.
33. Jones G, Johnson WO, Hanson TE, Christensen R. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 2010; **66**(3):855–863.
34. Gustafson P. On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables. *Statistical Science* 2005; **20**(2):111–129.
35. Gustafson P. The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard. *Statistics in Medicine* 2005; **24**(8):1203–1217.

H. D. PEREIRA DA SILVA *ET AL.*

36. Toft N, Jørgensen E, Højsgaard S. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine* 2005; **68**(1):19–33.
37. Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2003; **52**(1):63–76.

B Paper II

RESEARCH ARTICLE

Prevalence and Diagnosis of Latent Tuberculosis Infection in Young Children in the Absence of a Gold Standard

Tomas Maria Perez-Porcuna^{1,2}*, Hélio Doyle Pereira-da-Silva¹, Carlos Ascaso^{1,3}, Adriana Malheiro^{4,5}, Samira Bühner⁶, Flor Martinez-Espinosa^{7,8}, Rosa Abellana^{1*}

1 Departament de Salut Pública, Facultat de Medicina, Universitat de Barcelona, Barcelona, Catalunya, Spain, **2** Servei de Pediatria, CAP Valldoreix, Unitat de Investigació Fundació Mútua Terrassa, Hospital Universitari Mútua Terrassa, Terrassa, Catalunya, Spain, **3** Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Catalunya, Spain, **4** Laboratório Multidisciplinar - Fundação de Hematologia e Hemoterapia do Amazonas, Manaus, Amazonas, Brazil, **5** Instituto de Ciências Biológicas, Universidade Federal do Amazonas, Manaus, Amazonas, Brazil, **6** Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiás, Goiânia, Brazil, **7** Fundação de Medicina Tropical Dr. Heitor Vieira Dourado (FMT-HVD), Manaus, Amazonas, Brazil, **8** Instituto Leônidas e Maria Deane - Fiocruz Amazônia, Fundação Oswaldo Cruz, Manaus, Amazonas, Brazil

* These authors contributed equally to this work.

* rabellana@ub.edu (RA); tomas.perez.porcuna@gmail.com (TPP)


 OPEN ACCESS

Citation: Perez-Porcuna TM, Pereira-da-Silva HD, Ascaso C, Malheiro A, Bühner S, Martinez-Espinosa F, et al. (2016) Prevalence and Diagnosis of Latent Tuberculosis Infection in Young Children in the Absence of a Gold Standard. *PLoS ONE* 11(10): e0164181. doi:10.1371/journal.pone.0164181

Editor: Joan A Caylà, Agencia de Salut Publica de Barcelona, SPAIN

Received: January 31, 2016

Accepted: September 21, 2016

Published: October 26, 2016

Copyright: © 2016 Perez-Porcuna et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from Figshare at DOI [10.6084/m9.figshare.3859179](https://doi.org/10.6084/m9.figshare.3859179).

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: BCG, Calmette-Guérin bacillus vaccine; CI, Confidence interval; IGRAs, Interferon- γ release assays; IQR, Interquartile range; HIV, Human immunodeficiency virus; LTBI, Latent

Abstract

Introduction

For adequate disease control the World Health Organization has proposed the diagnosis and treatment of latent tuberculous infection (LTBI) in groups of risk of developing the disease such as children. There is no gold standard (GS) test for the diagnosis of LTBI. The objective of this study was to estimate the prevalence of LTBI in young children in contact with a household case of tuberculosis (TB-HCC) and determine the accuracy and precision of the Tuberculin Skin Test (TST) and QuantiFERON-TB Gold in-tube (QFT) used in the absence of a GS.

Methods

We conducted a cross-sectional study in children up to 6 years of age in Manaus/Brazil during the years 2009–2010. All the children had been vaccinated with the BCG and were classified into two groups according to the presence of a TB-HCC or no known contact with tuberculosis (TB). The variables studied were: the TST and QFT results and the intensity and length of exposure to the index tuberculosis case. We used the latent class model to determine the prevalence of LTBI and the accuracy of the tests.

Results

Fifty percent of the children with TB-HCC had LTBI, with the prevalence depending on the intensity and length of exposure to the index case. The sensitivity and specificity of TST were 73% [95% confidence interval (CI): 53–91] and 97% (95%CI: 89–100), respectively,

tuberculous infection; Mtb, *Mycobacterium tuberculosis*; MTC-score, *Mycobacterium tuberculosis* contact score; PPD, Purified protein derivatives; QFT, QuantiFERON-TB Gold in-tube; TB, Tuberculosis; TB-HCC, Household case of tuberculosis; TST, Tuberculin Skin Test; TU, Tuberculin units; WHO, World Health Organization.

versus 53% (95%CI: 41–66) and 81% (95%CI:71–90) for QFT. The positive predictive value of TST in children with TB-HCC was 91% (95%CI: 61–99), being 74% for QFT (95% CI: 47–95).

Conclusions

This is one of the first studies to estimate the prevalence of LTBI in children and the parameters of the main diagnostic tests using a latent class model. Our results suggest that children in contact with an index case have a high risk of infection. The accuracy and the predictive value of the two tests did not significantly differ. Combined use of the two tests showed scarce improvement in the diagnosis of LTBI.

Introduction

Tuberculosis (TB) is one of the most prevalent contagious infections worldwide[1]. Nonetheless, although the incidence of TB is decreasing by approximately 1.5% annually, according to the elimination target set for 2050 proposed by the World Health Organization (WHO) this reduction is insufficient[2]. This may be explained by the underdiagnosis of the disease which has been estimated to be around one third of the incidence cases[1]. For adequate disease control the WHO has proposed the need for not only adequate diagnosis and treatment of TB cases but also the diagnosis and treatment of latent tuberculous infection (LTBI) in groups of risk of developing the disease such as children[2–4]. Most cases of pediatric age are due to household contact with adults with TB[5,6]. Therefore, pediatric control measures involve contact screening to detect and treat children with TB and LTBI in order to prevent the disease in those infected[7].

The tuberculin skin test (TST) and Interferon- γ release assays (IGRAs) are used for the diagnosis of LTBI and are based on challenging the immune system in order to detect whether it recognizes the antigens of *Mycobacterium tuberculosis* (Mtb)[8,9]. However, relatively few studies have been carried out in children, and the results described suggest that the performance of these assays differs from those obtained in adults, especially in countries with a high incidence of TB and TB in young children[10]. The absence of a reference test for these techniques makes their evaluation and comparison difficult[8]. With the latent class model defined by Dendukuri and Joseph (2001) the prevalence and the performance parameters of several diagnostic tests without gold standard can be estimated[11].

The objective of the present study was to estimate the prevalence of LTBI and determine the accuracy and safety of the TST and QuantiFERON-TB Gold In-Tube (QFT) tests in young children taking into account the intensity of exposure of the children to an index case. The population studied included children up to 6 years of age vaccinated with the BCG in Manaus (Brazil) during the years 2009–2010.

Material and Methods

Study design and population

We conducted a cross-sectional study of children to calculate the prevalence and accuracy of the TST and QFT. The study included a sample of children of up to 6 years of age without recent known contact with an index TB case and another group of children in contact with an index case within the previous 12 months. Case recruitment was performed in the *Policlínica*

Cardoso Fontes (regional reference center for TB) and in the *Fundação de Medicina Tropical Dr. Heitor Vieira Dourado* (FMT-HVD), Manaus, Amazonas, Brazil from March 2009 to February 2010. Adults (greater than 12 years of age) diagnosed with TB in both centers were questioned about contact with children from 0 to 6 years of age. Those responding affirmatively were asked to bring the children to the center for evaluation and were invited to participate in the study. All the adult index cases were sputum smear and/or culture positive. Subjects receiving treatment or prophylaxis for TB were excluded. The controls were selected among children attended in both centers for reasons other than TB. The study was undertaken at an outpatient level.

Ethics Statement

The study was approved by the Ethical Committee in Investigation of the FMT-HVD, October 26, 2007 (Protocol 2865–07). The legal guardians of all the participants provided written informed consent for inclusion in the study.

Data collection

Demographic data as well as the epidemiologic history of exposure to an index case of TB and the clinical history of the patient and physical examination were recorded. Chest X-ray, the TST, and blood analysis (prior to the TST) were performed. The human immunodeficiency virus (HIV) test was not obligatory but was recommended to all the participants.

The *Mycobacterium tuberculosis* contact score (MTC-score) was used to evaluate the intensity of exposure. The MTC-score is from 0 to 15 and is based on the assumption that the gradient of Mtb exposure is a composite function of the infectivity of the index case (0–4), the duration of exposure in hours per day (0–4), the relationship to the index case (0–4) and the type of exposure (0–3) [12].

The time (months) from symptom onset to the time of initiation of treatment of the index case was calculated to measure the total exposure time (contagion) of the child to the index case.

Procedures

The TST was performed with an intradermic injection of 2 tuberculin units (TU) of PPD RT23 (Statens Serum Institut, Copenhagen, Denmark) and read 72 hours thereafter. A strong TST reaction (TST+) was defined when induration was ≥ 10 mm [7].

The QFT test (Cellestis, Carnegie, Australia) was carried out and interpreted according to the manufacturer's instructions by experienced laboratory technicians who were unaware of the data of the study subjects.

Statistical analysis

Categorical variables were described using absolute frequencies and percentages and quantitative variables using the 25, 50 and 75 percentiles. The chi-square test was used to study the association between the results of the two tests and the variables of gender and age groups (less than or equal to 24 months and greater than 24) and the Mann Whitney U test was used for the exposure time and the MTC-score variables. Concordance between the QFT and the TST tests was measured by the kappa index.

The statistical analysis was divided into two phases. In the first phase the accuracy of the two tests was estimated, and the prevalence of LTBI was determined based on whether the children had been in contact with an index case or not. To estimate these parameters the proposal

of Dendukuri N. and Joseph L. (2001) was used[11], stratifying the model to estimate the prevalence of LTBI in children with and without contact with an index case. Then, it was assumed that the combinations of both test results were distributed by a multinomial distribution with probabilities equal to:

$$\begin{aligned}
 P(\text{QFT} = +, \text{TST} = +) &= P_i \cdot [(S_1 \cdot S_2) + \text{cov}S_{12}] + (1 - P_i) \cdot [((1 - C_1) \cdot (1 - C_2)) + \text{cov}C_{12}] \\
 P(\text{QFT} = +, \text{TST} = -) &= P_i \cdot [(S_1 \cdot (1 - S_2)) - \text{cov}S_{12}] + (1 - P_i) \cdot [((1 - C_1) \cdot C_2) - \text{cov}C_{12}] \\
 P(\text{QFT} = -, \text{TST} = +) &= P_i \cdot [(1 - S_1) \cdot S_2 - \text{cov}S_{12}] + (1 - P_i) \cdot [(C_1 \cdot (1 - C_2)) - \text{cov}C_{12}] \\
 P(\text{QFT} = -, \text{TST} = -) &= P_i \cdot [(1 - S_1) \cdot (1 - S_2)) + \text{cov}S_{12}] + (1 - P_i) \cdot [(C_1 \cdot C_2) + \text{cov}C_{12}]
 \end{aligned}$$

Where P_i is the prevalence of LTBI for the i th stratum (1: children with contact with an index case, and 2: without contact with an index case), S_1 and S_2 are the sensitivity of QFT and the TST test respectively, and C_1 and C_2 are the specificities. $\text{Cov}S_{12}$ and $\text{cov}C_{12}$ are the covariance between tests among subjects with latent infection and non-latent infection, respectively.

Both prevalences were assumed to have a Beta(1,1) prior distribution. Based on the literature of LTBI, we reasonably considered that the sensitivity and the specificity of the two tests lie within the range of 50%-100%. Consequently, the prior distribution for these parameters was a Beta(8.25, 2.75). The covariance parameters are taken to have Beta(1,1) prior distribution, and their feasible ranges were given as $0 \leq \text{cov}S_{12} \leq \min(S_1, S_2) - S_1 S_2$ and $0 \leq \text{cov}C_{12} \leq \min(C_1, C_2) - C_1 C_2$,

According to the proposal of Martinez E.Z. (2008),[13], in the second phase the previous model is widened and rather than study the prevalence based on whether there was or was not exposure to an index case it is studied based on the magnitude of intensity and the time of exposure to an index case. Denoting the true infection status (latent) of the k th subject as D_k (0: among non-infected subjects and 1 among infected subjects), the prevalence of LTBI is the probability of being truly infected $P(D_k = 1)$, and it is given as covariates:

$$P(D_k = 1) = \Phi(\beta_0 + \beta_1 \times \text{ExpositionTime} + \beta_2 \times \text{MTC} - \text{Score})$$

Φ : represents the cumulative distribution function of the Normal(0,1) distribution

The conditional dependence between tests is modeled using a random effects model[14]. In this case, the sensitivities and specificities of the tests are modeled as a function of subject-specific random effect, i_k . Thus, the probability that the k th subject has a positive result in the j th test (1: QFT and 2: TST) is given as:

$$P(T_{jk} = + | D_k = 1, I_k = i_k) = \Phi(\beta_{3j} + b_1 i_k)$$

And has a negative result

$$P(T_{jk} = - | D_k = 0, I_k = i_k) = \Phi(\beta_{4j} + b_0 i_k)$$

where:

i_k : is the intensity of the k th subject, and it is a random variable following a Normal(0,1) distribution.

Thus, the mean sensitivity of the j th test in all the subjects was:

$$S_j = P(T_j = + | D = 1) = \int_{-\infty}^{\infty} P(T_{jk} = + | D_k = 1, I_k = i_k) d\Phi(i_k) = \Phi\left(\frac{\beta_{3j}}{\sqrt{1 + b_1^2}}\right)$$

and the mean specificity of the j th test in all the subjects was:

$$C_j = P(T_j = - | D = 0) = \int_{-\infty}^{\infty} P(T_{jk} = - | D_k = 0, I_k = i_k) d\Phi(i_k) = \Phi\left(-\frac{\beta_{4j}}{\sqrt{1 + b_0^2}}\right)$$

Using the estimates obtained in the first phase of the prevalence, sensitivity, specificity and the covariance between test, the parameters of the prior distributions from the constants of the model were $\beta_0 \sim \text{Normal}(-1.59, 1.8)$, $\beta_{31} \sim \text{Normal}(0.24, 0.32)$, $\beta_{32} \sim \text{Normal}(0.83, 0.49)$, $\beta_{41} \sim \text{Normal}(1.15, 0.32)$ and $\beta_{42} \sim \text{Normal}(2.01, 0.48)$ and for the b_1 and b_0 were $\text{Normal}(0.77, 0.5)$ and $\text{Normal}(0.99, 0.5)$, respectively. The prior distribution for the coefficient of the exposure time and MTC-score was a Normal $(0, 0.1)$, respectively [15].

Results

A total of 121 children were studied, 92 (76.0%) of whom had had contact with an index case and 29 (24.0%) had not. All the children studied had received the BCG vaccination in the first 2 months of life. Sixteen children (13.2%) were excluded from the analysis because they presented an indeterminate result with QFT. Of the 105 children analyzed, 41.9% were male and 23.8% were less than 24 months of age. Twenty-five children (23.8%) had had no contact with an index case but in 3 children result was positive for the QFT test but not for the TST test (Fig 1). There were no significant differences when comparing age and gender in exposed children and controls. Of the 80 (76.2%) children having contact with an index case, the median MTC-score was 12.0 points (IQR = 3.75), and the time of exposure was 2 months (IQR = 4.0). Thirty-one children (38.8%) showed a positive QFT result and 33 (41.2%) did so with the TST, but only 19 children (23.9%) were positive for both tests. None of the positive TST results were weak (between 5–9 mm). The two tests showed a low concordance of 0.364 ($p < 0.001$).

The results of the two tests were not associated with either the gender or age of the children. The time and the intensity of the exposure were associated with the QFT and TST results (Table 1). The time and the intensity of the exposure of the children with a positive QFT result (median time of exposure = 2; median MTC-score = 12.0) was greater than that of children with a negative QFT (median time of exposure = 1.0; $p = 0.026$; median MTC-score = 10.0, $p = 0.021$). The time and the intensity of the exposure of the children with a positive TST result (median time of exposure = 4; median MTC-score = 12.0) was greater than in those with a negative TST result (median time of exposure = 1.0, $p < 0.001$, median MTC-score = 9.5, $p < 0.001$).

Bayesian Analysis

The prevalence of infection in the group of children without contact with an index case was 0.04 (95% CrI [0.00, 0.20], being 0.50 in children with contact (95% CrI [0.28, 0.81]). The sensitivity of the QFT test was 0.58 (95% CrI [0.41, 0.78]) and 0.75 (95% CrI [0.49, 0.94]) for the TST test. The specificity of the QFT was 0.79 (95% CrI [0.67, 0.91]) and 0.92 (95% CrI [0.78, 0.98]) for the TST test (Table 2). The positive predictive values of both tests were very low in the group of children with no contact with an index case and thus, the prevalence of infection was very low. In the group of children with contact with an index case, the probability of a child diagnosed as positive really being infected was 0.74 (95% CrI [0.47, 0.95]) for QFT and 0.91 (95% CrI [0.61, 0.99]) for TST. Moreover, the probability of a child diagnosed as negative really not being infected was 0.65 (95% CrI [0.27, 0.88]) for QFT and 0.79 (95% CrI [0.31, 0.96]) for TST. On evaluating the positive predictive value with the results of both tests together we found that the probability of really being infected when the two tests were positive was 0.93

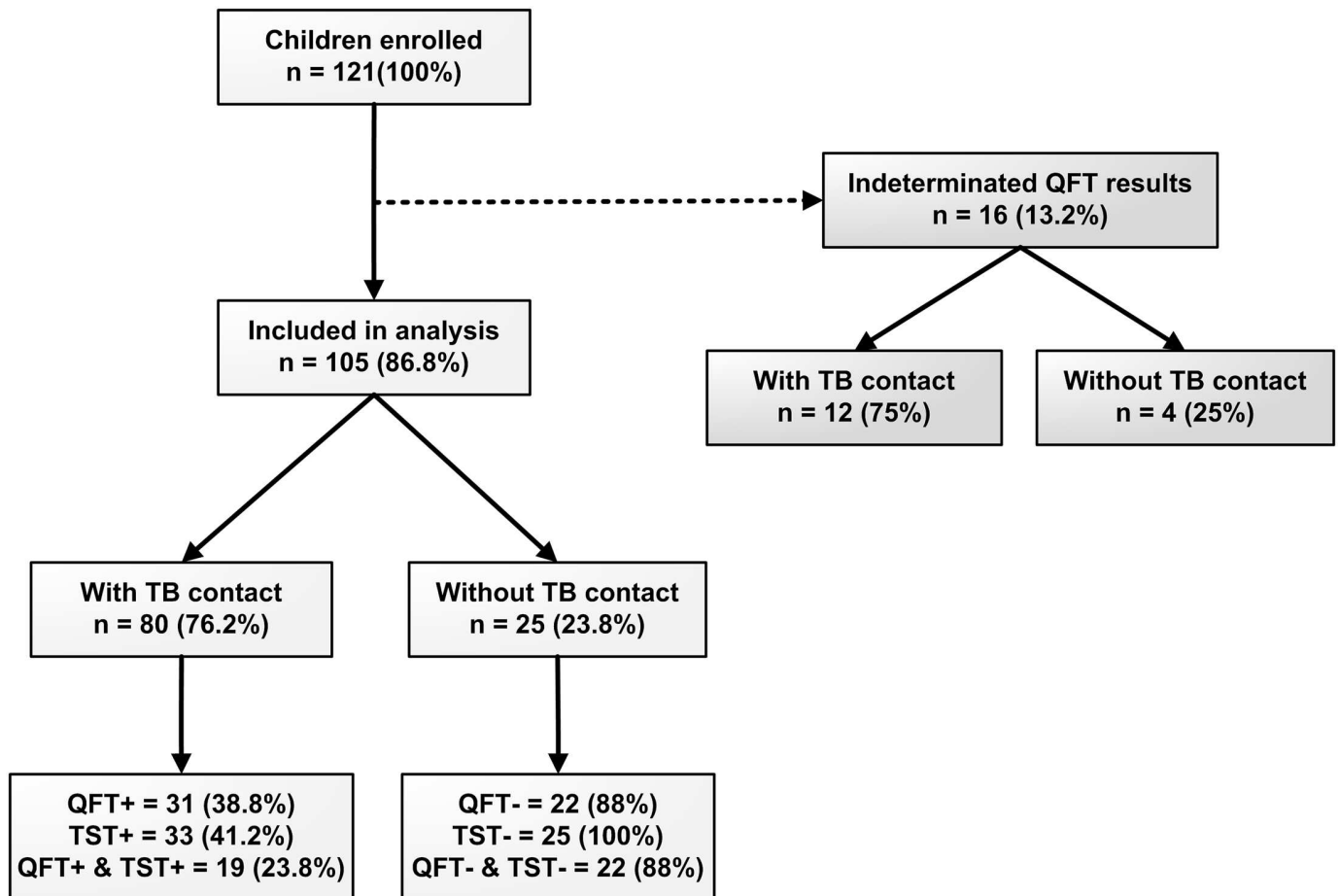


Fig 1. Flow diagram of enrollment. TB: tuberculosis, QFT: QuantiFERON-TB Gold In-Tube, TST: Tuberculin Skin Test.

doi:10.1371/journal.pone.0164181.g001

Table 1. Descriptive analysis and comparisons between the results of the QFT and the TST test in children up to 6 years of age in Manaus, Brazil.

	QFT			TST		
	+	-	p-value	+	-	p-value
Male	14 (31.8%)	30 (68.2%)	0.917	14 (31.8%)	30 (68.2%)	0.942
Female	20 (32.8%)	41 (67.2%)		19 (31.1%)	42 (68.9%)	
Age ≤ 24 months	7 (28.0%)	18 (72.0%)	0.592	8 (32.0%)	17 (68.0%)	0.944
Age >24 months	27 (33.8%)	53 (66.2%)		25 (31.2%)	55 (68.8%)	
Age (months)*	47.76 (22.5)	43.68 (21.5)	0.384	43.91 (22.7)	45.50 (22.4)	0.737
Exposure Time**	2 (5.25–1)	1 (0–3)	0.026	4 (1–10)	1 (0–3)	<0.001
MTC-score**	12 (14.0–8.75)	10 (0–12)	0.021	12 (10–14)	9.5 (0–12)	<0.001

* mean (standard deviation). Student t.

**median (quartile 25–quartile 75). Mann-Whitney U.

TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, MTC-SCORE: *Mycobacterium tuberculosis* contact score.

doi:10.1371/journal.pone.0164181.t001

Table 2. Results of the prevalence of LTBI according to exposure to an index TB case and the sensitivity and the specificity of QFT and TST in children up to 6 years of age in Manaus, Brazil (first phase).

	Median	95% CrI
P(LTBI) Unknown exposure of children	0.04	0.00–0.20
P(LTBI) Exposed Children	0.50	0.28–0.81
Sensitivity of QFT	0.58	0.41–0.78
Specificity of QFT	0.79	0.67–0.91
Sensitivity of TST	0.75	0.49–0.94
Specificity of TST	0.92	0.78–0.98

CrI: Credibility interval, TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, MTC-SCORE: *Mycobacterium tuberculosis* contact score, LTBI: latent tuberculosis infection.

doi:10.1371/journal.pone.0164181.t002

(95% CrI [0.69, 0.99]). When only the QFT was positive this value was 0.35 (95% CrI [0.04, 0.90]), being 0.90 (95% CrI [0.30, 0.99]) when only the TST was positive. In relation to the negative predictive values, the probability of not being infected when both tests were negative was 0.83 (95% CrI [0.37, 0.97]), and when only the QFT was negative this value was 0.1 (95% CrI [0.01, 0.7]), being 0.65 (95% CrI [0.10, 0.96]) (Table 3) when only the TST was negative.

Taking into account the magnitude of the intensity (MTC-score) and the time of exposure, according to the model used, the prevalence of LTBI increased with the intensity ($\beta = 0.14$, 95% CrI [0.06, 0.23]) and the time of exposure ($\beta = 0.19$, 95% CrI [0.05, 0.49]) (Tables 4 and 5). Fig 2 shows the values of prevalence for the times of exposure of 0, 1, 3, 6 and 12 months and MTC-score values of 4, 8, 13 and 15. Children with an exposure time of 12 months and MTC-score greater than 8 presented prevalences greater than 75% while children with low exposure times (less than or equal to 1 month) and a MTC-score less than 5 presented prevalences of less than 25%. On taking into account the intensity and time of exposure in this model, the sensitivity and specificity of the tests was slightly modified. Fig 3A and 3B show the positive predictive values based on the intensity and time of exposure. The results indicate that with high exposure times (greater than or equal to 6 months) and exposures of greater than or equal to 13, the predictive value of QFT was greater than 0.70, being greater than 0.9 for TST. The negative predictive values of both tests showed a similar behavior, being high with a low time and intensity of exposure and low when the time and intensity were high (Fig 3C and 3D).

Table 3. Probability of LTBI in children up to 6 years of age according to the results of both tests in the latent class model (first phase).

	Unknown exposure of children		Exposed Children	
	Median	95% CrI	Median	95% CrI
P(LTBI QFT+)	0.01	0.00–0.51	0.74	0.47–0.95
P(LTBI TST+)	0.27	0.01–0.79	0.91	0.61–0.99
P(LTBI QFT+ TST+)	0.31	0.01–0.83	0.93	0.69–0.99
P(LTBI QFT+ TST-)	0.02	0.00–0.34	0.35	0.04–0.90
P(LTBI QFT- TST+)	0.29	0.01–0.82	0.90	0.31–0.99

CrI: Credibility interval, TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, MTC-SCORE: *Mycobacterium tuberculosis* contact score, LTBI: latent tuberculosis infection.

doi:10.1371/journal.pone.0164181.t003

Table 4. Probability of non LTBI (LTBI^c) in children up to 6 years of age according to the results of both tests in the latent class model (first phase).

	Unknown exposure of children		Exposed Children	
	Median	95% CrI	Median	95% CrI
P(LTBI ^c QFT-)	0.98	0.88–1.00	0.65	0.27–0.88
P(LTBI ^c TST-)	0.99	0.90–1.00	0.79	0.31–0.96
P(LTBI ^c QFT+ TST-)	0.98	0.66–1.00	0.64	0.10–0.96
P(LTBI ^c QFT- TST+)	0.76	0.18–0.99	0.10	0.01–0.69
P(LTBI ^c QFT- TST-)	0.99	0.93–1.00	0.83	0.37–0.97

CrI: credibility interval, TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, MTC-SCORE: *Mycobacterium tuberculosis* contact score, LTBI: latent tuberculosis infection.

doi:10.1371/journal.pone.0164181.t004

Discussion

Estimation of the prevalence of a disease and the parameters of sensitivity, specificity and predictive values of the tests used to study this disease in the absence of a gold standard are an important advance for the evaluation of diagnostic tests. This is of special relevance in the current epidemiological and diagnostic context of the study of LTBI in which it is impossible to determine if an individual is really infected or whether this person presents immune response to the previous infection.

The prevalence of LTBI in the group exposed to an index case was 50.4% (27.1%–83.3%), and this prevalence or probability of becoming infected increases with both the intensity (MTC-score) and the time of exposure to the index case[16]. The prevalence in the control group was 3.9% (0.10%–20.6%).

The prevalence obtained suggests that young children are very susceptible to exposure to an index case, with this prevalence possibly being greater than in other population groups[17]. On the other hand, most of the cases of LTBI and TB disease in young children are the result of exposure to a known close index case of TB[5]. These two reasons reinforce the crucial importance of diagnostic strategies of infection and disease in household pediatric contacts in order to control TB[18].

Knowledge of the prevalence of LTBI based on the time and intensity of exposure allows estimation of the prevalence and thus, the risk of LTBI in a young child before obtaining the result of the diagnostic tests. Indeed, we found that the high intensity and time of exposure result in very high prevalences (greater than 70%).

Table 5. Estimation of the sensitivity and specificity of QFT and TST, and the coefficients of regression of the variables of time of exposure and MTC-Score associated with the prevalence (second phase).

Variable	Median	95% CrI
Exposure Time	0.19	0.05–0.49
MTC-Score	0.14	0.06–0.23
Sensitivity of QFT	0.53	0.41–0.66
Specificity of QFT	0.81	0.71–0.90
Sensitivity of TST	0.73	0.53–0.91
Specificity of TST	0.97	0.89–1.00

CrI: credibility interval, TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, MTC-SCORE: *Mycobacterium tuberculosis* contact score, LTBI: latent tuberculosis infection.

doi:10.1371/journal.pone.0164181.t005

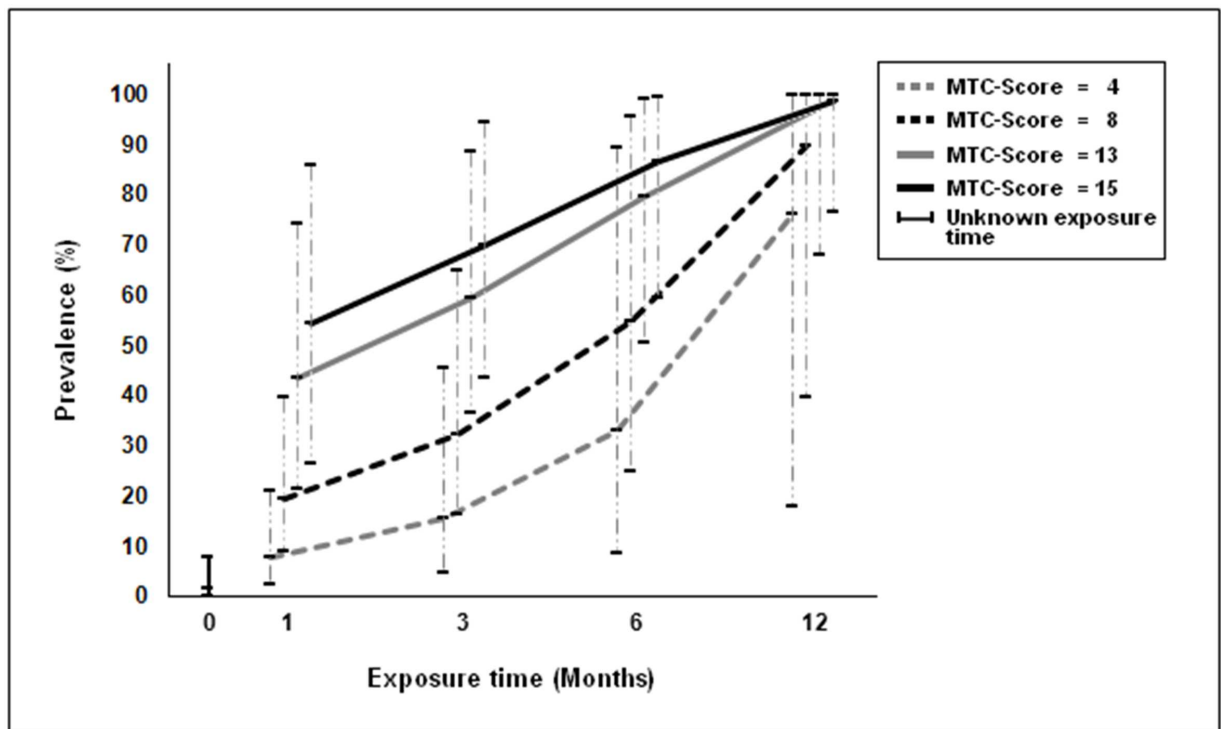


Fig 2. Prevalence of LTBI according to the time of exposure of 0, 1, 3, 6 and 12 months and the MTC-score values of 4, 8, 13 and 15 in children up to 6 years of age in Manaus, Brazil. MTC-SCORE: *Mycobacterium tuberculosis* contact score.

doi:10.1371/journal.pone.0164181.g002

Since it is difficult to precisely know at what time an individual is contagious, we estimated the period of contagion from the time at which the index case shows symptomatology compatible with TB until the time of treatment. Likewise, to evaluate the intensity of exposure with the index case we used the MTC-score which assesses the capacity of contagion of the index case and the grade of relationship with the same. Our results suggest that the use of the time of exposure and MTC-score are feasible for study models of LTBI in children.

Concordance between the TST and QFT was lower than expected. This has been reported previously in young children, especially in regions with a high incidence of TB[19] and may be related to coinfection with helminths, malnutrition, short exposure times, immunosuppression and reversion phenomena in the IGRAs[16,20,21].

Despite the low concordance, the tests presented similar sensitivities and specificities, without significant differences. It is of note that the specificity of the TST was not lower than QFT. This result is surprising in view of the widespread opinion that the BCG vaccine produces false positives in the TST[22]. Some authors have suggested that this effect is lower in tropical regions[23], and the type of BCG strain vaccinated may also have different effects on the TST [24].

The predictive values which depend on the prevalence and accuracy of the tests did not show significant differences according to the test applied, despite the TST being superior in all the values. Likewise, the results obtained suggest that both tests have a very high predictive capacity when the intensity and time of exposure are high and are therefore useful for diagnosis in the study of contacts. In the case of controls we observed very low positive predictive values thereby indicating the scarce utility of these tests in population screenings [2].

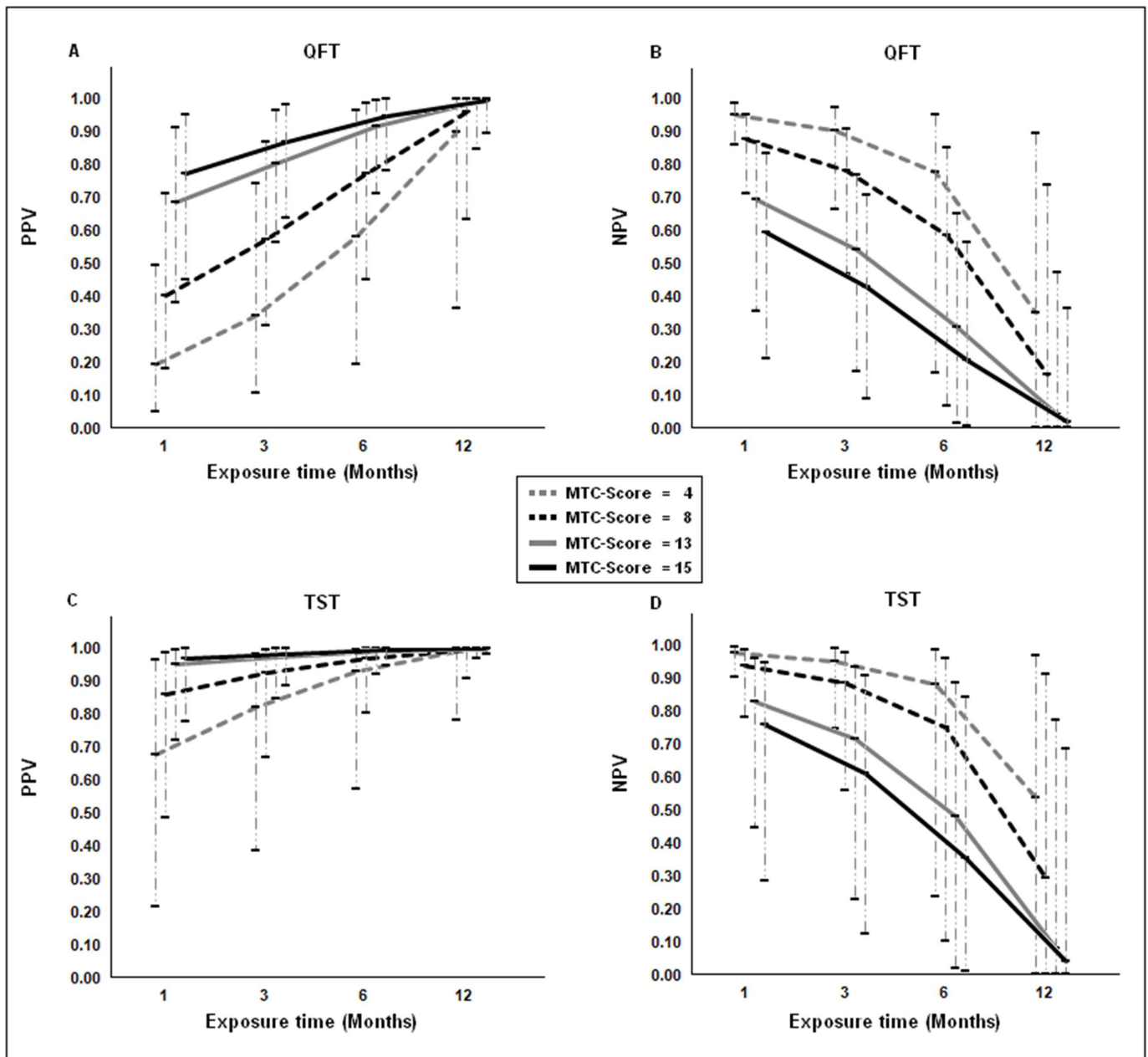


Fig 3. Positive and negative predictive values for QFT for times of exposure of 1, 3, 6, and 12 months and intensity of exposure of 4, 8, 13 and 15. TST: Tuberculin Skin Test, QFT: QuantiFERON-TB Gold In-Tube, PPV: positive predictive values, NPV: negative predictive values.

doi:10.1371/journal.pone.0164181.g003

On evaluating the results of the predictive values in the two tests combined it was found that these values in the two tests alone improved, albeit not significantly. According to the latent variable model, when the results between the tests were discordant, the TST presented a greater probability of correctly classifying the individual. Thus, in young children QFT results discordant with those of TST should be interpreted with caution since the predictive values fall markedly. This may explain, in part, the discordance observed between the two tests in short times of exposure similar to what has previously been described in this group [16].

The indeterminate results for QFT were high in our study. The proportion of indeterminate results varies greatly in studies in children [25–27] and may be associated with technical problems and probably also factors which alter Th1 immune response[28] such as age, the presence of helminths, and immunosuppression[29].

This study has several limitations. The relatively small sample size may have limited the power of the study to find significant differences. Nonetheless, there are few studies in small children and the results obtained are biologically logical. In addition, the sensitivity and specificity may vary widely among studies and this may affect the bayesian analysis. However, we have attempted to control this with the use of wide priors for the sensitivity and specificity of both tests. The third limitation was that we used the WHO definition of TST positivity (≥ 10 mm), and other cut off points might have shown different results.

On the other hand, the strengths of this study are: 1. There are few studies in small children who are the most vulnerable to *M. tuberculosis*. 2. The methodology used is novel in TB but has been well contrasted in other infectious diseases [30]. This study allows calculation of the prevalence of the disease which, to date, has not been done since LTBI cannot currently be microbiologically demonstrated due to its low bacillary load. Up to now both the prevalence and the accuracy of the tests has, in many cases, been calculated using an imperfect gold standard (one of the two tests) or has been extrapolated from ill children, who are clinically different from those who are infected. 3. The results obtained are biologically logical since the test results are correlated with the time and intensity of exposure.

Conclusions

This is one of the first studies to estimate the prevalence of LTBI in children and the parameters of its main diagnostic tests using a latent class model. The study of LTBI in young children is of great importance due to the high possibility of developing the disease and increasing the reservoir of cases which these cases may represent in the future if not treated[5,6]. The results of the present study suggest that children in contact with an index case have a high risk of infection according to the intensity and length of exposure to the index case, and this should be taken into account in the design of TB control strategies.

The TST and QFT showed similar characteristics and behavior but concordance was lower than expected. Combined use of the two tests in our study showed scarce improvement in the diagnosis of LTBI. We know that diagnostic tests of LTBI are not perfect. However, the latent variable model allows better evaluation of these tests and estimation of the prevalence of LTBI, thereby facilitating the development of more adequate approaches to this crucial problem of TB control. The diagnosis of LTBI and disease is the key to the control of pediatric TB and requires more and greater efforts to this end.

Author Contributions

Conceptualization: RA TPP HDP CA.

Formal analysis: RA HDP.

Funding acquisition: RA TPP HDP CA AM SB FME.

Investigation: TPP AM SB FME CA.

Methodology: RA TPP HDP CA AM SB FME.

Project administration: RA TPP.

Resources: RA TPP HDP CA AM SB FME.

Software: RA HDP CA.

Supervision: RA TPP.

Validation: RA TPP CA.

Writing – original draft: RA TPP HDP.

Writing – review & editing: RA TPP HDP CA AM SB FME.

References

1. WHO | Global tuberculosis report 2014. In: WHO [Internet]. [cited 21 Jan 2016]. Available: http://www.who.int/tb/publications/global_report/en/
2. WHO | Systematic screening for active tuberculosis: principles and recommendations. In: WHO [Internet]. [cited 21 Jan 2016]. Available: <http://www.who.int/tb/tbscreening/en/>
3. Marais BJ, Gie RP, Schaaf HS, Hesselning AC, Obihara CC, Starke JJ, et al. The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. *Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis*. 2004; 8: 392–402. PMID: [15141729](#)
4. Lewinsohn DA, Lewinsohn DM. Immunologic susceptibility of young children to *Mycobacterium tuberculosis*. *Pediatr Res*. 2008; 63: 115. doi: [10.1203/PDR.0b013e3181652085](#) PMID: [18209663](#)
5. Dodd PJ, Gardiner E, Coghlan R, Seddon JA. Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. *Lancet Glob Health*. 2014; 2: e453–459. doi: [10.1016/S2214-109X\(14\)70245-1](#) PMID: [25103518](#)
6. Starke JR. Pediatric tuberculosis: time for a new approach. *Tuberc Edinb Scotl*. 2003; 83: 208–212. doi: [10.1016/S1472-9792\(02\)00088-4](#) PMID: [12758213](#)
7. WHO | Guidance for national tuberculosis programmes on the management of tuberculosis in children. In: WHO [Internet]. [cited 21 Jan 2015]. Available: http://www.who.int/tb/publications/childtb_guidelines/en/
8. Pai M, Riley LW, Colford JM. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. *Lancet Infect Dis*. 2004; 4: 761–776. doi: [10.1016/S1473-3099\(04\)01206-X](#) PMID: [15567126](#)
9. Starke JR, Committee On Infectious Diseases. Interferon- γ release assays for diagnosis of tuberculosis infection and disease in children. *Pediatrics*. 2014; 134: e1763–1773. doi: [10.1542/peds.2014-2983](#) PMID: [25422024](#)
10. Sollai S, Galli L, de Martino M, Chiappini E. Systematic review and meta-analysis on the utility of Interferon-gamma release assays for the diagnosis of *Mycobacterium tuberculosis* infection in children: a 2013 update. *BMC Infect Dis*. 2014; 14 Suppl 1: S6. doi: [10.1186/1471-2334-14-S1-S6](#) PMID: [24564486](#)
11. 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: 263–272. PMID: [7840100](#)
12. Hesselning AC, Mandalakas AM, Kirchner HL, Chegou NN, Marais BJ, Stanley K, et al. Highly discordant T cell responses in individuals with recent exposure to household tuberculosis. *Thorax*. 2009; 64: 840–846. doi: [10.1136/thx.2007.085340](#) PMID: [18682523](#)
13. Martinez EZ, Louzada-Neto F, Derchain SFM, Achcar JA, Gontijo RC, Sarian LOZ, et al. Bayesian estimation of performance measures of cervical cancer screening tests in the presence of covariates and absence of a gold standard. *Cancer Inform*. 2008; 6: 33–46. PMID: [19259401](#)
14. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*. 2001; 57: 158–167. doi: [10.1111/j.0006-341X.2001.00158.x](#) PMID: [11252592](#)
15. Gelman A, Rubin DB. Markov chain Monte Carlo methods in biostatistics. *Stat Methods Med Res*. 1996; 5: 339–355. doi: [10.1177/096228029600500402](#) PMID: [9004377](#)
16. Pérez-Porcuna TM, Ascaso C, Malheiro A, Abellana R, Martins M, Sardinha JFJ, et al. *Mycobacterium tuberculosis* infection in young children: analyzing the performance of the diagnostic tests. *PLoS One*. 2014; 9: e97992. doi: [10.1371/journal.pone.0097992](#) PMID: [24879374](#)
17. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. *Eur Respir J*. 2013; 41: 140–156. doi: [10.1183/09031936.00070812](#) PMID: [22936710](#)

18. Cruz AT, Starke JR, Lobato MN. Old and new approaches to diagnosing and treating latent tuberculosis in children in low-incidence countries. *Curr Opin Pediatr*. 2014; 26: 106–113. doi: [10.1097/MOP.000000000000049](https://doi.org/10.1097/MOP.000000000000049) PMID: [24299911](https://pubmed.ncbi.nlm.nih.gov/24299911/)
19. Machingaidze S, Wiysonge CS, Gonzalez-Angulo Y, Hatherill M, Moyo S, Hanekom W, et al. The utility of an interferon gamma release assay for diagnosis of latent tuberculosis infection and disease in children: a systematic review and meta-analysis. *Pediatr Infect Dis J*. 2011; 30: 694–700. doi: [10.1097/INF.0b013e318214b915](https://doi.org/10.1097/INF.0b013e318214b915) PMID: [21427627](https://pubmed.ncbi.nlm.nih.gov/21427627/)
20. Thomas TA, Mondal D, Noor Z, Liu L, Alam M, Haque R, et al. Malnutrition and helminth infection affect performance of an interferon gamma-release assay. *Pediatrics*. 2010; 126: e1522–1529. doi: [10.1542/peds.2010-0885](https://doi.org/10.1542/peds.2010-0885) PMID: [21059723](https://pubmed.ncbi.nlm.nih.gov/21059723/)
21. Pai M, Joshi R, Dogra S, Zwerling AA, Gajalakshmi D, Goswami K, et al. T-cell assay conversions and reversions among household contacts of tuberculosis patients in rural India. *Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis*. 2009; 13: 84–92. PMID: [19105884](https://pubmed.ncbi.nlm.nih.gov/19105884/)
22. Seddon JA, Paton J, Nademi Z, Keane D, Williams B, Williams A, et al. The impact of BCG vaccination on tuberculin skin test responses in children is age dependent: evidence to be considered when screening children for tuberculosis infection. *Thorax*. 2016; 71: 932–939. doi: [10.1136/thoraxjnl-2015-207687](https://doi.org/10.1136/thoraxjnl-2015-207687) PMID: [27335104](https://pubmed.ncbi.nlm.nih.gov/27335104/)
23. Hill PC, Ota MOC. Tuberculosis case-contact research in endemic tropical settings: design, conduct, and relevance to other infectious diseases. *Lancet Infect Dis*. 2010; 10: 723–732. doi: [10.1016/S1473-3099\(10\)70164-X](https://doi.org/10.1016/S1473-3099(10)70164-X) PMID: [20883968](https://pubmed.ncbi.nlm.nih.gov/20883968/)
24. Ritz N, Dutta B, Donath S, Casalaz D, Connell TG, Tebruegge M, et al. The influence of bacille Calmette-Guerin vaccine strain on the immune response against tuberculosis: a randomized trial. *Am J Respir Crit Care Med*. 2012; 185: 213–222. doi: [10.1164/rccm.201104-0714OC](https://doi.org/10.1164/rccm.201104-0714OC) PMID: [22071384](https://pubmed.ncbi.nlm.nih.gov/22071384/)
25. Bergamini BM, Losi M, Vaianti F, D'Amico R, Meccugni B, Meacci M, et al. Performance of commercial blood tests for the diagnosis of latent tuberculosis infection in children and adolescents. *Pediatrics*. 2009; 123: e419–424. doi: [10.1542/peds.2008-1722](https://doi.org/10.1542/peds.2008-1722) PMID: [19254978](https://pubmed.ncbi.nlm.nih.gov/19254978/)
26. Connell TG, Ritz N, Paxton GA, BATTERY JP, Curtis N, Ranganathan SC. A three-way comparison of tuberculin skin testing, QuantiFERON-TB gold and T-SPOT.TB in children. *PLoS ONE*. 2008; 3: e2624. doi: [10.1371/journal.pone.0002624](https://doi.org/10.1371/journal.pone.0002624) PMID: [18612425](https://pubmed.ncbi.nlm.nih.gov/18612425/)
27. Hausteiner T, Ridout DA, Hartley JC, Thaker U, Shingadia D, Klein NJ, et al. The likelihood of an indeterminate test result from a whole-blood interferon-gamma release assay for the diagnosis of *Mycobacterium tuberculosis* infection in children correlates with age and immune status. *Pediatr Infect Dis J*. 2009; 28: 669–673. doi: [10.1097/INF.0b013e3181a16394](https://doi.org/10.1097/INF.0b013e3181a16394) PMID: [19633512](https://pubmed.ncbi.nlm.nih.gov/19633512/)
28. Grare M, Derelle J, Dailoux M, Laurain C. QuantiFERON-TB Gold In-Tube as help for the diagnosis of tuberculosis in a French pediatric hospital. *Diagn Microbiol Infect Dis*. 2010; 66: 366–372. doi: [10.1016/j.diagmicrobio.2009.11.002](https://doi.org/10.1016/j.diagmicrobio.2009.11.002) PMID: [20071130](https://pubmed.ncbi.nlm.nih.gov/20071130/)
29. Thomas TA, Mondal D, Noor Z, Liu L, Alam M, Haque R, et al. Malnutrition and helminth infection affect performance of an interferon gamma-release assay. *Pediatrics*. 2010; 126: e1522–1529. doi: [10.1542/peds.2010-0885](https://doi.org/10.1542/peds.2010-0885) PMID: [21059723](https://pubmed.ncbi.nlm.nih.gov/21059723/)
30. Pai M, Dendukuri N, Wang L, Joshi R, Kalantri S, Rieder HL. Improving the estimation of tuberculosis infection prevalence using T-cell-based assay and mixture models. *Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis*. 2008; 12: 895–902. PMID: [18647448](https://pubmed.ncbi.nlm.nih.gov/18647448/)

C Código WinBUGS para el modelo propuesto (SUB×TST)

```

1 # Modelo: dos efectos aleatorios, individuo y individuo-test
2 {
3   # Distribución prior para prevalencia, sensibilidad y
4     especificidad
5   prev~dbeta(alpha.prev,beta.prev)
6   a11~dnorm(A11,tau.A11)
7   a21~dnorm(A21,tau.A21)
8   a10~dnorm(A10,tau.A10)
9   a20~dnorm(A20,tau.A20)
10  b21~djl.dnorm.trunc(B21,tau.B21,-10,10)
11  b20~djl.dnorm.trunc(B20,tau.B20,-10,10)
12  b11~djl.dnorm.trunc(B11,tau.B11,-10,10)
13  b10~djl.dnorm.trunc(B10,tau.B10,-10,10)
14
15  # Definiendo la sensibilidad y la especificidad de cada individuo
16  # k = n° de individuos; j = n° de tests; r = n° de replicas
17  for (i in 1:k*j*r) {
18    sen1[i]<-phi(a11+b11*t.i[ind[i]]+b21*t.r[tr[i]])
19    one.minus.spec1[i]<-phi(-a10-b10*t.i[ind[i]]-b20*t.r[tr[i]])
20    spec1[i]<-1-one.minus.spec1[i]
21    sen2[i]<-phi(a21+b11*t.i[ind[i]]+b21*t.r[tr[i]])
22    one.minus.spec2[i]<-phi(-a20-b10*t.i[ind[i]]-b20*t.r[tr[i]])
23    spec2[i]<-1-one.minus.spec2[i]
24  }
25  # k individuos
26  for (i in 1:k) {t.i[i]~dnorm(0,1)}
27  # j tests
28  for (i in 1:k*j) {t.r[i]~dnorm(0,1)}
29
30  # Probabilidad condicional de una observación positiva
31  for (i in 1:k*j*r) {
32    pv[i,2,1]<-sen1[i] # sensibilidad del test 1
33    pv[i,1,1]<-1-spec1[i] # 1 - especificidad del test 1
34    pv[i,2,2]<-sen2[i] # sensibilidad del test 2
35    pv[i,1,2]<-1-spec2[i] # 1 - especificidad del test 2
36  }
37
38  # Relación entre datos observados e cond. probab. de positivo obs.
39  for (i in 1:k*j*r) {
40    D1[i]~dbern(prev)
41    D[i]<-D1[i]+1
42    resultat[i]~dbern(pv[i,D[i],Tst[i]])
43  }
44  S[1]<-phi(a11/sqrt(1+b11*b11+b21*b21))
45  S[2]<-phi(a21/sqrt(1+b11*b11+b21*b21))
46  C[1]<-phi(a10/sqrt(1+b10*b10+b20*b20))
47  C[2]<-phi(a20/sqrt(1+b10*b10+b20*b20))
48 }

```

D Código WinBUGS para el modelo de efectos fijos con estratificación de la población (dos prevalencias)

```

1 #LIKELIHOOD - 2 DEPENDENT TESTS
2 model {
3   ##### priors #####
4   us<- min(S[1],S[2])-(S[1]*S[2])
5   uc<- min(C[1],C[2])-(C[1]*C[2])
6   covs12 ~ dbeta(alpha.covS,beta.covS)
7   covc12 ~ dbeta(alpha.covC,beta.covC)
8   covS<- covs12*us
9   covC<- covc12*uc
10  for (t in 1:2)
11  {
12    p[t] ~ dbeta(1,1)
13    S[t] ~ dbeta(alpha.S[t],beta.S[t])
14    C[t] ~ dbeta(alpha.C[t],beta.C[t])
15
16    #predict values (+ and -)
17    for (st in 1:2){
18      PPV[st,t] <- p[st]*S[t] / (p[st]*S[t] + (1-p[st])*(1-C[t]))
19                # +
20      NPV[st,t] <- (1-p[st])*C[t] / (p[st]*(1-S[t]) + (1-p[st])*C[t])
21                # -
22    }
23
24    #combined predict positive values
25    PPVcov[1,1] <- p[1]*(S[1]*S[2]+covS) / (p[1]*(S[1]*S[2]+
26      covS) + (1-p[1])*((1-C[1])*(1-C[2])+covC)) # ++
27    PPVcov[1,2] <- p[1]*(S[1]*(1-S[2])-covS) / (p[1]*(S[1]*(1-S
28      [2])-covS) + (1-p[1])*((1-C[1])*C[2]-covC)) # +-
29    PPVcov[1,3] <- p[1]*((1-S[1])*S[2]-covS) / (p[1]*((1-S[1])*S
30      [2]-covS) + (1-p[1])*C[1]*(1-C[2]-covC)) # -+
31    PPVcov[1,4] <- p[1]*((1-S[1])*(1-S[2])+covS) / (p[1]*((1-S[1])*(1-
32      S[2])+covS) + (1-p[1])*C[1]*C[2]+covC)) # --
33
34    PPVcov[2,1] <- p[2]*(S[1]*S[2]+covS) / (p[2]*(S[1]*S[2]+
35      covS) + (1-p[2])*((1-C[1])*(1-C[2])+covC)) # ++
36    PPVcov[2,2] <- p[2]*(S[1]*(1-S[2])-covS) / (p[2]*(S[1]*(1-S
37      [2])-covS) + (1-p[2])*((1-C[1])*C[2]-covC)) # +-
38    PPVcov[2,3] <- p[2]*((1-S[1])*S[2]-covS) / (p[2]*((1-S[1])*S
39      [2]-covS) + (1-p[2])*C[1]*(1-C[2]-covC)) # -+
40    PPVcov[2,4] <- p[2]*((1-S[1])*(1-S[2])+covS) / (p[2]*((1-S[1])*(1-
41      S[2])+covS) + (1-p[2])*C[1]*C[2]+covC)) # --
42
43    #combined predict negative values
44    NPVcov[1,1] <- (1-p[1])*((1-C[1])*(1-C[2])+covC) / (p[1]*(S[1]*S
45      [2]+covS) + (1-p[1])*((1-C[1])*(1-C[2])+covC)) # ++
46    NPVcov[1,2] <- (1-p[1])*((1-C[1])*C[2]-covC) / (p[1]*(S[1]*(1-S
47      [2])-covS) + (1-p[1])*((1-C[1])*C[2]-covC)) # +-
48    NPVcov[1,3] <- (1-p[1])*C[1]*(1-C[2]-covC) / (p[1]*((1-S[1])*S
49      [2]-covS) + (1-p[1])*C[1]*(1-C[2]-covC)) # -+
50    NPVcov[1,4] <- (1-p[1])*C[1]*C[2]+covC) / (p[1]*((1-S[1])*
51      (1-S[2])+covS) + (1-p[1])*C[1]*C[2]+covC)) # --
52
53    NPVcov[2,1] <- (1-p[2])*((1-C[1])*(1-C[2])+covC) / (p[2]*(S[1]*S
54      [2]+covS) + (1-p[2])*((1-C[1])*(1-C[2])+covC)) # ++
55    NPVcov[2,2] <- (1-p[2])*((1-C[1])*C[2]-covC) / (p[2]*(S[1]*(1-S
56      [2])-covS) + (1-p[2])*((1-C[1])*C[2]-covC)) # +-
57    NPVcov[2,3] <- (1-p[2])*C[1]*(1-C[2]-covC) / (p[2]*((1-S[1])*S
58      [2]-covS) + (1-p[2])*C[1]*(1-C[2]-covC)) # -+
59    NPVcov[2,4] <- (1-p[2])*C[1]*C[2]+covC) / (p[2]*((1-S[1])*(1-
60      S[2])+covS) + (1-p[2])*C[1]*C[2]+covC)) # --
61
62  }
63 }

```

```

40 NPVcov[2,1] <- (1-p[2])*((1-C[1])*(1-C[2])+covC) / (p[2]*(S[1]*S
      [2]+covS) + (1-p[2])*((1-C[1])*(1-C[2])+covC)) # ++
41 NPVcov[2,2] <- (1-p[2])*((1-C[1])*C[2]-covC) / (p[2]*(S[1]*(1-S
      [2])-covS) + (1-p[2])*((1-C[1])*C[2]-covC)) # +-
42 NPVcov[2,3] <- (1-p[2])*C[1]*(1-C[2])-covC / (p[2]*((1-S[1])*
      S[2]-covS) + (1-p[2])*C[1]*(1-C[2])-covC) # -+
43 NPVcov[2,4] <- (1-p[2])*C[1]*C[2]+covC / (p[2]*((1-S[1])*
      (1-S[2])+covS) + (1-p[2])*C[1]*C[2]+covC) # --
44
45 ##### likelihood #####
46 for (i in 1:N){
47   for (l in 1:2){ #l=1 => testel
48     CPyx1[i,l] <- S[l]*equals(T[i,l],1) + (1-S[l])*equals(T[i,l]
      [l],0)
49     CPyx0[i,l] <- (1-C[l])*equals(T[i,l],1) + C[l]*equals(T[i,l]
      [l],0)
50     sig[i,l] <- pow(-1,equals(T[i,l],0)) #signos iguales -> +;
      signos diferentes -> -
51   }
52   L[i]<-p[CI[i]]*(CPyx1[i,1]*CPyx1[i,2]+(covS*sig[i,1]*sig[i,2])) +
      (1-p[CI[i]])*(CPyx0[i,1]*CPyx0[i,2]+(covC*sig[i,1]*sig[i,2]))
      ) #modelo de dependencia completa
53   ones[i] <- 1
54   ones[i] ~ dbern(L[i])
55 }
56 rhoS <- covS / sqrt(S[1]*(1-S[1])*S[2]*(1-S[2]))
57 rhoC <- covC / sqrt(C[1]*(1-C[1])*C[2]*(1-C[2]))
58 }

```

E Código WinBUGS para el modelo de efecto aleatorio individuo incluyendo covariables en la prevalencia

```

1 # Sensibilidad y Especificidad - Efecto Aleatorio (probit)
2 model{
3 #=====
4 # Prior distributions of prevalence, sensitivities and specificities
5 #=====
6 b1~djl.dnorm.trunc(B1,tau.B1,-1,2.2)
7 b0~djl.dnorm.trunc(B0,tau.B0,-1,2)
8 alpha0~djl.dnorm.trunc(MeanP1,TauP1,-3.7,-1) #prevalencia en t=0 (p
   [1] de sim.dep)
9 beta01~djl.dnorm.trunc(0,0.1,-0.2,0.7) #tiempo
10 beta02~djl.dnorm.trunc(0,0.1,-0.05,0.35) #STB
11 a11~djl.dnorm.trunc(A11,tau.A11,-0.7,1)
12 a21~djl.dnorm.trunc(A21,tau.A21,-0.5,2.5)
13 a10~djl.dnorm.trunc(A10,tau.A10,0.3,2)
14 a20~djl.dnorm.trunc(A20,tau.A20,1,4)
15 #=====
16 # Defining the sensitivity and specificity of each subject
17 #=====
18 for (i in 1:N) {
19   t.i[i]~dnorm(0,1)
20   prev[i] <-phi(alpha0 + beta01*tiempo[i] + beta02*STB[i])#
   prevalencia
21   sen1[i]<-phi(a11 + b1*t.i[i])
22   one.minus.spec1[i]<-phi(-a10 -b0*t.i[i]) # 1 - Especificidad
   Quantiferon
23   spec1[i]<-1-one.minus.spec1[i]
24   sen2[i]<-phi(a21 + b1*t.i[i])
25   one.minus.spec2[i]<-phi(-a20 -b0*t.i[i])
26   spec2[i]<-1-one.minus.spec2[i]
27 }
28 #=====
29 # Conditional probability of a positive observation
30 #=====
31 for (i in 1:N) {
32   pv1[i,2]<-sen1[i] # sensibilidad del test 1
33   pv1[i,1]<-1-spec1[i] # 1 - especificidad del test 1
34   pv2[i,2]<-sen2[i] # sensibilidad del test 2
35   pv2[i,1]<-1-spec2[i] # 1 - especificidad del test 2
36 }
37 #=====
38 # Relation between observed data and conditional probability of a
   positive observation
39 #=====
40 for (i in 1:N){
41   D1[i]~dbern(prev[i])
42   D[i]<-D1[i]+1
43   res1[i]~dbern(pv1[i,D[i]])
44   res2[i]~dbern(pv2[i,D[i]])
45 }
46 # PREVALENCIA en el tiempo t=0 y score s=0; t=1,3,6,12 en s=4,8,13,15
47 p00 <- phi(alpha0) #prevalencia no tempo 0
48 for (t in 1:4){
49   for (s in 1:4){

```

```

50   p[t,s]<-phi(alpha0 + beta01*qTmp[t] + beta02*qSTB[s])
51   }
52 }
53
54 # vPP vPN unidimensional
55 for (t in 1:4){
56   for (s in 1:4){
57     vPP1[t,s]<-(p[t,s]*S[1])/(p[t,s]*S[1] + (1-p[t,s])*(1-C[1]))
58     vPN1[t,s]<-((1-p[t,s])*C[1])/(p[t,s]*(1-S[1]) + (1-p[t,s])*C[1])
59     vPP2[t,s]<-(p[t,s]*S[2])/(p[t,s]*S[2] + (1-p[t,s])*(1-C[2]))
60     vPN2[t,s]<-((1-p[t,s])*C[2])/(p[t,s]*(1-S[2]) + (1-p[t,s])*C[2])
61   }
62 }
63
64 # vPP, vPN cuando score=0
65 PPV00[1] <- p00*(S[1]*S[2]) / (p00*(S[1]*S[2]) + (1-p00)*
      ((1-C[1])*(1-C[2]))) # ++
66 PPV00[2] <- p00*(S[1]*(1-S[2])) / (p00*(S[1]*(1-S[2])) + (1-p00)
      *((1-C[1])*C[2])) # +-
67 PPV00[3] <- p00*((1-S[1])*S[2]) / (p00*((1-S[1])*S[2]) + (1-p00)
      *(C[1]*(1-C[2]))) # -+
68 PPV00[4] <- p00*((1-S[1])*(1-S[2])) / (p00*((1-S[1])*(1-S[2])) + (1-
      p00)*(C[1]*C[2])) # --
69 NPV00[1] <- (1-p00)*(C[1]*C[2]) / (p00*((1-S[1])*(1-S[2])) +
      (1-p00)*(C[1]*C[2])) # ++
70 NPV00[2] <- (1-p00)*(C[1]*(1-C[2])) / (p00*((1-S[1])*S[2]) + (1-
      p00)*(C[1]*(1-C[2]))) # +-
71 NPV00[3] <- (1-p00)*((1-C[1])*C[2]) / (p00*(S[1]*(1-S[2])) + (1-
      p00)*((1-C[1])*C[2])) # -+
72 NPV00[4] <- (1-p00)*((1-C[1])*(1-C[2])) / (p00*(S[1]*S[2]) + (1-p00)*
      ((1-C[1])*(1-C[2]))) # --
73
74 # vPP, vPN cuando score>0 en distintos puntos de score y tiempo
75 for (t in 1:4){
76   for (s in 1:4){
77     PPVts[1,t,s] <- p[t,s]*(S[1]*S[2]) / (p[t,s]*(S[1]*S[2])
      + (1-p[t,s])*((1-C[1])*(1-C[2]))) # ++
78     PPVts[2,t,s] <- p[t,s]*(S[1]*(1-S[2])) / (p[t,s]*(S[1]*(1-S
      [2])) + (1-p[t,s])*((1-C[1])*C[2])) # +-
79     PPVts[3,t,s] <- p[t,s]*((1-S[1])*S[2]) / (p[t,s]*((1-S[1])*S
      [2]) + (1-p[t,s])*(C[1]*(1-C[2]))) # -+
80     PPVts[4,t,s] <- p[t,s]*((1-S[1])*(1-S[2])) / (p[t,s]*((1-S[1])*
      (1-S[2])) + (1-p[t,s])*(C[1]*C[2])) # --
81     NPVts[1,t,s] <- (1-p[t,s])*(C[1]*C[2]) / (p[t,s]*((1-S
      [1])*(1-S[2])) + (1-p[t,s])*(C[1]*C[2])) # ++
82     NPVts[2,t,s] <- (1-p[t,s])*(C[1]*(1-C[2])) / (p[t,s]*((1-S
      [1])*S[2]) + (1-p[t,s])*(C[1]*(1-C[2]))) # +-
83     NPVts[3,t,s] <- (1-p[t,s])*((1-C[1])*C[2]) / (p[t,s]*(S[1]*
      (1-S[2])) + (1-p[t,s])*((1-C[1])*C[2])) # -+
84     NPVts[4,t,s] <- (1-p[t,s])*((1-C[1])*(1-C[2])) / (p[t,s]*(S[1]*S
      [2]) + (1-p[t,s])*((1-C[1])*(1-C[2]))) # --
85   }
86 }
87
88 P<-mean(prev[])
89 S[1]<-phi(a11/sqrt(1+b1*b1))
90 S[2]<-phi(a21/sqrt(1+b1*b1))
91 C[1]<-phi(a10/sqrt(1+b0*b0))

```

```
92 C[2] <- phi(a20/sqrt(1+b0*b0))
93 Sen[1] <- mean(sen1[])
94 Sen[2] <- mean(sen2[])
95 Spec[1] <- mean(spec1[])
96 Spec[2] <- mean(spec2[])
97 }
```

F Funciones R utilizadas

```

1 #####
2 # GENERA LOS PARAMETROS BETA (IC 95%) - parBeta(li,ls)
3 #####
4
5 parBeta<-function(li,ls) {
6   a <- -(1/2)*(5*li^3+11*li^2*ls-8*li^2+11*li*ls^2-16*li*ls+5*ls^3-8*
7     ls^2)/(li^2-2*li*ls+ls^2)
8   b <- a*(2-li-ls)/(li+ls)
9   return(round(cbind(a,b),6))
10 }
11 #####
12 # CALCULA COVARIANZAS
13 #####
14
15 parCovSC<-function(nsim=100000,b1=0.668,b1r=0,b0=0.861,b0r=0,S1=0.25,
16   S2=0.80,C1=0.95,C2=0.70)
17 {
18   set.seed(1)
19   x<-rnorm(nsim,0,1)
20   y<-rnorm(nsim,0,1)
21   a11<-qnorm(S1)*sqrt(1+b1^2+b1r^2)
22   a21<-qnorm(S2)*sqrt(1+b1^2+b1r^2)
23   a10<-qnorm(C1)*sqrt(1+b0^2+b0r^2)
24   a20<-qnorm(C2)*sqrt(1+b0^2+b0r^2)
25   covS<-mean(pnorm(a11+b1*x+b1r*y)*pnorm(a21+b1*x+b1r*y))-
26     S1*S2
27   covC<-mean(pnorm(a10+b0*x+b0r*y)*pnorm(a20+b0*x+b0r*y))-
28     C1*C2
29   parS<-data.frame("a11"=a11,"a21"=a21,"b1"=b1,"b1r"=b1r,"CovS"=covS)
30   parC<-data.frame("a10"=a10,"a20"=a20,"b0"=b0,"b0r"=b0r,"CovC"=covC)
31   parSC<-cbind(parS,parC)
32   return(parSC)
33 }
34 #####
35 # CALCULA VPC - (b1,b1r,b0,b0r son las sds)
36 #####
37
38 VPCprobit<-function(Nsim,M,m,a,sigd,sigu) {
39   VPC1<-VPC2<-VPC3<-VPC12<-VPC13<-VPC23<-0
40   VPCBr1<-VPCBr2<-VPCBr3<-VPCRep<-0
41   for (nsim in 1:Nsim) {
42     print(paste("Nsim = ",nsim,"/",Nsim,sep=""))
43     set.seed(nsim)
44     uk<-rnorm(M,mean=0,sd=sigd)
45     pijk<-matrix(NA,M,m)
46     for (i in 1:M){
47       set.seed(i*123456)
48       vk<-rnorm(m,mean=0,sd=sigu)
49       pijk[i,] <- pnorm(a+uk[i]+vk) # probit
50 #       pijk[i,] <- 1/(1+exp(-(a+uk[i]+vk))) # logit
51     }
52     rm(uk,vk)
53     #r = 1...T, T = M x n, M = higher level, m = over dispersion (

```



```

        individuo)
52  var1<-pijk*(1-pijk)
53  v1<-mean(var1)          #nivel 1 (Error)
54  rm(var1)
55  pijkR<-rowMeans(pijk)
56  v2<-var(c(pijk))       #nivel 2 (Replica + Individuo)
57  rm(pijk)
58  v3<-var(pijkR)        #nivel 3 (Individuo)
59  rm(pijkR)
60  var2_3 <- v2-v3 # Efecto Replica (sin efecto individuo)
61  VPCBr3[nsim]<-(v3)/(v2+v1) # VPC Individuo
62  VPCRep[nsim]<-(v2-v3)/(v2+v1) # VPC Replica
63  piMarginal <- (pnorm(a/(sqrt((sigu^2) + (sigd^2) + 1))))
64  varMarginal <- piMarginal*(1-piMarginal)
65  varTotal <- v1+v2
66  }
67  VPCBrow1<-mean(VPCBr1); sdBrow1<-sd(VPCBr1)/sqrt(nsim)
68  VPCBrow2<-mean(VPCBr2); sdBrow2<-sd(VPCBr2)/sqrt(nsim)
69  VPCBrow3<-mean(VPCBr3); sdBrow3<-sd(VPCBr3)/sqrt(nsim)
70  VPCREP <-mean(VPCRep); sdRep <-sd(VPCRep)/sqrt(nsim)
71  res<-list(
72      "var1"=v1, "var2"=v2, "var3"=v3, "var2_3"=var2_3,
73      "VPCBrow3"=VPCBrow3,
74      "varMarginal"=varMarginal, "varTotal"=varTotal, "VPCrep"=
          VPCREP)
75  return(res)
76 }
77
78 #####
79 # APROXIMA LA INTEGRAL - (bisectional search algorithm).
80 #####
81
82 parIndRep<-function(nsim=100000,b1r=0,b0r=0,S1=0.25,S2=0.80,C1=0.95,
      C2=0.70,covS=0.025,covC=0.0175)
83 {
84  a11<-a21<-b1<-a10<-a20<-b0<-res1<-res0<-0
85  set.seed(1)
86  x<-rnorm(nsim,0,1)
87  y<-rnorm(nsim,0,1)
88  for (i in 0:5000){
89    b1[i]<-i/1000
90    b0[i]<-i/1000
91    a11[i]<-qnorm(S1)*sqrt(1+b1[i]^2 + b1r^2)
92    a21[i]<-qnorm(S2)*sqrt(1+b1[i]^2 + b1r^2)
93    a10[i]<-qnorm(C1)*sqrt(1+b0[i]^2 + b0r^2)
94    a20[i]<-qnorm(C2)*sqrt(1+b0[i]^2 + b0r^2)
95    res1[i]<-mean(pnorm(a11[i]+b1[i]*x+b1r*y)*pnorm(a21[i]+b1[i]*x+
      b1r*y)) - S1*S2 - covS
96    res0[i]<-mean(pnorm(a10[i]+b0[i]*x+b0r*y)*pnorm(a20[i]+b0[i]*x+
      b0r*y)) - C1*C2 - covC
97  }
98  Res1<-data.frame("a11"=a11,"a21"=a21,"b1"=b1,"Res1"=res1)
99  parS<-data.frame(Res1[abs(Res1$Res1)==min(abs(Res1$Res1))],)
100 Res0<-data.frame("a10"=a10,"a20"=a20,"b0"=b0,"Res0"=res0)
101 parC<-data.frame(Res0[abs(Res0$Res0)==min(abs(Res0$Res0))],)
102 parSC<-cbind(parS,parC)
103 return(parSC)
104 }

```

```

105
106 #####
107 # GENERA DATOS DE LAS SIMULACIONES DEL ARTICULO I
108 #####
109
110 simRNDrep<-function(n,num.rep=3,k=2,a10,a11,a20,a21,b1,b0,b1r,b0r,
    prev) {
111   X<-matrix(0,num.rep*k*n,n)
112   pos<-0
113   for (i in 1:n){
114     pos[i]<-num.rep*k*i
115     if (i==1) X[1:pos[1],i] <- 1
116     if (i>=2) X[(pos[i-1]+1):pos[i],i] <- 1
117   }
118
119   Y<-matrix(0,n*k*num.rep,n*k)
120   j=0
121   for(l in 1:n){
122     if(l==1) j<-rep(c(1,2),num.rep)
123     if(l!=1) j<-j+2
124     for(i in 1:(k*num.rep)){
125       Y[i+(l-1)*(k*num.rep),j[i]]<-1
126     }
127   }
128
129   ti<-rnorm(n,0,1)
130   tr<-rnorm(n*k,0,1)
131   s1<-pnorm(a11 + b1*X%*%ti + b1r*Y%*%tr)
132   c1<-pnorm(a10 + b0*X%*%ti + b0r*Y%*%tr)
133   s2<-pnorm(a21 + b1*X%*%ti + b1r*Y%*%tr)
134   c2<-pnorm(a20 + b0*X%*%ti + b0r*Y%*%tr)
135   p11<-prev*(s1*s2)+(1-prev)*((1-c1)*(1-c2))
136   p10<-prev*(s1*(1-s2))+(1-prev)*((1-c1)*c2)
137   p01<-prev*((1-s1)*s2)+(1-prev)*(c1*(1-c2))
138   p00<-prev*((1-s1)*(1-s2))+(1-prev)*(c1*c2)
139
140   Ind <-array(NA,dim=c(2*n*num.rep,1))
141   Tst<-array(NA,dim=c(2*n*num.rep,1))
142   Rep <-array(NA,dim=c(2*n*num.rep,1))
143   dd <-array(NA,dim=c(2*n*num.rep,1))
144
145   v<-1
146   for (j in 1:n){
147     for (i in 1:num.rep){
148       resultat<-rmultinom(1, 1, prob=c(p11[v],p10[v],p01[v],p00[v]))
149       m<-which(resultat==1)
150       Ind[v,1]<-j
151       Rep[v,1]<-i
152       Tst[v,1]<-1
153       if (m==1) dd[v,1]<-1
154       if (m==2) dd[v,1]<-1
155       if (m==3) dd[v,1]<-0
156       if (m==4) dd[v,1]<-0
157       v<-v+1
158       Ind[v,1]<-j
159       Rep[v,1]<-i
160       Tst[v,1]<-2
161       if (m==1) dd[v,1]<-1

```

```
162     if (m==2) dd[v,1] <-0
163     if (m==3) dd[v,1] <-1
164     if (m==4) dd[v,1] <-0
165     v<-v+1
166   }
167 }
168
169 dat<-data.frame ("Res"=dd, "Rep"=Rep, "Tst"=Tst, "Ind"=Ind)
170 o<-order (dat$Ind, dat$Tst, dat$Rep)
171 dat<-data.frame (dat[o, ])
172 return (dat)
173
174 }
```

Bibliografía

- Achcar, J. A., Martinez, E. Z., & Louzada-Neto, F. (2005). Binary data in the presence of covariates and misclassifications: A bayesian approach. *Brazilian Journal of Probability and Statistics*, 19, 65–84. 67
- Albert, P. S. (2009). Estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard. *Statistics in Medicine*, 28(5), 780–797. 92
- Bayes, M. & Price, M. (1763). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s. *Philosophical Transactions*, 53, 370–418. 33
- Bergamini, B. M., Losi, M., Vaienti, F., D’Amico, R., Meccugni, B., Meacci, M., De Giovanni, D., Rumpianesi, F., Fabbri, L. M., Balli, F., & Richeldi, L. (2009). Performance of commercial blood tests for the diagnosis of latent tuberculosis infection in children and adolescents. *Pediatrics*, 123(3), e419–e424. 108
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media. 34
- Bossuyt, P. M. M. (2008). Interpreting diagnostic test accuracy studies. *Seminars in Hematology*, 45(3), 189–195. 29, 30
- Box, G. E. & Tiao, G. C. (2011). *Bayesian inference in statistical analysis*, volume 40. John Wiley & Sons. 32
- Brooks, S. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455. 39
- Browne, W. J., Subramanian, S. V., Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(3), 599–613. 78
- Chib, S. & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4), 327–335. 35, 37
- Connell, T. G., Ritz, N., Paxton, G. A., Buttery, J. P., Curtis, N., & Ranganathan, S. C. (2008). A three-way comparison of tuberculin skin testing, quantiferon-tb gold and t-spot. tb in children. *Plos One*, 3(7). 109

- Cruz, A. T., Starke, J. R., & Lobato, M. N. (2014). Old and new approaches to diagnosing and treating latent tuberculosis in children in low-incidence countries. *Current Opinion in Pediatrics*, 26(1), 106–113. 107
- de Araujo Pereira, G., Louzada-Neto, F., de Fátima Barbosa, V., Ferreira-Silva, M. M., & Moraes-Souza, H. (2012). A general latent class model for performance evaluation of diagnostic tests in the absence of a gold standard: an application to chagas disease. *Computational and Mathematical Methods in Medicine*, 2012. 68
- Dendukuri, N., Hadgu, A., & Wang, L. (2009). Modeling conditional dependence between diagnostic tests: A multiple latent variable model. *Statistics in Medicine*, 28(3), 441–461. 92, 93
- Dendukuri, N. & Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*, 57(1), 158–167. 31, 55, 61, 71, 72, 81, 92, 96, 99, 100
- Dodd, P. J., Gardiner, E., Coghlan, R., & Seddon, J. A. (2014). Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. *Lancet Global Health*, 2(8), E453–E459. 95, 107
- Engel, B., Swildens, B., Stegeman, A., Buist, W., & de Jong, M. (2006). Estimation of sensitivity and specificity of three conditionally dependent diagnostic tests in the absence of a gold standard. *Journal of Agricultural Biological and Environmental Statistics*, 11(4), 360–380. 54
- Espeland, M. A., Murphy, W. C., & Leverett, D. H. (1988). Assessing diagnostic reliability and estimating incidence rates associated with a strictly progressive disease: Dental caries. *Statistics in Medicine*, 7(3), 403–416. 92
- Fox, G. J., Barry, S. E., Britton, W. J., & Marks, G. B. (2013). Contact investigation for tuberculosis: a systematic review and meta-analysis. *European Respiratory Journal*, 41(1), 140–156. 107
- Fryback, D. (1978). Bayes' theorem and conditional non-independence of data in medical diagnosis. *Computers and Biomedical Research*, 11(5), 423–434. 55
- Gelfand, A. E., Hills, S. E., Poon, A. R., & Smith, A. F. M. (1990). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412), 972–985. 35
- Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to computing marginal densities. *Journal of the American Statistical Association*, 85(410), 398–409. 35
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Taylor & Francis. 34
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. 36, 39
- Gelman, A. & Rubin, D. B. (1996). Markov chain monte carlo methods in biostatistics. *Statistical Methods in Medical Research*, 5(4), 339–355. 101

- Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-6*(6), 721–741. 35, 49
- Georgiadis, M. P., Johnson, W. O., Gardner, I. A., & Singh, R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1), 63–76. 93
- Goldstein, H. (2011). *Multilevel statistical models*, volume 922. John Wiley & Sons. 78
- Gonçalves, A. Q., Abellana, R., da Silva, H. D. P., Santos, I., Serra, P. T., ao, G. R. J., Orlandi, P. P., & Ascaso, C. (2014). Comparison of the performance of two spontaneous sedimentation techniques for the diagnosis of human intestinal parasites in the absence of a gold standard. *Acta Tropica*, 131(0), 63 – 70. 76
- Grare, M., Derelle, J., Dailloux, M., & Laurain, C. (2010). Quantiferon (r)-tb gold in-tube as help for the diagnosis of tuberculosis in a french pediatric hospital. *Diagnostic Microbiology and Infectious Disease*, 66(4), 366–372. 109
- Gustafson, P. (2005a). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Statistical Science*, 20(2), 111–129. 93
- Gustafson, P. (2005b). The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard. *Statistics in Medicine*, 24(8), 1203–1217. 93
- Hadgu, A. & Qu, Y. (1998). A biomedical application of latent class models with random effects. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 47(4), pp. 603–616. 54, 55
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1), 97–109. 35, 37
- Haustein, T., Ridout, D. A., Hartley, J. C., Thaker, U., Shingadia, D., Klein, N. J., Novelli, V., & Dixon, G. L. J. (2009). The likelihood of an indeterminate test result from a whole-blood interferon-gamma release assay for the diagnosis of mycobacterium tuberculosis infection in children correlates with age and immune status. *Pediatric Infectious Disease Journal*, 28(8), 669–673. 109
- Hesseling, A. C., Mandalakas, A. M., Kirchner, H. L., Chegou, N. N., Marais, B. J., Stanley, K., Zhu, X., Black, G., Beyers, N., & Walzl, G. (2009). Highly discordant t cell responses in individuals with recent exposure to household tuberculosis. *Thorax*, 64(10), 840–846. 98
- Hill, P. C. & Ota, M. O. C. (2010). Tuberculosis case-contact research in endemic tropical settings: design, conduct, and relevance to other infectious diseases. *Lancet Infectious Diseases*, 10(10), 723–732. 108
- Hoffman-Jørgensen, J. (1994a). *Probability with a view towards statistics*, volume 1. CRC Press. 33
- Hoffman-Jørgensen, J. (1994b). *Probability with a view towards statistics*, volume 2. CRC Press. 33

- Hui, S. L. & Walter, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, 36(1), pp. 167–171. 68
- Hui, S. L. & Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research*, 7(4), 354–370. 92
- Joseph, L., Gyorkos, T. W., & Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, 141(3), 263–272. 31, 43, 44, 92, 94, 96
- Kaldor, J. & Clayton, D. (1985). Latent class analysis in chronic disease epidemiology. *Statistics in Medicine*, 4(3), 327–335. 43
- Lazarsfeld, P. F. (1950). The logical and mathematical foundations of latent structure analysis. *In Studies in Social Psychology in World War II (Vol. IV: Measurement and Prediction)*, S. A., (pp. 362–412). 43
- Lewinsohn, D. A. & Lewinsohn, D. M. (2008). Immunologic susceptibility of young children to mycobacterium tuberculosis. *Pediatric Research*, 63(2), 115. 95
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs - a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337. 38, 74
- Machingaidze, S., Wiysonge, C. S., Gonzalez-Angulo, Y., Hatherill, M., Moyo, S., Hanekom, W., & Mahomed, H. (2011). The utility of an interferon gamma release assay for diagnosis of latent tuberculosis infection and disease in children a systematic review and meta-analysis. *Pediatric Infectious Disease Journal*, 30(8), 694–700. 108
- Marais, B., Gie, R., Schaaf, H., Hesselning, A., Obihara, C., Starke, J., Enarson, D., Donald, P., & Beyers, N. (2004). The natural history of childhood intra-thoracic tuberculosis: a critical review of literature from the pre-chemotherapy era. *International Journal of Tuberculosis and Lung Disease*, 8(4), 392–402. 95
- Martinez, E. Z., Achcar, J. A., & Louzada-Neto, F. (2005). Bayesian estimation of diagnostic tests accuracy for semi-latent data with covariates. *Journal of Biopharmaceutical Statistics*, 15(5), 809–821. 31, 67
- Martinez, E. Z., Louzada-Neto, F., Achcar, J. A., Syrjanen, K. J., Derchain, S. F. M., Gontijo, R. C., & Sarian, L. O. Z. (2009). Bayesian estimation of performance measures of screening tests in the presence of covariates and absence of a gold standard. *Brazilian Journal of Probability and Statistics*, 23(1), 68–81. 67
- Martinez, E. Z., Louzada-Neto, F., Derchain, S. F. M., Achcar, J. A., Gontijo, R. C., Sarian, L. O. Z., & Syrjanen, K. J. (2008). Bayesian estimation of performance measures of cervical cancer screening tests in the presence of covariates and absence of a gold standard. *Cancer Informatics*, 6, 33. 31, 67, 92, 99
- McCutcheon, A. L. & Hagenars, J. A. (2002). *Applied latent class analysis*. Cambridge etc.: Cambridge University Press. 43

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6), 1087–1092. 35
- Moons, K., Biesheuvel, C., & Grobbee, D. (2004). Test research versus diagnostic research. *Clinical Chemistry*, 50(3), 473–476. 29
- Moons, K., van Es, G., Michel, B., Buller, H., Habbema, J., & Grobbee, D. (1999). Redundancy of single diagnostic test evaluation. *Epidemiology*, 10(3), 276–281. 29
- Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS*, volume 698. John Wiley & Sons. 32, 35
- Organization, W. H. et al. (2014a). *Guidance for national tuberculosis programmes on the management of tuberculosis in children*. World Health Organization. 96, 98
- Organization, W. H. et al. (2014b). *Systematic screening for active tuberculosis: principles and recommendations*. World Health Organization. 95, 108
- Organization, W. H. et al. (2015). *Global tuberculosis report 2015*. Technical report. 95
- Pai, M., Joshi, R., Dogra, S., Zwerling, A. A., Gajalakshmi, D., Goswami, K., Reddy, M. V. R., Kalantri, S., Hill, P. C., Menzies, D., & Hopewell, P. C. (2009). T-cell assay conversions and reversions among household contacts of tuberculosis patients in rural india. *International Journal of Tuberculosis and Lung Disease*, 13(1), 84–92. 108
- Pai, M., Riley, L., & Colford, J. (2004). Interferon- γ assays in the immunodiagnosis of tuberculosis: a systematic review. *Lancet Infectious Diseases*, 4(12), 761–776. 96
- Perez-Porcuna, T. S. M., Ascaso, C., Malheiro, A., Abellana, R., Martins, M., Jardim Sardinha, J. F., Quinco, P., Antunes, I. A., Garrido, M. d. S., Buehrer-Sekula, S., & Martinez-Espinosa, F. E. (2014). Mycobacterium tuberculosis infection in young children: Analyzing the performance of the diagnostic tests. *Plos One*, 9(5). 107, 108
- Qu, Y., Tan, M., & Kutner, M. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52(3), 797–810. 31, 54, 55, 61
- R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 63, 74
- Reitsma, J. B., Rutjes, A. W. S., Khan, K. S., Coomarasamy, A., & Bossuyt, P. M. (2009). A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *Journal of Clinical Epidemiology*, 62(8), 797–806. 27
- Ritz, N., Dutta, B., Donath, S., Casalaz, D., Connell, T. G., Tebruegge, M., Robins-Browne, R., Hanekom, W. A., Britton, W. J., & Curtis, N. (2012). The influence of bacille calmette-guerin vaccine strain on the immune response against tuberculosis a randomized trial. *American Journal of Respiratory and Critical Care Medicine*, 185(2), 213–222. 108
- Rujes, A. W. S., Reitsma, J. B., Coomarasamy, A., Khan, K. S., & Bossuyt, P. M. M. (2007). Evaluation of diagnostic test when there is no gold standard. a review of methods. *Health Technology Assessment*, 11(50), 1–+. 27, 30

- Sollai, S., Galli, L., de Martino, M., & Chiappini, E. (2014). Systematic review and meta-analysis on the utility of interferon-gamma release assays for the diagnosis of mycobacterium tuberculosis infection in children: a 2013 update. *BMC Infectious Diseases*, 14(1). 96
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). Winbugs user manual. 81
- Starke, J. (2003). Pediatric tuberculosis: time for a new approach. *Tuberculosis*, 83(1-3), 208–212. 4th World Congress on Tuberculosis, WASHINGTON, D.C., JUN 03-05, 2002. 95
- Starke, J. R., Byington, C. L., Maldonado, Y. A., Barnett, E. D., Davies, H. D., Edwards, K. M., Jackson, M. A., Murray, D. L., Rathore, M. H., Sawyer, M. H., et al. (2014). Interferon- γ release assays for diagnosis of tuberculosis infection and disease in children. *Pediatrics*, 134(6), e1763–e1773. 96
- Sturtz, S., Ligges, U., & Gelman, A. E. (2005). R2winbugs: a package for running winbugs from r. *Journal of Statistical Software*, 12(3), 1–16. 74
- Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), pp. 528–540. 35, 43
- Thisted, R. A. (1988). *Elements of statistical computing: numerical computation*, volume 1. CRC Press. 63, 77
- Thomas, T. A., Mondal, D., Noor, Z., Liu, L., Alam, M., Haque, R., Banu, S., Sun, H., & Peterson, K. M. (2010). Malnutrition and helminth infection affect performance of an interferon γ -release assay. *Pediatrics*, 126(6), E1522–E1529. 108, 109
- Toft, N., Jørgensen, E., & Højsgaard, S. (2005). Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine*, 68(1), 19 – 33. 93
- Torrance-Rynard, V. & Walter, S. (1997). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine*, 16(19), 2157–2175. 31
- Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*, 41(4), 959–968. 31, 55, 58
- Valenstein, P. (1990). Evaluating diagnostic tests with imperfect standards. *American Journal of Clinical Pathology*, 93(2), 252–258. 27
- van Smeden, M., Naaktgeboren, C. A., Reitsma, J. B., Moons, K. G. M., & de Groot, Joris, A. (2014). Latent class models in diagnostic studies when there is no reference standard - a systematic review. *American Journal of Epidemiology*, 179(4), 423–431. 54, 93
- Walter, S. & Irwig, L. (1988). Estimation of test error rates, disease prevalence and relative risk from misclassified data: A review. *Journal of Clinical Epidemiology*, 41(9), 923–937. 43
- Walter, S., Macaskill, P., Lord, S. J., & Irwig, L. (2012). Effect of dependent errors in the assessment of diagnostic or screening test accuracy when the reference standard is imperfect. *Statistics in Medicine*, 31(11-12), 1129–1138. 31
- Whiting, P., Rutjes, A., Reitsma, J., Glas, A., Bossuyt, P., & Kleijnen, J. (2004). Sources of variation and bias in studies of diagnostic accuracy - a systematic review. *Annals of Internal Medicine*, 140(3), 189–202. 30

- Yang, I. & Becker, M. P. (1997). Latent variable modeling of diagnostic accuracy. *Biometrics*, 53(3), pp. 948–958. [54](#)
- Zhang, B., Chen, Z., & Albert, P. S. (2012). Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. *Biometrics*, 68(4), 1294–1302. [92](#)