



TRABAJO FIN DE MÁSTER EN BIOESTADÍSTICA

**Estimación Bayesiana de la prevalencia  
de Estrongiloidiasis en población infantil  
de Cubal (Angola) y de los parámetros de  
los tests diagnósticos en ausencia de *gold  
standard***

Septiembre 2017

Elena Dacal Picazo

**Tutoras:**

Rosario Susi García

Carmen Nieto Zayas

## Índice

1. Introducción .....	2
1.1. Ciclo biológico del parásito.....	3
1.2. Manifestaciones clínicas y tratamiento .....	4
1.3. Diagnóstico .....	5
1.4. Sensibilidad y Especificidad .....	6
1.5. La estadística bayesiana como herramienta metodológica.....	7
2. Objetivos .....	10
3. Metodología .....	11
3.1. Población y área de estudio .....	11
3.2. Técnicas diagnósticas.....	12
3.2.1. Técnica de Baermann .....	12
3.2.2. Técnica molecular .....	13
3.3. Metodología Estadística .....	14
3.3.1. Modelo matemático.....	16
3.3.2. Distribuciones a priori de la prevalencia y de la especificidad y sensibilidad de las pruebas.....	19
3.3.3. Cálculo de las distribuciones a posteriori.....	20
3.3.4. Estudio de la convergencia y comparación de modelos.....	21
4. Resultados y discusión .....	22
4.1. Población de estudio.....	22
4.2. Desarrollo de los modelos planteados .....	23
4.2.1. Distribuciones a priori.....	23
4.2.2. Distribuciones a posteriori .....	31
4.3. Estudio de convergencia para los tres modelos.....	37
4.3.1. Estudio de convergencia: Trazas .....	37
4.3.2. Estudio de convergencia: criterio Geweke y PSRF.....	42
4.4. Comparación de modelos .....	45
4.5. Conclusiones: .....	45
Anexo 1 .....	47
Bibliografía .....	61

## Índice de figuras

<b>Figura 1.</b> Mapa de distribución geográfica de <i>Strongyloides</i> spp. ....	2
<b>Figura 2.</b> Ciclo biológico de <i>Strongyloides</i> spp. (CDC) .....	3
<b>Figura 3.</b> Número de artículos por año publicados para la búsqueda “Bayesian estimation” .....	7
<b>Figura 4.</b> Número de artículos de Estadística Bayesiana publicados por año para las diferentes disciplinas.....	9
<b>Figura 5.</b> Función de densidad de la distribución a priori Beta de la sensibilidad de la PCR....	25
<b>Figura 6.</b> Función de densidad de la distribución a priori Beta de la especificidad de la PCR..	26
<b>Figura 7.</b> Función de densidad de la distribución a priori Beta de la sensibilidad de la técnica Baermann. ....	27
<b>Figura 8.</b> Función de densidad de la distribución a priori Beta de la especificidad de la técnica Baermann. ....	27
<b>Figura 9.</b> Función de densidad de la distribución a priori Beta de la sensibilidad de la PCR....	28
<b>Figura 10.</b> Función de densidad a priori Beta de la especificidad de la especificidad de la PCR. ....	29
<b>Figura 11.</b> Función de densidad de la distribución a priori Beta de la sensibilidad de la técnica Baermann. ....	29
<b>Figura 12.</b> Función de densidad de la distribución a priori Beta de la especificidad de la técnica Baermann. ....	29
<b>Figura 13.</b> Función de densidad de la distribución a priori Beta de la prevalencia.....	30
<b>Figura 14.</b> Funciones de densidad de los parámetros del modelo 1. ....	32
<b>Figura 15.</b> Funciones de densidad de los parámetros del modelo 2. ....	33
<b>Figura 16.</b> Funciones de densidad de los parámetros del modelo 3. ....	35

## Índice de Tablas

<b>Tabla 1.</b> Resultado de un test diagnóstico vs enfermedad.....	7
<b>Tabla 2.</b> Comparación de las técnicas diagnósticas empleadas.....	14
<b>Tabla 3.</b> Tabla de contingencia de comparación de técnicas.....	16
<b>Tabla 4.</b> Tablas de contingencia de comparación de técnicas diferenciando las dos poblaciones. .....	18
<b>Tabla 5.</b> Número de positivos y negativos para ambas técnicas en la población 1 y en la población 2. ....	23
<b>Tabla 6.</b> Artículos resultantes de la búsqueda bibliográfica con los datos de prevalencia para los distintos países africanos. MO: Observación microscópica, TB: Técnica de Baermann. ....	24
<b>Tabla 7.</b> Resultados del modelo “intermedio” para los parámetros estudiados.....	28
<b>Tabla 8.</b> Valores de los parámetros $\alpha$ y $\beta$ de las distribuciones Beta a priori.....	30
<b>Tabla 9.</b> Resultados del modelo 1 para los parámetros estudiados. ....	31
<b>Tabla 10.</b> Resultados del modelo 2 para los parámetros estudiados. ....	33
<b>Tabla 11.</b> Resultados del modelo 2 para los parámetros estudiados. ....	34
<b>Tabla 12.</b> Valores $\alpha$ y $\beta$ de la distribución a priori y la media más desviación estándar (D.E.) de la sensibilidad de la PCR y del Baermann, la especificidad de la PCR y Baermann y prevalencias de población de niños descalzos (1) y de niños calzados (2) de la distribución a posteriori. ....	36
<b>Tabla 13.</b> Valores del criterio Geweke para los distintos parámetros en las tres cadenas del modelo 1.....	42
<b>Tabla 14.</b> Valores del criterio PSRF para los distintos parámetros del modelo 1. ....	42
<b>Tabla 15.</b> Valores del criterio Geweke para los distintos parámetros en las tres cadenas del modelo 2.....	43
<b>Tabla 16.</b> Valores del criterio PSRF para los distintos parámetros del modelo 2. ....	43
<b>Tabla 17.</b> Valores del criterio Geweke para los distintos parámetros en las tres cadenas del modelo 3.....	44
<b>Tabla 18.</b> Valores del criterio PSRF para los distintos parámetros del modelo 3. ....	44

## Resumen

El diagnóstico de certeza de *Strongyloides stercoralis* es la detección de larvas por microscopía; sin embargo su sensibilidad es baja. La técnica de Baermann mejora dicha sensibilidad, pero es insuficiente por las limitaciones que presenta. Como alternativa, las técnicas moleculares suponen una notable mejora de sensibilidad, especificidad y rapidez. Sin embargo, actualmente no existe un *gold standard*. La Estadística Bayesiana permite abordar esta problemática en ausencia de *gold standard* estimando la sensibilidad y la especificidad de estas pruebas diagnósticas y la prevalencia de la enfermedad en contexto de zonas endémicas. Por esto nos planteamos estimar la prevalencia de la estrongiloidiasis en población infantil de una zona rural de Cubal (Angola) y la sensibilidad y especificidad de dos pruebas diagnósticas específicas de esta enfermedad, las técnicas de Baermann y PCR en tiempo real (PCRrt), empleadas en el contexto de ausencia de *gold standard*. Para ello la población se dividió en niños que caminaban descalzos y niños que caminaban calzados. Se utilizó inferencia Bayesiana para estimar tres modelos dependiendo de las distribuciones a priori fijadas para los parámetros de interés, a partir de: 1) una búsqueda bibliográfica; 2) datos de un estudio de una población de Etiopía cuyas a priori se determinaron a partir de la búsqueda bibliográfica; 3) una combinación de los dos anteriores. Tras observar los datos de la población de Cubal se obtuvieron las distribuciones a posteriori generando muestras aleatorias por el método de Monte Carlo de Cadenas de Markov. El modelo 2 presentó un bajo valor de DIC (32.02) y una mayor precisión de los intervalos de probabilidad. La sensibilidad para la técnica Baermann y su intervalo de probabilidad al 95% fue 51.54% (35.72%-69.14%), mientras que para PCRrt fue de 81.75% (67.54%-92.85%). La especificidad para Baermann fue de 94.66% (92.27%-96.73%) y para PCRrt 89.90% (85.70%-93.91%). Las prevalencias estimadas fueron 14.15% (9.02%-20.00%) para la población que camina descalza y 11.10% (6.48%-16.71%) para la que camina calzada.

A la vista de estos resultados se puede concluir que la Estadística Bayesiana permite estimar la validez de las pruebas diagnóstico en estudio así como determinar la prevalencia de estrongiloidiasis en ausencia de *gold standard*.

**Palabras clave:** *Strongyloides*, PCRrt, Baermann, sensibilidad, especificidad, prevalencia, Estadística Bayesiana.

## 1. Introducción

La estrongiloidiasis es una enfermedad parasitaria producida por *Strongyloides stercoralis* y *Strongyloides fuelleborni*. En un primer momento denominada “anguilulosis”, fue descubierta en 1876 en un grupo de soldados que había regresado de la CochinChina (Sur de Vietnam) que presentaban un cuadro clínico diarreico, por lo que también es conocida como “diarrea de la CochinChina” (Globe, 1990).

Aproximadamente entre unos 30-100 millones de personas están afectadas por la enfermedad en el mundo, la mayoría de zonas tropicales y subtropicales, destacando el Sudeste Asiático, Sur de EEUU, América Latina, India y África subsahariana (**Figura 1**), ([http://www.who.int/intestinal\\_worms/epidemiology/strongyloidiasis/en/](http://www.who.int/intestinal_worms/epidemiology/strongyloidiasis/en/)).

En referencia al nivel de prevalencia de países del África subsahariana no existen demasiados estudios al respecto. Recientemente se ha publicado un estudio realizado en Mozambique en el que desean estimar la prevalencia para un conjunto de helmintiasis y detectan para las distintas especies de *Strongyloides* (*Strongyloides* spp.) una prevalencia del 48% (Meurs *et al.*, 2017). Actualmente no se dispone de un test diagnóstico de referencia para determinar la prevalencia de la estrongiloidiasis. Asimismo, la estrongiloidiasis es una enfermedad subestimada, ya que una gran parte de casos cursan de forma asintomática (Requena-Méndez *et al.*, 2013) y los métodos de diagnóstico tienen una baja sensibilidad y/o especificidad; por lo que son numerosos los casos que no se detectan o que se confunden con otras helmintiasis. En este contexto es de especial interés el utilizar métodos de estimación que permitan determinar la prevalencia de la enfermedad así como de los parámetros de los test diagnósticos, como son la sensibilidad y especificidad de los mismos.



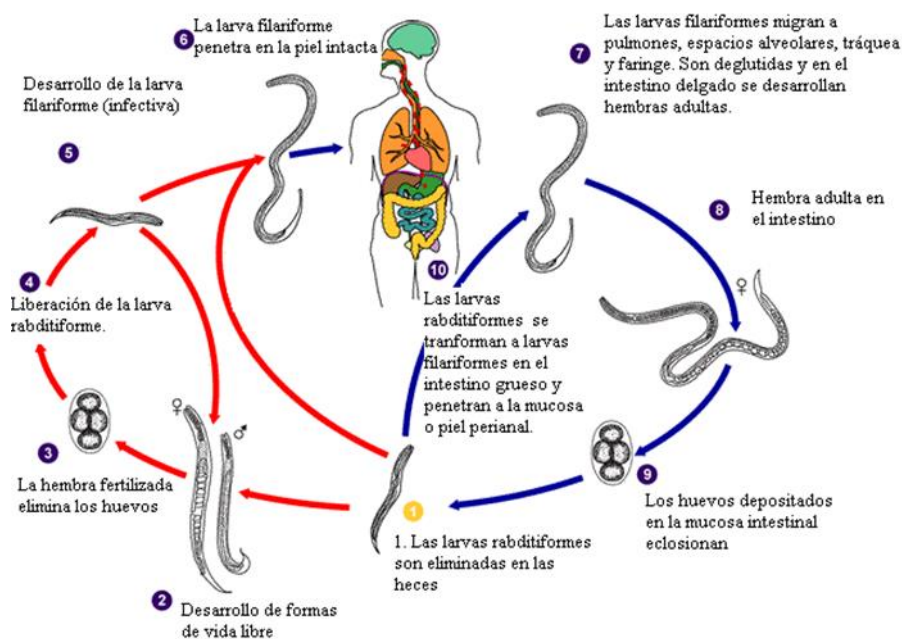
**Figura 1.** Mapa de distribución geográfica de *Strongyloides* spp.

## 1.1. Ciclo biológico del parásito

El ciclo de vida del parásito (**Figura 2**) comienza con la expulsión de larvas rabditiformes (no infectivas) en las heces de individuos infectados. Una vez en el suelo si las condiciones medioambientales son adecuadas, esta larva tras sucesivas mudas se transformará en larva filariforme (infectiva). En caso de que la larva encuentre un hospedador al que infectar, ésta penetrará por contacto a través de la piel intacta y se desplazará a los pulmones, tráquea y faringe. Posteriormente, será deglutida y se instalará en el intestino en forma de hembra adulta. La hembra produce huevos por partenogénesis, los cuales son depositados en la mucosa intestinal donde eclosionan y las larvas rabditiformes son eliminadas en las heces.

En el caso de no encontrar hospedador las larvas en el suelo con condiciones medioambientales adecuadas, seguirán mudando hasta convertirse en adultos que copularán, producirán huevos y mudarán a las distintas fases larvarias hasta encontrar un hospedador, produciéndose así un ciclo de vida libre.

Existe además el caso de autoinfección en el que la muda a larva infectiva se produce en el interior del hospedador sin necesidad de producirse en el suelo. Este proceso explica que la enfermedad pueda aparecer años después de que los individuos dejen de residir en áreas endémicas.



**Figura 2.** Ciclo biológico de *Strongyloides* spp. (CDC)

## 1.2. Manifestaciones clínicas y tratamiento

La mayoría de los pacientes infectados permanecen asintomáticos durante años. No obstante, en infecciones agudas el cuadro clínico se caracteriza por:

-Manifestaciones cutáneas. Como consecuencia de la penetración de la larva filariforme a través de la piel pueden aparecer: i) erupciones urticariformes, posiblemente provocadas por la sensibilización a los antígenos del parásito, y ii) una dermatitis característica provocada por la migración subcutánea de larvas filariformes (larva *currens*). En este último caso, los pacientes presentan una lesión cutánea serpiginosa de varios centímetros de tamaño.

-Manifestaciones pulmonares. Pueden aparecer los primeros días tras la migración larvaria, con tos, infiltrados pulmonares y acompañadas de eosinofilia intensa (síndrome de Löeffler). Estas manifestaciones aparecen de forma más severa en pacientes con neumonía, EPOC y asma bronquial. Como consecuencia de la migración pueden encontrarse larvas en muestras de esputo (García, 2007).

-Manifestaciones gastrointestinales. Aunque el tiempo puede variar para cada individuo, los síntomas gastrointestinales como diarrea, estreñimiento, y dolor abdominal suelen empezar dos semanas después de la infección, y las larvas empiezan a ser detectables en las heces después de 3 o 4 semanas.

-Síndrome de hiperinfección. Aparece cuando el individuo está inmunodeprimido (tratamiento con corticoides, HTLV-I, etc.). El ciclo de autoinfección se acelera, por lo que la carga parasitaria es mayor pudiendo provocar una infección diseminada y una localización ectópica del parásito. La tasa de mortalidad del síndrome de hiperinfección es del 15% y se incrementa hasta un 87% en los casos de diseminación (Vadlamudi *et al.*, 2006).

El tratamiento se recomienda a todas las personas infectadas, presenten o no síntomas, puesto que como ya se ha mencionado existe la posibilidad de hiperinfección y/o infección diseminada (<http://www.cdc.gov/parasites/strongyloides/treatment.html>). El tiabendazol, albendazol e ivermectina son los tratamientos más efectivos, siendo este último el fármaco de elección (en una sola dosis, 200 µg/kg vía oral durante 1-2 días).

### 1.3. Diagnóstico

#### *Diagnóstico directo*

El diagnóstico de certeza de *S. stercoralis* es la detección de larvas por microscopía. Sin embargo, la sensibilidad es baja, alrededor del 30% en el examen directo de una muestra de heces, o del 50% en muestras seriadas (Ruano *et al.*, 2005). Dicha sensibilidad puede mejorarse mediante el empleo de técnicas de migración larvaria como Baermann y Harada Mori. Aunque con distintos protocolos, ambas técnicas se basan en el tropismo de las larvas hacia el agua, lo que permite su aislamiento y visualización. También, la técnica de cultivo en agar, gracias a las marcas características que deja la larva a su paso por la superficie de agar, permite el diagnóstico de *Strongyloides* spp., así como diferenciarlas por microscopía de las larvas de otros helmintos.

Estas técnicas parasitológicas de diagnóstico directo son muy específicas pero poco sensibles, puesto que además la expulsión de las larvas se produce de forma intermitente (Dreyer *et al.*, 1996; Requena-Méndez *et al.*, 2013). Aunque se ha demostrado que las técnicas de concentración mejoran la sensibilidad, la gran mayoría de ellas están diseñadas para la búsqueda de huevos en las muestras de heces y en este caso lo que se detectan son larvas. Por todo ello estas técnicas requieren personal especializado y su realización es laboriosa. Además, es necesario que la larva sea viable, lo que dificulta mucho más su diagnóstico, cuestionando de esta forma la fiabilidad de las técnicas parasitológicas como “gold standard” o prueba de referencia para determinar la enfermedad.

Otra técnica de diagnóstico directo es la detección molecular mediante la técnica de reacción en cadena de la polimerasa (PCR). Durante los últimos años el desarrollo de este tipo de tecnología ha supuesto un gran avance y para el caso particular de *Strongyloides* spp. se han diseñado diferentes formatos de PCR: convencional, anidada, a tiempo real o múltiple, frente a distintas dianas, destacando el ADNribosomal, los segmentos espaciadores transcritos internos (ITS) o la subunidad 18S (Levenhagen y Costa-Cruz, 2014; Hasegawa *et al.*, 2009; Moghaddassani *et al.*, 2011). Son numerosas las ventajas que presentan las técnicas moleculares, en especial las basadas en tiempo real, destacando una mejora en la sensibilidad y la especificidad frente a las técnicas parasitológicas, así como mayor rapidez en la realización de la técnica y menor probabilidad de contaminación de las muestras. Una de las principales PCR a tiempo real

desarrollada en los últimos años fue descrita por Verweij et al. en el año 2009 y amplifica un fragmento de la subunidad 18 S de *S. stercoralis*, sirviendo de base para el desarrollo de diferentes protocolos de PCR (Sultana *et al.*, 2013; Saugar *et al.*, 2015) en el que se obtienen distintos valores de sensibilidad y especificidad.

#### *Diagnóstico indirecto*

De forma paralela a las técnicas parasitológicas, se utilizan técnicas serológicas para el diagnóstico de esta parasitosis. El diagnóstico serológico supone la determinación de los anticuerpos anti-*Strongyloides* en el suero del paciente. Para ello, la técnica más ampliamente utilizada es el *Enzyme-Linked Immunoabsorbent Assay* (ELISA). Dicho ensayo de inmunoabsorción es muy sensible pero poco específico debido a la existencia de reacciones cruzadas con otras helmintiasis afines: filarias, *Toxocara*, *Ascaris*, *Schistosoma*... Además, no discrimina entre infecciones pasadas y activas, lo que conlleva la necesidad de una mejora del diagnóstico serológico de esta parasitosis (Lindo *et al.*, 1994; Costa *et al.*, 2003; Requena-Méndez *et al.*, 2013).

Otra técnica complementaria de diagnóstico indirecto es el Western Blot, que permite la confirmación de los resultados obtenidos por otras técnicas (Pereira Silva *et al.*, 2003, Sudré *et al.*, 2007).

#### **1.4. Sensibilidad y Especificidad**

De forma paralela en la evaluación de las pruebas diagnósticas hacemos uso de dos conceptos clave, el de sensibilidad y especificidad. La sensibilidad es la probabilidad de que un individuo dé positivo en el test estando enfermo y la especificidad es la probabilidad de que un individuo dé negativo en el test estando sano. Se definen como Verdaderos Positivos (VP) aquellos casos de enfermos en los que el test ha dado positivo, como Verdaderos Negativos (VN) aquellos casos de sanos en los que el test ha dado negativo, como Falsos Positivos (FP) aquellos casos de sanos en los que el test ha dado positivo y como Falsos Negativos (FN) aquellos casos de enfermos en los que el test ha dado negativo. Así la sensibilidad se puede definir como el cociente entre los VP y VP+FN, y la especificidad como el cociente entre VN y VN+FP. Todos estos conceptos se resumen en la **tabla 1**.

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

Resultado test diagnóstico	Diagnóstico de referencia	
	Enfermo	Sano
Positivo	VP	FP
Negativo	FN	VN

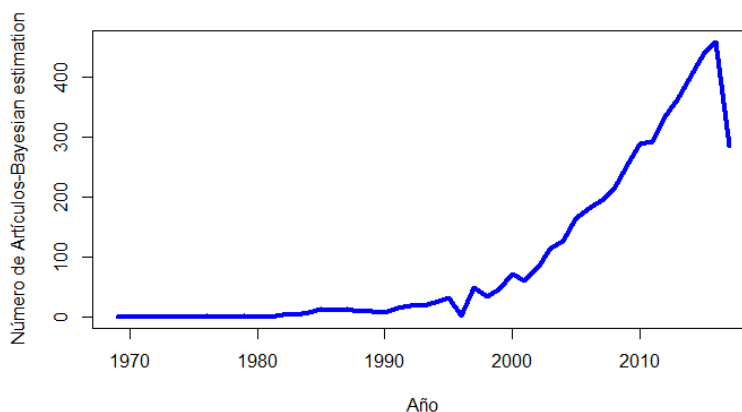
**Tabla 1.** Resultado de un test diagnóstico vs enfermedad.

En términos de diagnóstico la mejor prueba sería aquella que tuviera elevados valores tanto de sensibilidad como de especificidad.

Estos dos valores son independientes de otro concepto importante, el de prevalencia que se define como la proporción de personas que sufren una enfermedad en un momento determinado con respecto al total de la población de estudio, por lo que los valores de sensibilidad y especificidad solo harán referencia a las técnicas diagnósticas y el de prevalencia a la población de estudio.

### 1.5. La estadística bayesiana como herramienta metodológica

El problema planteado en este trabajo de fin de máster ante la ausencia de una técnica de *gold standard* para el diagnóstico de la estrongiloidiasis se resolverá utilizando la metodología Bayesiana. El primer artículo del que se tiene constancia cuando se introducen los términos “Bayesian estimation” en la base de datos PubMed data de 1969 y ha existido una evolución exponencial, especialmente en los últimos años y particularmente a partir del año 2000 (**Figura 3**), apareciendo 469 artículos en el año 2016 y 286 en lo que llevamos del año 2017.



**Figura 3.** Número de artículos por año publicados para la búsqueda “Bayesian estimation”.

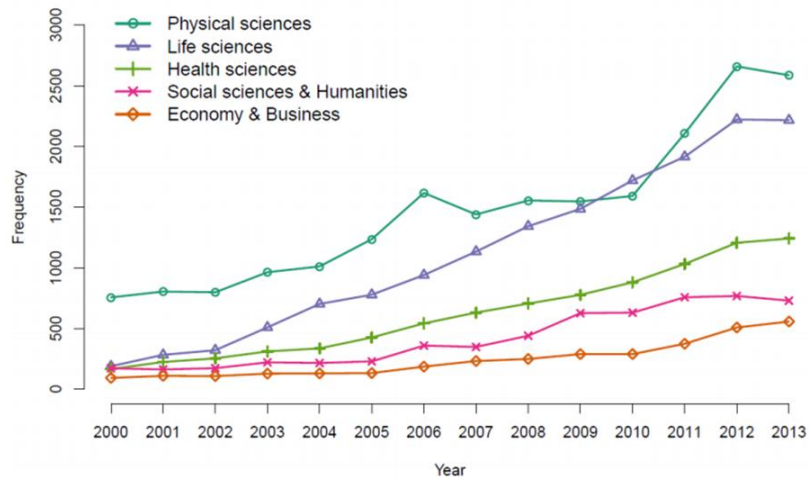
La Estadística Bayesiana se apoya en la interpretación subjetiva de la probabilidad, es decir, en el grado de creencia que se tiene sobre la ocurrencia de un suceso y tiene como eje el Teorema de Bayes, desarrollado por Thomas Bayes y publicado, tras su muerte, en 1763. Expresa la probabilidad de que ocurra un suceso A dado que ha ocurrido un suceso B en función de la probabilidad de que haya ocurrido el suceso B dado que ha ocurrido el suceso A. Es decir:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad \text{con } P(B) > 0 \quad (1)$$

Una interpretación de este teorema, es que permite “actualizar” la probabilidad de un suceso A,  $P(A)$ , cuando se conoce que ha ocurrido un suceso B, pudiéndose calcular así la probabilidad de A condicionada a que ha ocurrido B,  $P(A|B)$  y obtener nuevas distribuciones (Silva, L.C. & Benavides, A., 2001).

Las ventajas que presenta la Estadística Bayesiana frente al enfoque clásico son numerosas, destacando que no se necesita un gran tamaño muestral, y parte de un conocimiento previo aprovechable. Este último hecho ha generado mucha controversia a este tipo de análisis puesto que en ocasiones se ve que la aportación subjetiva del investigador podría llegar a ser una manipulación de los datos. Sin embargo, esto no tiene por qué ser así, una buena justificación de los conocimientos que se tienen a priori respaldado por una búsqueda bibliográfica adecuada puede dar a nuestro conocimiento la base sólida para considerar esta información inicial como certera y/o fiable.

Cada vez es más frecuente el uso de la Estadística Bayesiana en el ámbito de las ciencias de la salud. En el siguiente gráfico (**Figura 4**) se muestra la evolución del número de artículos publicados de Estadística Bayesiana para las distintas ramas (Depaoli, S. & van de Schoot, R., 2017), siendo las ramas de ciencias físicas, de la vida y de la salud las tres primeras y dejando en segundo plano a las ciencias sociales.



**Figura 4.** Número de artículos de Estadística Bayesiana publicados por año para las diferentes disciplinas.

Actualmente existe una lista conocida por sus siglas en inglés como la WAMBS (*When to worry and how to Avoid the Misuse of Bayesian Statistics*) donde se describen las principales razones en la que el uso de la estadística bayesiana está justificada (Depaoli, S. & van de Schoot, R., 2017): algunos modelos complejos no pueden ser estimados con la metodología frecuentista; algunos investigadores opinan que los parámetros de la población deben ser vistos como aleatorios por la propia naturaleza del problema; se puede incorporar la información que se tiene de los parámetros y que este conocimiento se puede ir actualizando.

La Estadística Bayesiana no necesita basarse en muestras de gran tamaño ya que se pueden obtener resultados razonables incluso con pequeños tamaños muestrales, especialmente cuando disponemos de un conocimiento a priori consistente. La Estadística Bayesiana se basa en realizar las inferencias sobre las cantidades desconocidas a partir de un modelo conjunto de probabilidad que relaciona los datos y los parámetros, considerando a estos últimos como aleatorios, dado el desconocimiento acerca de los mismos. La regla de Bayes proporciona un modo natural de actualización de las creencias cuando aparece nueva información. Este proceso de aprendizaje inductivo por medio de la regla de Bayes es la base de la Inferencia Bayesiana.

La metodología Bayesiana consta de los siguientes pasos fundamentales:

1. Definir el modelo de probabilidad correspondiente a las observaciones recogidas donde se relacionen los datos y los parámetros desconocidos (verosimilitud).

2. Especificar un modelo o distribución de probabilidad a priori para los parámetros de interés que recoja el conocimiento previo que se tiene de ellos. Se denomina elicitación (*elicitation*) a la determinación de los parámetros (hiperparámetros) que definen la distribución a priori de los parámetros de interés (Van de Schoot *et al.*, 2014). Existen tres clases de información a priori dependiendo del grado de información disponible: (1) a priori no informativa (2) a priori débilmente informativa y (3) a priori informativa. La a priori informativa es crucial para estimar el modelo y puede tener un gran impacto en la estimación final (Van de Schoot *et al.*, 2014), sin embargo la distribución a priori no informativa a menudo se asemeja a un análisis frecuentista, en la que son los datos los que determinan la distribución a posteriori del parámetro de interés. Para determinados parámetros como por ejemplo las proporciones, estos tres niveles de información son creados al modificar los hiperparámetros de la distribución a priori. Cada parámetro en el modelo necesita de una distribución a priori, y el uso de la distribución beta como a priori está justificado en el caso de proporciones por su propia definición, además dadas las propiedades de dicha distribución es posible simplificar los cálculos en la aplicación de la metodología bayesiana.

3. Actualizar el conocimiento de los parámetros, condicionando el modelo a los datos observados (distribución a posteriori). En la mayoría de los casos, la distribución a posteriori se obtiene por simulación, utilizando el método de Monte Carlo de Cadenas de Markov (MCMC) usando el muestreador de Gibbs (Enoe *et al.*, 2000) u otros algoritmos como Metrópolis-Hastings. Existen numerosos paquetes estadísticos para implementar la metodología Bayesiana: paquetes de R, AMOS, Mplus, BIEMS, SAS.

4. Evaluar el ajuste del modelo a los datos y la sensibilidad del modelo a modificaciones en los parámetros.

## 2. Objetivos

El objetivo de este trabajo es estimar tanto la prevalencia de la estrongiloidiasis en población infantil de una zona rural de Cubal (Angola) como la sensibilidad y especificidad de dos pruebas diagnósticas específicas de esta enfermedad, la técnica de Baermann y PCR en tiempo real, empleadas en el contexto de ausencia de *gold standard*.

### 3. Metodología

#### 3.1. Población y área de estudio

Se recogieron un total de 315 muestras de heces pertenecientes a población infantil de edades comprendidas entre 5 y 14 años. El área geográfica escogida para el estudio fue una zona rural del municipio de Cubal (Angola) que pertenece a la provincia de Benguela y situado a 146 kms de esta ciudad y a unos 700 kms de la capital (Luanda). Se estima que hay 320.000 habitantes de los cuales el 47% son niños. Este municipio está dividido en cuatro comunas: Cubal, Tumbulo, Capupa y Yambala. El principal biotopo es la sabana y hay una fuerte alternancia entre la estación húmeda y seca, siendo los meses de noviembre a abril cuando tiene lugar la época de lluvias y los meses de mayo y octubre cuando se produce la estación seca.

Las muestras se recogieron en el período de enero a junio de 2015 en un total de 16 escuelas distribuidas de forma igualitaria en las 4 comunas: comuna Cubal (Garcia Neto, Tomás Ferreira, Comandante Bula, Teresianas), comuna Tumbulo (Tumbulo Sede, Kalamuma, Kambinji, Lulambo), comuna Capupa (Loneta Halu, Hoji-ha-henda, Caviva Sul, Cassua) y comuna Yambala (Kambondiongolo, Kassalasisitu, Yamba Sede, Atiopo). De forma paralela a la recogida de muestras se obtuvo información a través de un cuestionario acerca de los diferentes hábitos de carácter social: utilización de zapatos, consumo de carne cruda, lavado de vegetales, lugar de baño...

Al disponer de datos sobre el uso o no de zapatos y sabiendo que la transmisión de la enfermedad se produce, en la mayoría de los casos por la penetración de la larva al caminar descalzo en terrenos contaminados con heces, tanto de animales como de personas infectadas y a su vez, teniendo en cuenta los numerosos artículos publicados en los que se observa una reducción en el número de casos de la enfermedad después de la utilización de calzado, se decidió dividir a la población en dos subgrupos (población 1 que camina descalzo y población 2 que camina con zapatos) y tratarlas como dos poblaciones diferentes en las que se estimarán las prevalencias de la enfermedad.

Tanto la técnica parasitológica de Baermann como la concentración y extracción de ADN de las muestras de heces se realizaron en el Hospital Nossa Senhora da Paz, que se encuentra a las afueras de Cubal, en Tchambungo. Este hospital de carácter privado está gestionado por las hermanas de la Compañía de Santa Teresa de Jesús y asociado a

una de las escuelas incluidas en el estudio (Teresianas). En su origen actuaba únicamente como dispensario pero las circunstancias de la guerra civil que duró casi 30 años (1975-2002) le obligaron a transformarse en hospital, siendo ahora uno de los principales centros de referencia de tuberculosis del país. La técnica molecular de la PCR se realizó en el Instituto de Salud Carlos III en Majadahonda (Centro Nacional de Microbiología).

Este estudio ha sido aprobado por el Comité Ético del Hospital Vall d'Hebron y por el Instituto regional de Salud y Educación de Angola.

## **3.2. Técnicas diagnósticas**

### **3.2.1. Técnica de Baermann**

Esta técnica se base en el termotropismo e hidrotropismo positivo que poseen las larvas de *Strongyloides stercoralis*. Las larvas migran desde las heces hacia el agua a 37°C, depositándose por gravedad en el fondo del vaso de precipitación. Primero, las muestras se mezclaron con una pequeña cantidad de agua (1/3) y con una espátula se pesaron 15 gramos de heces aproximadamente, que fueron mezclados con carbón activado en una placa Petri y se incubó durante 24 horas a 26°C. Esta muestra se colocó durante 1 hora sobre un colador en la parte de arriba de un embudo con agua a 37°C. El embudo en su parte posterior está conectado con una goma de unos 35 cm aproximadamente, cerrada en su extremo por una llave de Mohr. En el caso de que existan larvas deberían migrar a la parte del agua quedando retenidas en la parte de la goma antes del cierre. Posteriormente, se puso una pinza hemostática en la parte de la goma más próxima al embudo y se quita la llave de Mohr, de manera que el agua que está retenida se va a recoger en un vaso de vidrio. Se pasó el agua recogida a un tubo cónico y se dejó sedimentar durante al menos 30 minutos y después se centrifugó a 2000 rpm durante 5 minutos. Se eliminó el sobrenadante y se visualizó el sedimento por microscopía (10x-40x) en un portaobjetos añadiendo una gota de lugol diluido al 20% en solución salina. Cada muestra se verificó tres veces por tres microscopistas diferentes y experimentados. Las larvas de *Strongyloides stercoralis* fueron identificadas por la cavidad bucal y los órganos genitales en el caso de larvas rabditiformes y por la punta posterior en las larvas filariformes.

### 3.2.2. Técnica molecular

#### *Extracción de ADN*

Previa a la extracción de ADN se realizó a través del sistema comercial Bioparaprep MINI® (Leti Diagnostics, Barcelona, España) una concentración de la muestra de heces a partir de 1 gramo disuelto con ayuda de un vórtex en 4 mL de solución salina al 0.9%. Posteriormente la mezcla se añadió al tubo con filtro provisto por el kit y se centrifugo durante 5 minutos a 2500 rpm. Una vez centrifugado, se eliminó el sobrenadante y se añadió 1 ml de solución salina. Se agitó de nuevo con un vórtex y por último se centrifugó durante 1 minuto a 13000 rpm y se eliminó el sobrenadante.

El sedimento se utilizó para la extracción de ADN que se realizó empleando el kit comercial QiaAmp Fast DNA Stool Mini kit (Qiagen, Hilen, Germany) y siguiendo las instrucciones del fabricante, eluyendo el ADN en 200 µl de agua destilada. Las muestras de ADN se mantuvieron a 4°C hasta su uso.

#### *Detección molecular mediante PCR a tiempo real*

Para realizar la amplificación se emplearon los cebadores diseñados y de secuencia descrita por Verweij *et al.*, (2009). Estos cebadores permiten amplificar una fracción del gen 18S rRNA, y presentan la siguiente secuencia: Cebador directo: 5' GAATTCCAAGTAAACGTAAGTCATTAGC 3'. Cebador reverso: 5' TGCCTCTGGATATTGCTCAGTTC 3'. La amplificación y detección se realizó en un volumen final de 25 µL. Para ello se emplearon 15 µl de un mezcla de reactivos formada por: Quantimix EasyProbes 1X (Biotools B& M Laboratorios, Madrid, España), que contiene los 4 nucleótidos necesarios (dNTPs), la enzima ADN polimerasa, MgCl<sub>2</sub> y el buffer; los cebadores directo y reverso a una concentración de 0.2 µM y el fluoróforo SYBR Green 0,3X (Invitrogen, San Diego, CA, EEUU). Para completar la reacción se añadió el ADN problema a una concentración final de 125 ng en un volumen máximo de 10 µL. Las condiciones de la amplificación fueron 15 minutos a 95°C, seguidos de 50 ciclos de 10 segundos a 95°C, 10 segundos a 60°C y 30 segundos a 72°C. La detección de la fluorescencia se realizó en cada ciclo. Para ello se utilizó un equipo CorbettRotorGene™ 6000 real-time PCR system (Qiagen, Hilen, Germany) y el software RotorGene 6000 Series version 1.7 para el análisis los datos. En cada uno de los ensayos se incluyeron 6 controles: un blanco formado por la mezcla de reactivos sin ADN, un control negativo, 3 controles positivos formados por ADN de *S. venezuelensis* (de los cuales, dos se emplearon a una concentración de 50 y 5 ng permitiendo así estandarizar

la curva), el último control formado por ADN del control negativo, al que se le añadió ADN de *S. venezuelensis*, como control de inhibición. Cada muestra problema se ensayó por triplicado. Dos de los análisis se emplearon para detectar ADN de *Strongyloides* spp., mientras que al tercero se le añadió ADN de *S. venezuelensis*, para detectar inhibición en la amplificación. Una vez finalizada la amplificación y detección, se realizó el análisis de la curva de melting, determinando así la Temperatura de Melting (Tm) de los productos amplificados y con ello la especificidad de la amplificación. Para que una muestra sea considerada positiva la fluorescencia debe cruzar el umbral de forma exponencial y el producto de amplificación debe rendir una Tm de  $85\pm 1^{\circ}\text{C}$ .

A continuación se muestra una tabla comparativa (**tabla 2**) de las principales características de las dos técnicas diagnósticas empleadas en el estudio.

	Baermann	PCR
Detección	Larvas vivas	ADN
Rapidez	++	+++
Necesidad de entrenamiento	+++	+
Sensibilidad	+	+++
Especificidad	+++	+++
Accesibilidad económica	+++	+

**Tabla 2.** Comparación de las técnicas diagnósticas empleadas.

### 3.3. Metodología Estadística

Como hemos expuesto anteriormente la metodología a utilizar está basada en el paradigma Bayesiano que permite actualizar la información de los parámetros de interés a partir de la información contenida en los datos mediante la expresión obtenida a partir del Teorema de Bayes:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad \text{con } P(B) > 0 \quad (2)$$

Según la definición de la probabilidad condicionada

$$P(\theta, D) = P(\theta) * P(D | \theta) \quad (3)$$

La  $P(\theta)$  corresponde a la distribución a priori del parámetro (o parámetros)  $\theta$  de interés y queremos conocer la  $P(\theta | D)$  que es la distribución a posteriori, conocido el valor de  $D$  que son nuestros datos y por lo tanto es un valor constante.

Como  $P(\theta, D) = P(D, \theta)$ , entonces

$$P(\theta) * P(D | \theta) = P(D) * P(\theta | D) \quad (4)$$

Despejando,

$$P(\theta | D) = \frac{P(\theta) * P(D | \theta)}{P(D)} \quad (5)$$

Al ser los datos  $D$  conocidos y por tanto fijos,  $P(D | \theta)$  se corresponde con la función de verosimilitud  $L(\theta; D)$  y  $P(D)$  es un valor constante, obteniéndose

$$P(\theta | D) = \frac{P(\theta) * L(\theta; D)}{k} \quad (6)$$

Entonces:

$$P(\theta | D) \propto P(\theta) * L(\theta; D) \quad (7)$$

Donde la distribución a posteriori es proporcional ( $\propto$ ) a la distribución a priori por la función de verosimilitud. En resumen:

**Distribución a posteriori  $\propto$  Distribución a priori \* Función de verosimilitud**

### 3.3.1. Modelo matemático

En el problema que nos ocupa, nuestros datos se corresponden al formato éxito-fracaso, existiendo únicamente dos posibilidades (positivo o negativo en el test, y presencia o ausencia de la enfermedad). Al tratar conjuntamente los resultados de dos pruebas diagnóstico, el esquema de los datos que tenemos en cada una de las dos poblaciones consideradas (pies descalzos y pies calzados) se muestra en la **tabla 3** (Dendukuri, N *et al.*, 2010):

	Baermann(+)	Baermann(-)	
PCR (+)	$n_{11}$	$n_{12}$	$n_{11} + n_{12}$
PCR (-)	$n_{21}$	$n_{22}$	$n_{21} + n_{22}$
	$n_{11} + n_{21}$	$n_{12} + n_{22}$	$n$

**Tabla 3.** Tabla de contingencia de comparación de técnicas.

En este sentido, los datos se corresponden con una distribución multinomial  $Y \sim \text{multinomial}(n, p_{11}, p_{12}, p_{21}, p_{22})$  con  $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$  (8)

Es decir:

$$P(Y = (n_{11}, n_{12}, n_{21}, n_{22})) = \frac{n!}{n_{11}! n_{12}! n_{21}! n_{22}!} p_{11}^{n_{11}} p_{12}^{n_{12}} p_{21}^{n_{21}} p_{22}^{n_{22}} \quad (9)$$

Donde

$$\sum_{i=1}^2 \sum_{j=1}^2 n_{ij} = n \quad (10)$$

La  $n$  representa el tamaño total de la población,  $p_{11}$  corresponde a la probabilidad de tener un resultado positivo en ambas pruebas,  $p_{12}$  a la probabilidad de dar un positivo por PCR y negativo por Baermann,  $p_{21}$  la probabilidad de dar un positivo por Baermann pero negativo por PCR y  $p_{22}$  la probabilidad de dar negativo en ambas técnicas.

Por el teorema de la probabilidad total es posible desarrollar las probabilidades tal y como se muestra a continuación, donde  $E$  representa a los enfermos ( $\bar{E}$  sanos) y  $+ o -$  el resultado positivo o negativo de los tests Baerman (B) y PCR:

$$p_{11} = P(PCR = +, B = +) \quad (11)$$

$$= P(PCR = +, B = +|E) * P(E) + P(PCR = +, B = +|\bar{E}) * P(\bar{E})$$

$$= P(PCR = +|E) * P(B = +|E) * P(E) + P(PCR = +|\bar{E}) * P(B = +|\bar{E}) * P(\bar{E})$$

$$p_{12} = P(PCR = +, B = -) \quad (12)$$

$$= P(PCR = +, B = -|E) * P(E) + P(PCR = +, B = -|\bar{E}) * P(\bar{E})$$

$$= P(PCR = +|E) * P(B = -|E) * P(E) + P(PCR = +|\bar{E}) * P(B = -|\bar{E}) * P(\bar{E})$$

$$p_{21} = P(PCR = -, B = +) \quad (13)$$

$$= P(PCR = -, B = +|E) * P(E) + P(PCR = -, B = +|\bar{E}) * P(\bar{E})$$

$$= P(PCR = -|E) * P(B = +|E) * P(E) + P(PCR = -|\bar{E}) * P(B = +|\bar{E}) * P(\bar{E})$$

$$p_{22} = P(PCR = -, B = -) \quad (14)$$

$$= P(PCR = -, B = -|E) * P(E) + P(PCR = -, B = -|\bar{E}) * P(\bar{E})$$

$$= P(PCR = -|E) * P(B = -|E) * P(E) + P(PCR = -|\bar{E}) * P(B = -|\bar{E}) * P(\bar{E})$$

Basándonos en el artículo de Branscum *et al.* del año 2005 se ha tomado el modelo propuesto por Hui and Walter (1980) para dos poblaciones y dos test diagnósticos condicionalmente independientes dada la enfermedad. Los dos test diagnósticos se aplican a las dos poblaciones y con esos resultados se construyen tablas de contingencia 2x2 a través de R mediante la función *table*, en las que aparecen reflejados el número de los resultados positivos y negativos para cada técnica en cada población. Este modelo asume que los test son independientes condicionalmente dada la enfermedad y que los valores de sensibilidad y especificidad de ambas técnicas son iguales para las dos poblaciones. Se asumió independencia condicional entre los test porque aunque ambos son técnicas de diagnóstico directo la base de la metodología es diferente, en la técnica de Baermann se visualiza el parásito mediante microscopía mientras que en la técnica de PCR se detecta ADN del parásito. En nuestro caso de dos técnicas diagnósticas y dos poblaciones en total tenemos 6 parámetros que queremos estimar:

- Sepcr: sensibilidad de la técnica PCR
- Sppcr: especificidad de la técnica PCR

- Sebaermann: sensibilidad de la técnica Baermann
- Spbaermann: especificidad de la técnica Baermann
- $\pi_1$ : de caminan descalzos prevalencia de la población 1
- $\pi_2$ : de caminan con zapatos prevalencia de la población 2

Dado que nuestro interés se centra en las dos poblaciones previamente descritas como niños que caminan descalzos y que caminan con zapatos, en la población  $k$  ( $k=1,2$ ) la variable  $Y_k$  sigue una distribución multinomial

$$Y_k \sim \text{multinomial}(n_k, p_{11k}, p_{12k}, p_{21k}, p_{22k}), \quad (15)$$

De forma que los datos recogidos se pueden estructurar tal y como se muestra en la **tabla 4**.

Población 1: Caminan descalzos		Técnica Baermann	
		Positivo	Negativo
Técnica PCR	Positivo	$n_{111}$	$n_{121}$
	Negativo	$n_{211}$	$n_{221}$

Población 2: Caminan con zapatos		Técnica Baermann	
		Positivo	Negativo
Técnica PCR	Positivo	$n_{112}$	$n_{122}$
	Negativo	$n_{212}$	$n_{222}$

**Tabla 4.** Tablas de contingencia de comparación de técnicas diferenciando las dos poblaciones.

Así, las expresiones de los  $p_{ijk}$  donde  $i,j,k=1,2$ , pueden expresarse en términos de nuestros parámetros de interés dados por las prevalencias de la estrongiloidiasis en ambas poblaciones y las sensibilidades y especificidades de ambas técnicas, de la siguiente manera:

Para  $k=1$  que corresponde a la población 1: Caminan descalzos

$$p_{111} = \pi_1 * Sepcr * Sebaermann + (1 - \pi_1) * (1 - Sppcr)(1 - Spbaermann) \quad (16)$$

$$p_{112} = \pi_1 * Sepcr * (1 - Spbaermann) + (1 - \pi_1) * (1 - Sppcr) * Spbaermann \quad (17)$$

$$p_{211} = \pi_1 * (1 - Sepcr) * Sebaermann + (1 - \pi_1) * Sppcr * (1 - Spbaermann) \quad (18)$$

$$p_{221} = \pi_1 * (1 - Sepcr) * (1 - Sebaermann) + (1 - \pi_1) * Sppcr * Spbaermann \quad (19)$$

Para  $k=2$  que corresponde a la población 2: Caminan con zapatos

$$p_{112} = \pi_2 * Sepcr * Sebaermann + (1 - \pi_2) * (1 - Sppcr)(1 - Spbaermann) \quad (20)$$

$$p_{122} = \pi_2 * Sepcr * (1 - Spbaermann) + (1 - \pi_2) * (1 - Sppcr) * Spbaermann \quad (21)$$

$$18 \quad (22)$$

$$(23)$$

$$p_{212} = \pi_2 * (1 - \text{Sepcr}) * \text{Sebaermann} + (1 - \pi_2) * \text{Sppcr} * (1 - \text{Spbaermann})$$

$$p_{222} = \pi_2 * (1 - \text{Sepcr}) * (1 - \text{Sebaermann}) + (1 - \pi_2) * \text{Sppcr} * \text{Spbaermann}$$

### 3.3.2. Distribuciones a priori de la prevalencia y de la especificidad y sensibilidad de las pruebas.

Se denomina “elicitación” (*elicitation*) a la determinación de los valores de los hiperparámetros que definen la distribución a priori de los parámetros.

Dado que los parámetros de interés son probabilidades (sensibilidad PCR, sensibilidad Baermann, especificidad PCR, especificidad Baermann, prevalencia población 1 y prevalencia población 2) es adecuado utilizar la distribución Beta. Una de las principales ventajas de esta distribución es su ajuste a una gran variedad de distribuciones empíricas, adoptando muy diversas formas dependiendo de los valores de los parámetros.

La función de densidad de una distribución Beta de parámetros  $(\alpha, \beta)$  con  $\alpha, \beta > 0$ , viene dada por la siguiente fórmula:

$$f(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1}, \quad \text{si } 0 \leq \theta \leq 1 \quad (24)$$

Existen diferentes formas de “elicitación” los parámetros  $\alpha$  y  $\beta$  de la distribución Beta. En general, bastarán dos restricciones (por ejemplo, sobre la expresión de la media, moda, varianza o cuantiles) para asignar los valores de la distribución a priori de cada parámetro.

En este trabajo fin de máster, se han establecido dos formas distintas de elicitación de los parámetros  $\alpha$  y  $\beta$  de las distribuciones Beta a priori que se asignarán a los parámetros de interés: a partir de bibliografía existente y a partir de datos de un estudio previo con población infantil de Etiopía.

El primero de ellos se basó en la búsqueda bibliográfica a través de la base de datos PubMed introduciendo varios criterios: *strongyloides* AND *prevalence*, *strongyloides* AND *sensitivity*, *strongyloides* AND *specificity* y únicamente se seleccionaron aquellos artículos llevados a cabo en países africanos por las posibles coincidencias en cuanto a condiciones de trabajo y hábitos de vida de los propios pacientes. Se han incluido todos los estudios que aparecían tras la búsqueda con los criterios de selección, independientemente de si proporcionaban todos los datos.

La segunda forma de elicitar las distribuciones a priori fue a través de las distribuciones a posteriori obtenidas con los datos recogidos en un estudio previo de condiciones similares realizado en un área rural de Etiopía. En este caso como hiperparámetros de la distribución a priori se utilizaron los mismos que en el caso anterior (datos bibliográficos).

La elicitación de todos los parámetros  $\alpha$  y  $\beta$  se realizó mediante la librería LearnBayes y la función incluida dentro de ésta de *beta.select* dentro del programa R (Version 1.0.44 – © 2009-2016 RStudio, Inc.) en la que se introduce el valor de dos cuantiles, uno inferior y otro superior, añadiendo la probabilidad de ambos.

Estas propuestas de elicitaciones posibles dan lugar a plantear tres posibles modelos distintos:

Modelo 1. Modelo con distribuciones a priori obtenidas a partir de búsqueda bibliográfica.

Modelo 2. Modelo con distribuciones a priori obtenidas a partir de los datos del estudio de una población de Etiopía.

Modelo 3. Modelo con distribuciones a priori para las sensibilidades y especificidades a partir de búsqueda bibliográfica y para las prevalencias a partir de los datos de Etiopía.

### **3.3.3. Cálculo de las distribuciones a posteriori.**

Se van a generar muestras de la distribución de probabilidad a posteriori de los parámetros a través del método de Monte Carlo de Cadenas de Markov (MCMC). Una Cadena de Markov es una secuencia de variables en la que cada valor generado depende sólo del anterior valor simulado y viene definida por su estado inicial y el núcleo de transición. Los métodos MCMC se basan en construir una Cadena de Markov de tal manera que converja a una distribución estacionaria que sea la distribución a posteriori de interés. Así, a partir de un momento, los valores de la Cadena de Markov serán una muestra de esta distribución a posteriori. Para que se produzca esta convergencia es necesario simular la cadena para un número elevado de iteraciones. Generalmente en estas cadenas se suele descartar un número alto de valores al principio de la misma, en lo que se denomina periodo de calentamiento o *burn-in*, ya que los primeros valores simulados no están en el estado estacionario.

En nuestro caso se han generado tres cadenas con distintos puntos iniciales elegidos aleatoriamente para asegurar la independencia del punto de partida y con un total de 100000 iteraciones y con un burn-in de las 10000 primeras. La estimación del modelo se realizó mediante la librería R2OpenBugs que llama al programa OpenBUGS a través de la función *bugs* incluidas en el programa R.

### 3.3.4. Estudio de la convergencia y comparación de modelos.

Para comprobar que el método MCMC ha generado Cadenas de Markov con distribución estacionaria es necesario estudiar la convergencia de cada cadena generada. Así, el estudio de la convergencia se realizó mediante dos vías: gráfica y analíticamente. La primera fue mediante comparación por visualización directa de los gráficos de traza proporcionados por el programa OpenBUGS para cada parámetro estimado (sensibilidad y especificidad de ambas técnicas y prevalencia de la enfermedad en ambas poblaciones) de las tres cadenas generadas. La traza consiste en el gráfico secuencial de los valores de la cadena simulada. La segunda comprobación de la convergencia de dichas cadenas fue a través de dos criterios analíticos: el criterio de convergencia PSRF (*Potencial Scale Reduction Factor*) o Brooks-Gelman-Rubin y el criterio de convergencia de Geweke, dichos criterios se calcularon con la librería *coda* (*Convergence Diagnosis and Output Analysis*) de R.

El criterio de convergencia PSRF se basa en comparar la variabilidad entre cadenas con la variabilidad dentro de las cadenas, de forma que los valores cercanos a 1 indican una buena convergencia de las cadenas. El criterio de convergencia de Geweke estudia la convergencia de cada cadena individualmente mediante la comparación de las medias de la misma cadena en diferentes puntos de ésta, por ejemplo la media de un porcentaje del principio de la cadena y un porcentaje del final. Si existe convergencia no se detectarán diferencias entre los valores medios de ambos tramos y, por tanto, el valor del estadístico Z de la normal debe estar entre (-1.96, 1.96).

Para poder comparar los 3 modelos enunciados que se obtienen se utilizó el criterio DIC (*Deviance Information Criterion*) que es una medida de la validez de ajuste del modelo a los datos asociando estos con la verosimilitud mediante la fórmula:

$$\text{Deviance} = -2 * \log(L(D; \theta)) , \quad (25)$$

siendo L la verosimilitud, D los datos y  $\theta$  los parámetros.

Este resultado se genera automáticamente en el programa OpenBUGS en el apartado de *Deviance Information* y aparece desglosado en dos:  $Dbar$  que es la media a posteriori de deviance y  $Dhat$  que es la deviance de la media a posteriori. Dados estos dos valores se define  $pD = Dbar - Dhat$  y el DIC se puede calcular mediante dos fórmulas:

$$DIC = Dbar + pD \quad (26)$$

o bien

$$DIC = Dhat + 2 * pD \quad (27)$$

Cuanto menor es el valor obtenido mejor es el ajuste, mayor verosimilitud de los datos con las distribuciones a posteriori de los parámetros, por lo que se escogerá aquel modelo que su DIC total sea menor.

Añadido a este criterio se observó la precisión de los intervalos de probabilidad de los distintos parámetros de interés. Esta precisión viene dada por la amplitud en el intervalo de probabilidad al 95%, obteniéndose mayor precisión en la estimación a posteriori cuanto menor es la amplitud del mismo.

Finalmente se escogerá el modelo que presente mejores valores del criterio DIC y un intervalo de probabilidad más estrecho.

## 4. Resultados y discusión

### 4.1. Población de estudio

La población total (n=324) tal y como viene especificado en el apartado de materiales y métodos se dividió en dos poblaciones atendiendo al criterio de si caminan descalzos (población 1) o caminan con zapatos (población 2). La población 1 consta de un total de 206 niños (89 hombres y 117 mujeres) de los cuales, 24 fueron positivos por la técnica Baermann mientras que por la técnica de PCR se obtuvo el doble de niños positivos (n=49). Estos datos representan un 11.11% en el caso de la técnica de Baermann y un 23.79% en PCR. La población 2 consta de un total de 118 niños (49 hombres y 69 mujeres) de los cuales, 12 fueron positivos por Baermann y 22 por PCR, representando respectivamente un 10.17% y un 18.64%. El número de positivos obtenidos por ambas técnicas fue mayor en la población 1 que en la 2 manteniéndose, en ambas poblaciones,

la proporción entre número de resultados positivos por PCR y por Baermann, siendo este primer resultado en ambos casos casi el doble (2.04 para la población 1 y 1.83 para la población 2).

Las tablas de contingencia que muestran el número de niños positivos y negativos para cada una de las técnicas (PCR y Baermann) de las dos poblaciones aparecen recogidas en las **tabla 5**.

Población 1: Caminan descalzos		Técnica Baermann	
		Positivo	Negativo
Técnica PCR	Positivo	14	35
	Negativo	10	147

Población 2: Caminan con zapatos		Técnica Baermann	
		Positivo	Negativo
Técnica PCR	Positivo	5	17
	Negativo	7	89

**Tabla 5.** Número de positivos y negativos para ambas técnicas en la población 1 y en la población 2.

Las variables Y1 que corresponde a la población 1 y Y2 que corresponde a la población 2 resultantes de ambas poblaciones se muestran a continuación:

$$Y1 \sim \text{multinomial}(n1 = 206, (p_{111}, p_{211}, p_{121}, p_{221})) \quad (28)$$

$$Y2 \sim \text{multinomial}(n2 = 118, (p_{112}, p_{212}, p_{122}, p_{222})) \quad (29)$$

## 4.2. Desarrollo de los modelos planteados

### 4.2.1. Distribuciones a priori

Tal como se explica en el apartado 3.3.2. del presente trabajo se han planteado tres modelos como resultado de las propuestas de elicitaciones posibles.

*4.2.1.1. Modelo 1. Modelo con distribuciones a priori obtenidas a partir de búsqueda bibliográfica.*

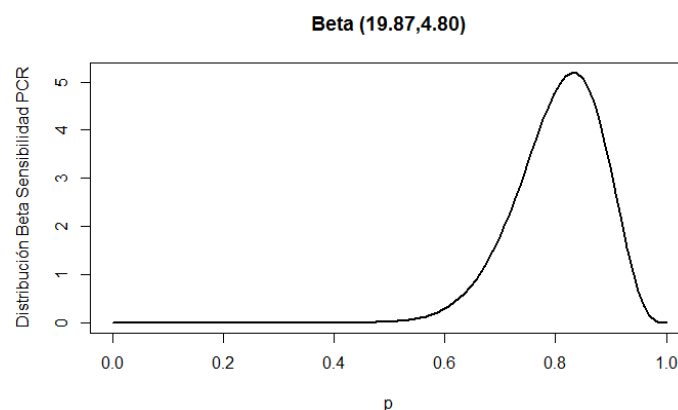
En la **Tabla 6** se recogen los resultados de los artículos encontrados para países africanos tras la búsqueda bibliográfica en la base de datos PubMed según lo indicado en el apartado 3.3.1. de Material y métodos. Se incluye también el valor de prevalencia de estrongiloidiasis en aquellos artículos que la indicaban expresamente.

Artículo	Prevalencia	País	Ref.
Hookworm Infection among School Age Children in Kintampo North Municipality, Ghana: Nutritional Risk Factors and Response to Albendazol Treatment.	Sin datos	Ghana	(Humphries D <i>et al.</i> , 2013)
Urogenital schistosomiasis and soil-transmitted helminthiasis (ATH) in Cameroon: An epidemiological update at Barombi Mbo and Barombi Kotto Crater lakes assesing.	0 %	Camerún	(Campbell SJ <i>et al.</i> , 2017)
Epidemiology of polyparasitism with <i>Taenia solium</i> , schistosomes and soil-transmitted helminths in the co-endemic village of Malanga, Democratic Republic of Congo.	Sin datos	R.D.C.	( Madinga J <i>et al.</i> , 2017)
Helminthiasis: Hookworm Infection Remains a Public Health Problem in Dera District, South Gondar, Ethiopia.	1,1 % (OM)	Etiopía	(Shiferaw MB, 2015)
High prevalence of <i>Strongyloides stercoralis</i> in school-age in a rural highland of north-western Ethiopia: the role of intensive diagnostic work-up.	20,70 % (PCR+OM)	Etiopía	(Amor <i>et al.</i> , 2016)
Hookworm infection is associated with decreased CD4+ T cells count in HIV-infected adult Ugandans	4% (PCR)	Uganda	( Morawski BM <i>et al.</i> , 2017)
Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming.	1 % (PCR)	Kenia	(Easton AV <i>et al.</i> , 2016)
Do shoes reduce hookworm infection in school-aged children on Pemba Island, Zanzibar? A pragmatic trial.	Sin datos	Zanzíbar	(Bird C <i>et al.</i> , 2014)
Diagnosing Polyparasitism in a High-Prevalence Setting in Beira, Mozambique: Detection of Intestinal Parasites in Fecal Samples by Microscopic and Real-Time PCR.	48 % (PCR) 21 % (TB)	Mozambique	(Meurs <i>et al.</i> , 2017)
Infection by intestinal parasites, stunting and anemia in school-age children from Southern Angola.	0.3 % (OM)	Angola	(Oliveira D <i>et al.</i> , 2015)

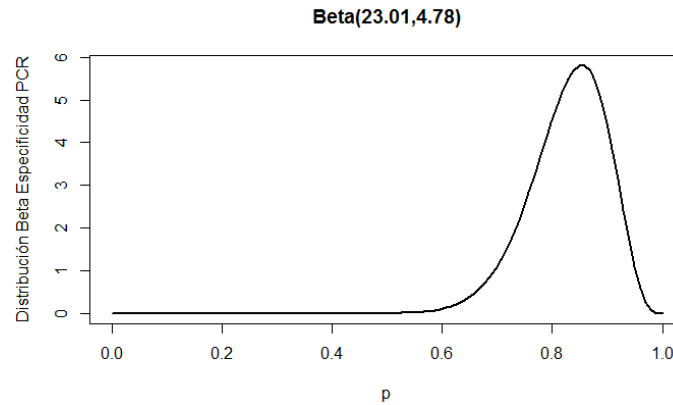
**Tabla 6.** Artículos resultantes de la búsqueda bibliográfica con los datos de prevalencia para los distintos países africanos. MO: Observación microscópica, TB: Técnica de Baermann.

Después de la lectura de los artículos, se consideró que al no haber una técnica *gold standard*, los datos de prevalencia proporcionados no son reales ya que en la mayoría de los casos este resultado únicamente refleja el número de positivos obtenidos por una de las técnicas diagnósticas. Es por este motivo por el que para la elicitación de los parámetros  $\alpha$  y  $\beta$  de las prevalencias de ambas poblaciones se utilizó una distribución uniforme (no informativa) Beta (1,1). Por el contrario, sí tenemos datos fiables de la

sensibilidad y especificidad de las técnicas diagnósticas. En el caso de la PCR a tiempo real se le dio mayor peso al artículo publicado por Saugar *et al.* en 2015 ya que esta técnica se realiza en las mismas condiciones, en el mismo lugar y con el mismo personal que la utilizada para este estudio. La sensibilidad y especificidad descritas en este artículo al compararlos con la combinación de tres técnicas parasitológicas fueron de 93.8% y 86.5%, respectivamente. Estos resultados fueron similares a los descritos por otros autores, Schär *et al.* 2013 (88.9% sensibilidad, 92.7% especificidad), Sharifdini *et al.* 2015 (84.7% sensibilidad y 95.8% especificidad), Martins de Paula *et al.* 2015 (90% sensibilidad, 85.7% especificidad). Con todos estos datos se estableció que los valores de sensibilidad de la PCR se encontraban entre 0.7 y 0.9 con una probabilidad de 0.8 dejando un 0.1 de probabilidad por debajo de 0.7 y un 0.1 por encima de 0.9. Los valores de especificidad de la PCR se encontraban entre 0.7 y 0.9 con una probabilidad de 0.8 y dejando por debajo del 0.7 una probabilidad de 0.05 y por encima de 0.9 una de 0.15. Con estos valores introducidos como cuantiles inferior y superior se obtiene un  $\alpha= 19.87$  y un  $\beta= 4.80$  para la sensibilidad y un  $\alpha= 23.01$  y un  $\beta= 4.78$  para la especificidad. Las funciones de densidad beta con estos dos parámetros aparecen representadas en la **figura 5 y 6**.



**Figura 5.** Función de densidad de la distribución a priori Beta de la sensibilidad de la PCR.



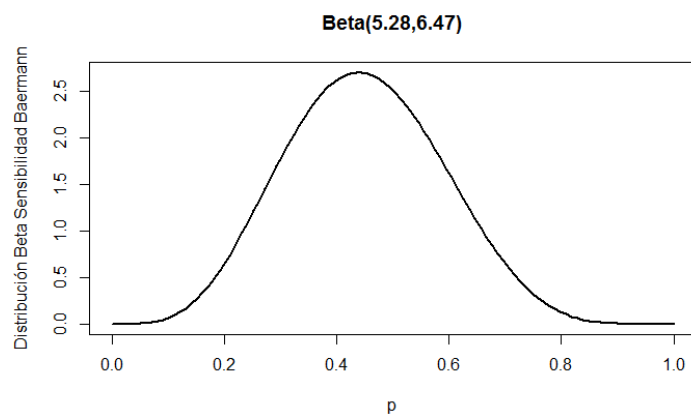
**Figura 6.** Función de densidad de la distribución a priori Beta de la especificidad de la PCR.

En ambos casos el parámetro  $\alpha$  es mayor que el  $\beta$ , por lo que ambas distribuciones se acumulan en la parte de la derecha.

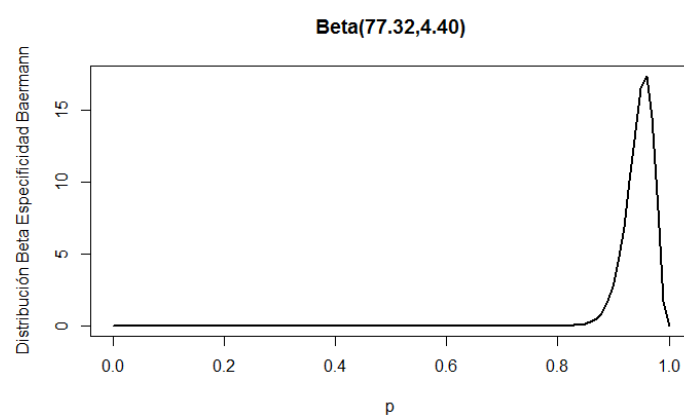
En el caso de la técnica de Baermann en la mayoría de los artículos se habla de casi un 100% de especificidad, ya que la probabilidad de confusión de la larva de *Strongyloides* spp. con otro parásito o artefacto es muy baja. En cuanto a la sensibilidad de la técnica, los datos proporcionados en la bibliografía son muy dispares, ya que además de ser una técnica limitado por las condiciones intrínsecas de la metodología (por ejemplo necesidad de tener la larva viva), depende en gran medida de la experiencia del investigador. De esta manera, se observan datos del 37.5% (Becker *et al.*, 2015), 59.52% (Lau *et al.*, 2005), 60% (Assefa *et al.*, 1991).

Se estableció que los valores de sensibilidad de la técnica Baermann se encontraban entre 0.3 y 0.6 con una probabilidad del 0.7 dejando un 0.15 de probabilidad por debajo de 0.3 y un 0.15 de probabilidad por encima de 0.7. Los valores de especificidad de la técnica Baermann se encontraban entre 0.9 y 0.98 con una probabilidad de 0.9 dejando un 0.05 por debajo de 0.9 y un 0.05 por encima de 0.98.

Con estos valores introducidos como cuantiles inferior y superior se obtiene un  $\alpha= 5.28$  y un  $\beta= 6.47$  para la sensibilidad y un  $\alpha= 77.32$  y un  $\beta= 4.40$  para la especificidad. Las funciones de densidad beta con estos dos parámetros aparecen representadas en la **figura 7 y 8**.



**Figura 7.** Función de densidad de la distribución a priori Beta de la sensibilidad de la técnica Baermann.



**Figura 8.** Función de densidad de la distribución a priori Beta de la especificidad de la técnica Baermann.

En este caso, el parámetro  $\alpha$  de la sensibilidad de la técnica Baermann es ligeramente menor que el  $\beta$ , por lo que la distribución está un poco desplazada hacia la izquierda; sin embargo, en la especificidad ocurre al contrario ( $\alpha > \beta$ ) encontrándose totalmente desplazada a la derecha. Estas dos elicitaciones van a ser utilizadas tanto en el modelo 1 (datos de Angola) como en el modelo “intermedio” para los datos de Etiopía y cuyas a posteriori nos permitirán elicitar el modelo 2, tal y como se recoge en el apartado siguiente.

#### *4.2.1.2. Modelo 2. Modelo con distribuciones a priori obtenidas a partir de los datos del estudio de una población de Etiopía.*

Para la elicitación de las distribuciones a priori de los parámetros de interés en este modelo (dos poblaciones de niños de Angola) es necesario el cálculo de las distribuciones a posteriori de dichos parámetros en un modelo “intermedio”, con una sola población de niños de Etiopía. En este modelo “intermedio” la elicitación de las distribuciones a priori se obtiene de la misma manera que en el modelo 1, es decir a partir

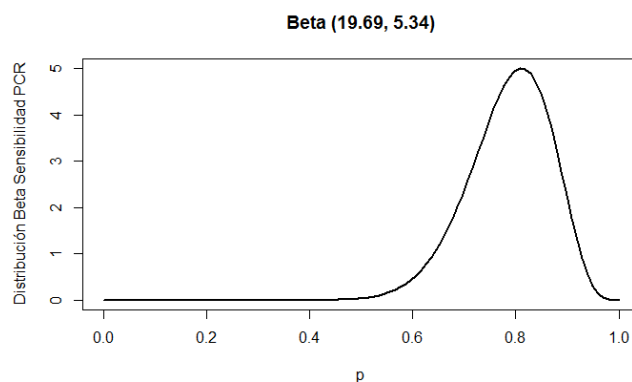
de la información bibliográfica, por lo que los hiperparámetros de los parámetros y respectivas distribuciones son las mismas que en el modelo anterior. Estas elicitaciones se incluyeron dentro del modelo “intermedio” y se calculó su verosimilitud introduciendo los datos de Etiopía para una sola población. Los valores de la distribución a posteriori se muestran a continuación:

	mean	sd	val2.5pc	median	val97.5pc
<b>Sebaermann</b>	0.5738	0.101	0.3848	0.5708	0.7766
<b>Sepcr</b>	0.786	0.08208	0.6087	0.7932	0.9244
<b>Spbaermann</b>	0.939	0.01475	0.9092	0.9394	0.9674
<b>Sppcr</b>	0.9322	0.01975	0.8923	0.9326	0.9695
<b>pi</b>	0.1084	0.03024	0.05733	0.1058	0.175

**Tabla 7.** Resultados del modelo “intermedio” para los parámetros estudiados.

Vamos a utilizar los valores de los dos cuantiles (val 2.5 pc y val 97.5 pc) que aparecen recogidos en la tabla anterior para elicitar los  $\alpha$  y  $\beta$  de los parámetros de interés.

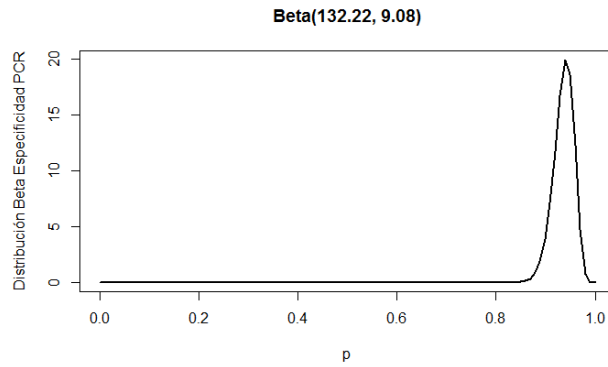
Se obtiene que para la sensibilidad de la técnica PCR el  $\alpha = 19.69$  y el  $\beta = 5.34$  y la función de densidad se muestra en la **figura 9**.



**Figura 9.** Función de densidad de la distribución a priori Beta de la sensibilidad de la PCR.

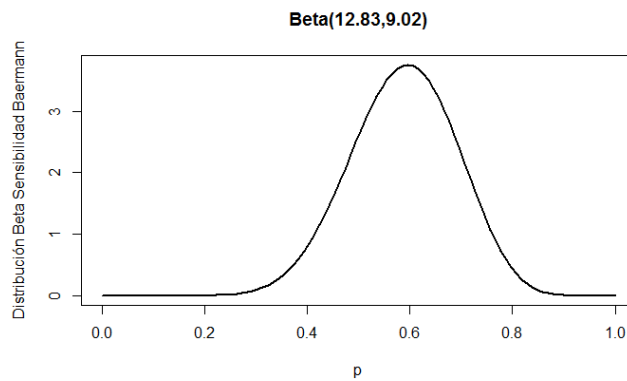
El valor de es  $\alpha > \beta$  por lo que la gráfica aparece desplazada hacia la derecha.

Para la especificidad de la técnica PCR se obtuvo un  $\alpha = 132.22$  y el  $\beta = 9.08$  y la función de densidad se muestra en la **figura 10**.



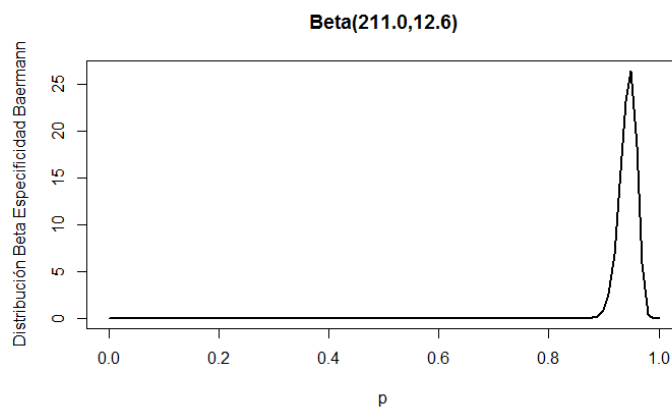
**Figura 10.** Función de densidad a priori Beta de la especificidad de la especificidad de la PCR.

Se obtuvo que para la sensibilidad de la técnica Baermann el  $\alpha = 12.83$  y el  $\beta = 9.02$  y la función de densidad se muestra en la **figura 11**.



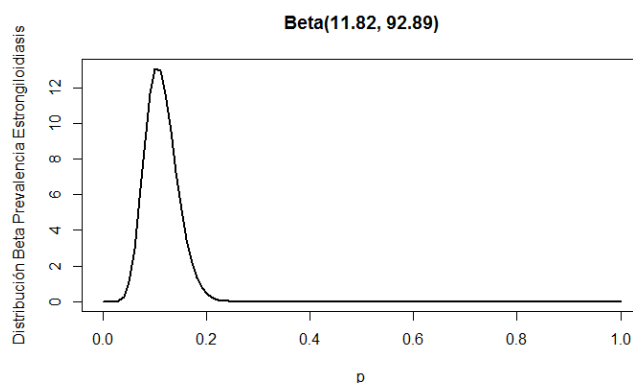
**Figura 11.** Función de densidad de la distribución a priori Beta de la sensibilidad de la técnica Baermann.

Se obtuvo que para la especificidad de la técnica Baermann el  $\alpha = 211.0$  y el  $\beta = 12.6$  y la función de densidad se muestra en la **figura 12**.



**Figura 12.** Función de densidad de la distribución a priori Beta de la especificidad de la técnica Baermann.

Se obtuvo para la prevalencia de la estrogiloidiasis un  $\alpha = 11.82$  y el  $\beta = 92.89$  y la función de densidad se muestra en la **figura 13**.



**Figura 13.** Función de densidad de la distribución a priori Beta de la prevalencia.

*4.2.1.3. Modelo 3. Modelo con distribuciones a priori para las sensibilidades y especificidades a partir de búsqueda bibliográfica y para las prevalencias a partir de los datos de Etiopía.*

La elicitación de los valores  $\alpha$  y  $\beta$  para los parámetros de sensibilidad y especificidad de las técnicas diagnósticas son los mismos que los obtenidos para el modelo 1. Sin embargo, para la prevalencia se utilizaron los valores  $\alpha$  y  $\beta$  obtenidos del modelo 2.

En la **tabla 8** aparece recogido una comparación de los  $\alpha$  y  $\beta$  de los parámetros de estudio obtenidos mediante el método bibliográfico y con el estudio previo de Etiopía.

	Modelo 1		Modelo 2		Modelo 3	
	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$
SePCR	19,87	4,8	19,69	5,34	19,87	4,8
SpPCR	23,01	4,78	132,22	9,08	23,01	4,78
SeBaermann	5,28	6,47	12,83	9,02	5,28	6,47
SpBaermann	77,32	4,4	211	12,06	77,32	4,4
pi1	1	1	11,82	92,89	11,82	92,89
pi2	1	1	11,82	92,89	11,82	92,89

**Tabla 8.** Valores de los parámetros  $\alpha$  y  $\beta$  de las distribuciones Beta a priori.

Cabe destacar que son los valores  $\alpha$  y  $\beta$  de especificidad los que difieren más entre ambas metodologías. La utilización de los datos del estudio realizado en Etiopía

proporcionó una distribución a priori de la prevalencia informativa, a diferencia de lo obtenido por el método bibliográfico.

#### 4.2.2. Distribuciones a posteriori

##### 4.2.2.1. Modelo 1. Modelo con distribuciones a priori obtenidas a partir de búsqueda bibliográfica.

En este caso utilizamos de información a priori los  $\alpha$  y  $\beta$  previamente elicitados para los parámetros de sensibilidad, especificidad de ambas técnicas y como distribución a priori de la prevalencia en ambas poblaciones de estudio una distribución uniforme no informativa, B(1,1). En el Anexo 1 se muestra el programa que permite realizar los métodos MCMC tras la especificación de los parámetros de interés, sus distribuciones y los datos recogidos en ambas poblaciones. Los resultados obtenidos muestran que la sensibilidad para la técnica Baermann se encuentra con un 95% de probabilidad entre 0.28-0.69, mientras que la sensibilidad de la PCR entre 0.63-0.92. La especificidad de Baermann con un 95% de probabilidad fue de 0.92-0.98 y la de la PCR 0.81 y 0.94. La prevalencia de la enfermedad para la población 1 está entre 0.08 y 0.29 y para la población 2 entre 0.03 y 0.24. En todos los parámetros la media y la mediana son similares.

La descripción de las muestras obtenidas de las distribuciones a posteriori de los parámetros de interés se muestran en la siguiente tabla:

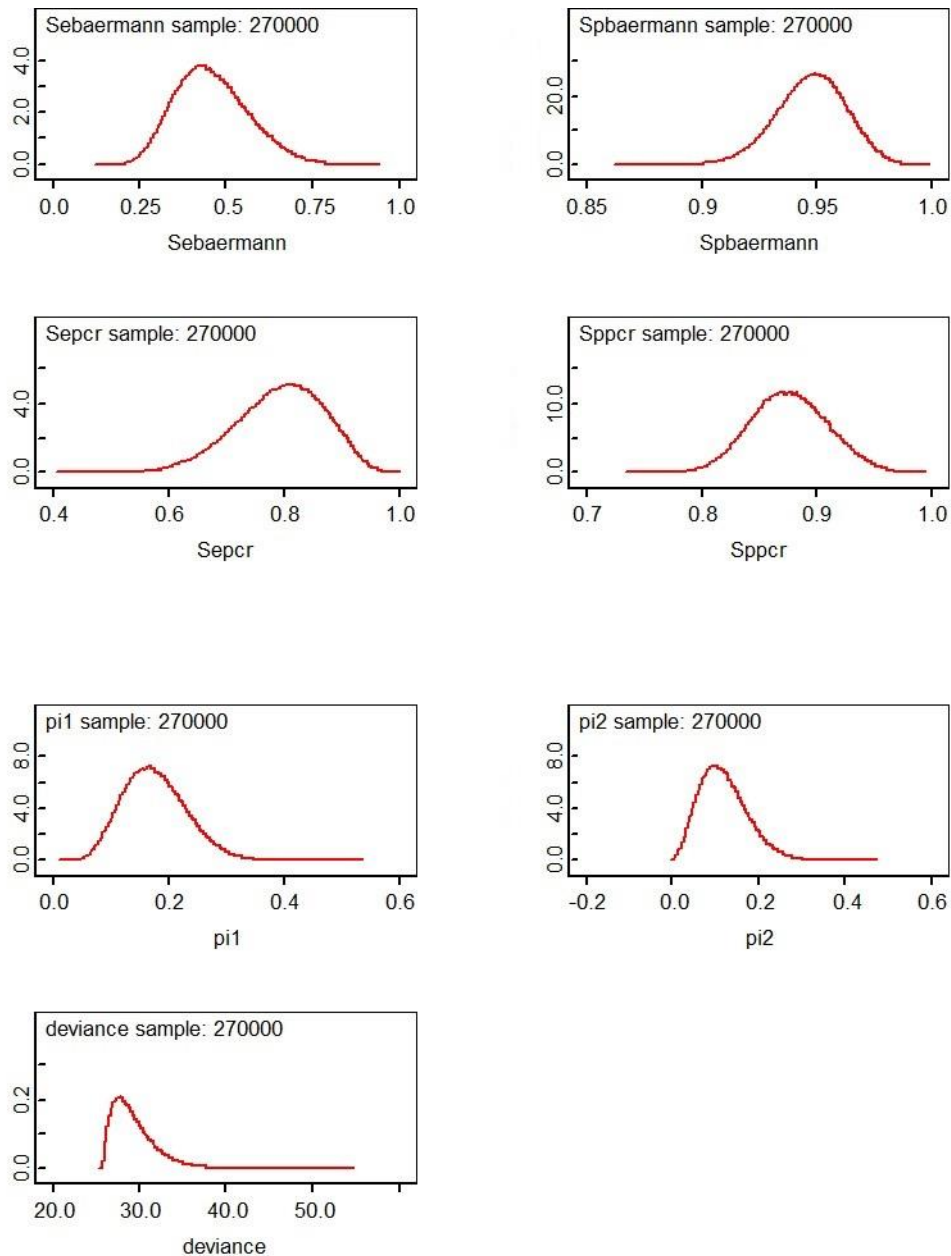
	mean	sd	val2.5pc	median	val97.5pc
<b>Sebaermann</b>	0.459	0.1065	0.2763	0.4499	0.6889
<b>SePCR</b>	0.7925	0.07716	0.6272	0.7985	0.9241
<b>Spbaermann</b>	0.9476	0.0153	0.9156	0.9483	0.9757
<b>SpPCR</b>	0.8766	0.03336	0.813	0.8759	0.9428
<b>pi1</b>	0.1764	0.05529	0.08079	0.1724	0.2947
<b>pi2</b>	0.1188	0.05588	0.02918	0.1126	0.2441

Deviance information

	Dbar	Dhat	DIC	pD
<b>y</b>	15.83	13.89	17.78	1.945
<b>z</b>	13.5	12.01	14.99	1.489
<b>total</b>	29.33	25.9	32.77	3.434

**Tabla 9.** Resultados del modelo 1 para los parámetros estudiados.

Las funciones de densidad a posteriori de los parámetros se muestran en la **figura 14**.



**Figura 14.** Funciones de densidad a posteriori de los parámetros del modelo 1.

*4.2.2.2. Modelo 2. Modelo con distribuciones a priori obtenidas a partir de los datos del estudio de una población de Etiopía.*

La sensibilidad para la técnica Baermann se encuentra con un 95% de probabilidad entre 0.36-0.69, mientras que la sensibilidad de la PCR entre 0.68-0.93. La especificidad de Baermann con un 95% de probabilidad fue de 0.92-0.97 y la de la PCR 0.85 y 0.94. La prevalencia de la enfermedad para la población 1 está entre 0.09 y 0.2 y para la población 2 entre 0.06 y 0.17. En todos los parámetros la media coincide con la mediana. Los resultados obtenidos de la distribución a posteriori del modelo se muestran en la **tabla 10**.

	mean	sd	val2.5pc	median	val97.5pc
Sebaermann	0.5154	0.0855	0.3572	0.5123	0.6914
Sepcr	0.8175	0.06555	0.6754	0.823	0.9285
Spbaermann	0.9466	0.01133	0.9227	0.9472	0.9673
Sppcr	0.899	0.02095	0.857	0.8994	0.9391
pi1	0.1415	0.02812	0.09026	0.1402	0.2
pi2	0.111	0.02619	0.06481	0.1093	0.1671

Deviance information

	Dbar	Dhat	DIC	pD
y	16.87	15.54	18.2	1.332
z	13.01	12.2	13.82	0.8081
total	29.88	27.74	32.02	2.14

Tabla 10. Resultados del modelo 2 para los parámetros estudiados.

Las funciones de densidad a posteriori de los parámetros se muestran en la **figura**

15.

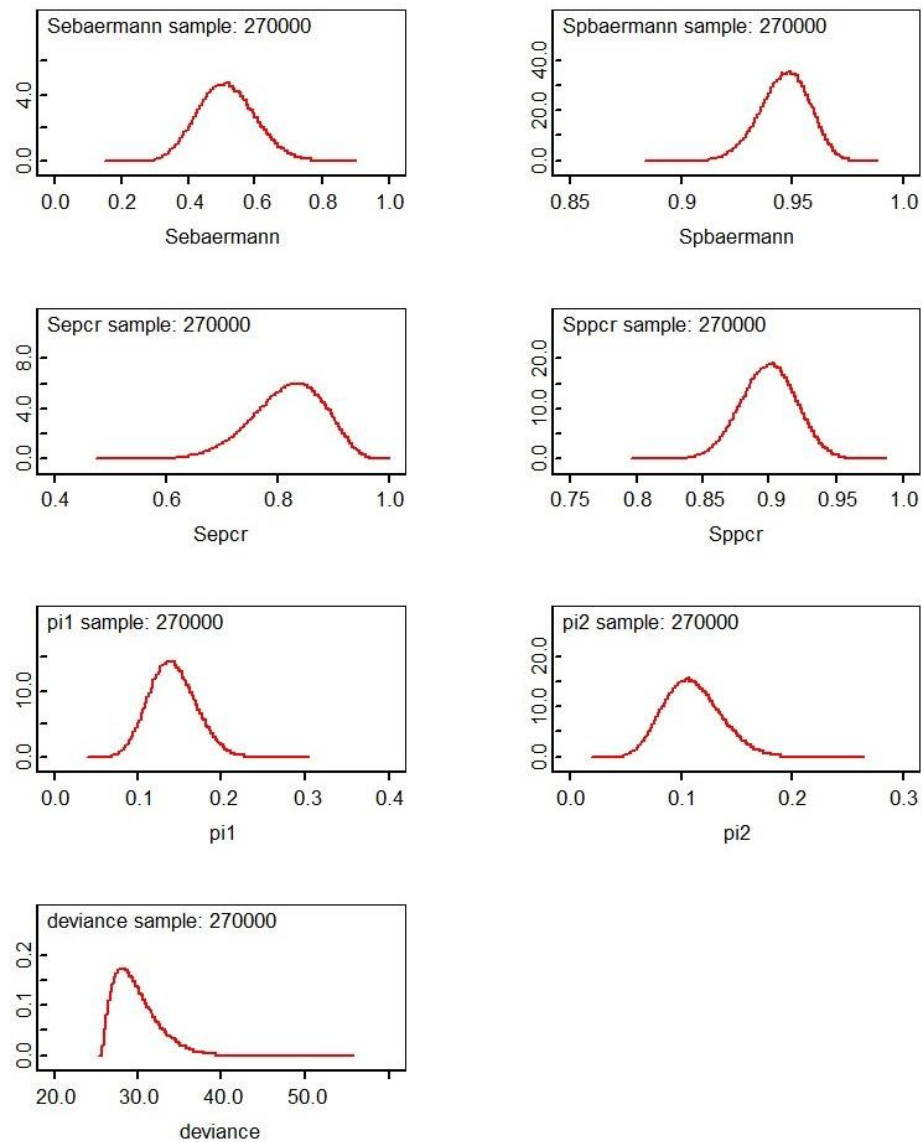


Figura 15. Funciones de densidad a posteriori de los parámetros del modelo 2.

4.2.2.3. *Modelo 3. Modelo con distribuciones a priori para las sensibilidades y especificidades a partir de búsqueda bibliográfica y para las prevalencias a partir de los datos de Etiopía.*

La sensibilidad para la técnica Baermann se encuentra con un 95% de probabilidad entre 0.32-0.72, mientras que la sensibilidad de la PCR entre 0.66-0.93. La especificidad de Baermann con un 95% de probabilidad fue de 0.91-0.97 y la de la PCR 0.81 y 0.91. La prevalencia de la enfermedad para la población 1 está entre 0.08 y 0.19 y para la población 2 entre 0.06 y 0.16. En todos los parámetros la media coincide con la mediana.

Los resultados obtenidos a posteriori para los parámetros del modelo se muestran en la **Tabla 11**:

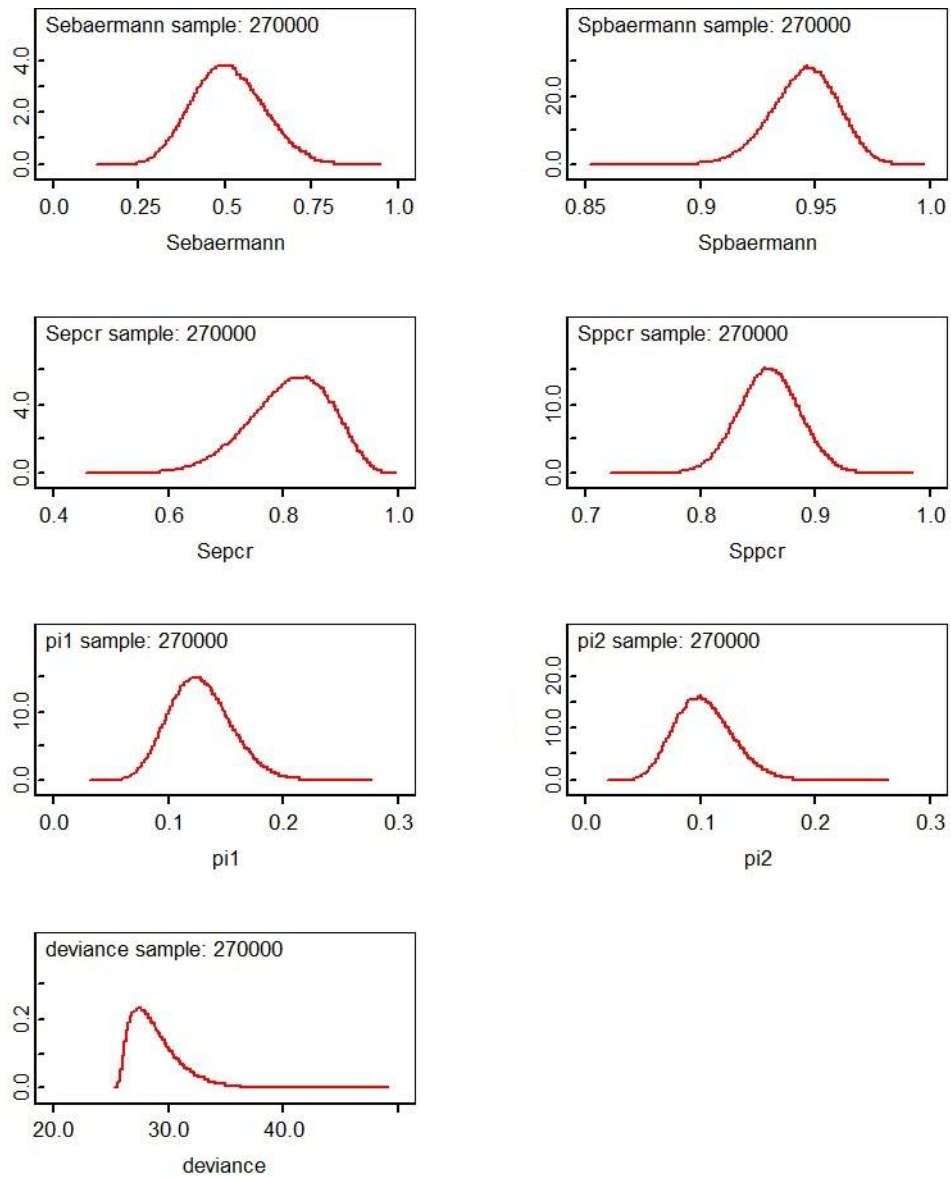
	<b>mean</b>	<b>sd</b>	<b>val2.5pc</b>	<b>median</b>	<b>val97.5pc</b>
<b>Sebaermann</b>	0.5089	0.1023	0.3207	0.5051	0.7181
<b>SePCR</b>	0.8108	0.0713	0.6563	0.8169	0.931
<b>Spbaermann</b>	0.9453	0.01463	0.9145	0.9461	0.9719
<b>SpPCR</b>	0.8601	0.02648	0.8077	0.8602	0.9117
<b>pi1</b>	0.1283	0.02703	0.07983	0.1267	0.1855
<b>pi2</b>	0.1038	0.02531	0.05944	0.102	0.1581

Deviance information

	<b>Dbar</b>	<b>Dhat</b>	<b>DIC</b>	<b>pD</b>
<b>y</b>	15.77	14.06	17.47	1.707
<b>z</b>	13.15	12.16	14.14	0.9892
<b>total</b>	28.92	26.22	31.62	2.697

**Tabla 11.** Resultados del modelo 3 para los parámetros estudiados.

Las funciones de densidad a posteriori de los parámetros se muestran a continuación:



**Figura 16.** Funciones de densidad a posteriori de los parámetros del modelo 3.

A continuación se muestra una tabla resumen (**tabla 12**) con los valores  $\alpha$  y  $\beta$  de la distribución a priori así como los valores obtenidos para cada uno de los modelos de la distribución a posteriori.

		Sensibilidad PCR	Sensibilidad Baermann	Especificidad PCR	Especificidad Baermann	Prevalencia 1	Prevalencia 2
<b>Modelo 1.</b> Modelo con distribuciones a priori obtenidas a partir de búsqueda bibliográfica.	<b>A priori</b> $\alpha; \beta$	19,87; 4,8	5,28; 6,47	23,01; 4,78	77,32; 4,4	1; 1	1; 1
	<b>A posteriori</b> Media $\pm$ DE IP95%	0,7925 $\pm$ 0,077 (0,6272-0,9241)	0,4590 $\pm$ 0,1065 (0,2763-0,6889)	0,8766 $\pm$ 0,0333 (0,8130-0,9428)	0,9476 $\pm$ 0,0153 (0,9156-0,9757)	0,1764 $\pm$ 0,0553 (0,0808-0,2947)	0,1188 $\pm$ 0,0559 (0,0292-0,2441)
<b>Modelo 2.</b> Modelo con distribuciones a priori obtenidas a partir de los datos del estudio de una población de Etiopía.	<b>A priori</b> $\alpha; \beta$	19,69; 5,34	12,83; 9,02	132,22; 9,08	211; 12,06	11,82; 92,89	11,82; 92,89
	<b>A posteriori</b> Media $\pm$ DE IP95%	0,8175 $\pm$ 0,0656 (0,6754-0,9285)	0,5154 $\pm$ 0,0855 (0,3572-0,6914)	0,8990 $\pm$ 0,0210 (0,8570-0,9391)	0,9466 $\pm$ 0,0113 (0,9227-0,9673)	0,1415 $\pm$ 0,0281 (0,0903-0,2000)	0,1110 $\pm$ 0,0262 (0,0648-0,1671)
<b>Modelo 3.</b> Modelo con distribuciones a priori para las sensibilidades y especificidades a partir de búsqueda bibliográfica y para las prevalencias a partir de los datos de Etiopía.	<b>A priori</b> $\alpha; \beta$	19,87; 4,8	5,28; 6,47	23,01; 4,78	77,32; 4,4	11,82; 92,89	11,82; 92,89
	<b>A posteriori</b> Media $\pm$ DE IP95%	0,8108 $\pm$ 0,0713 (0,6563-0,9310)	0,5089 $\pm$ 0,1023 (0,3207-0,7181)	0,8601 $\pm$ 0,0265 (0,8077-0,9117)	0,9453 $\pm$ 0,0146 (0,9145-0,9719)	0,1283 $\pm$ 0,0270 (0,0798-0,1855)	0,1038 $\pm$ 0,0253 (0,0594-0,1581)

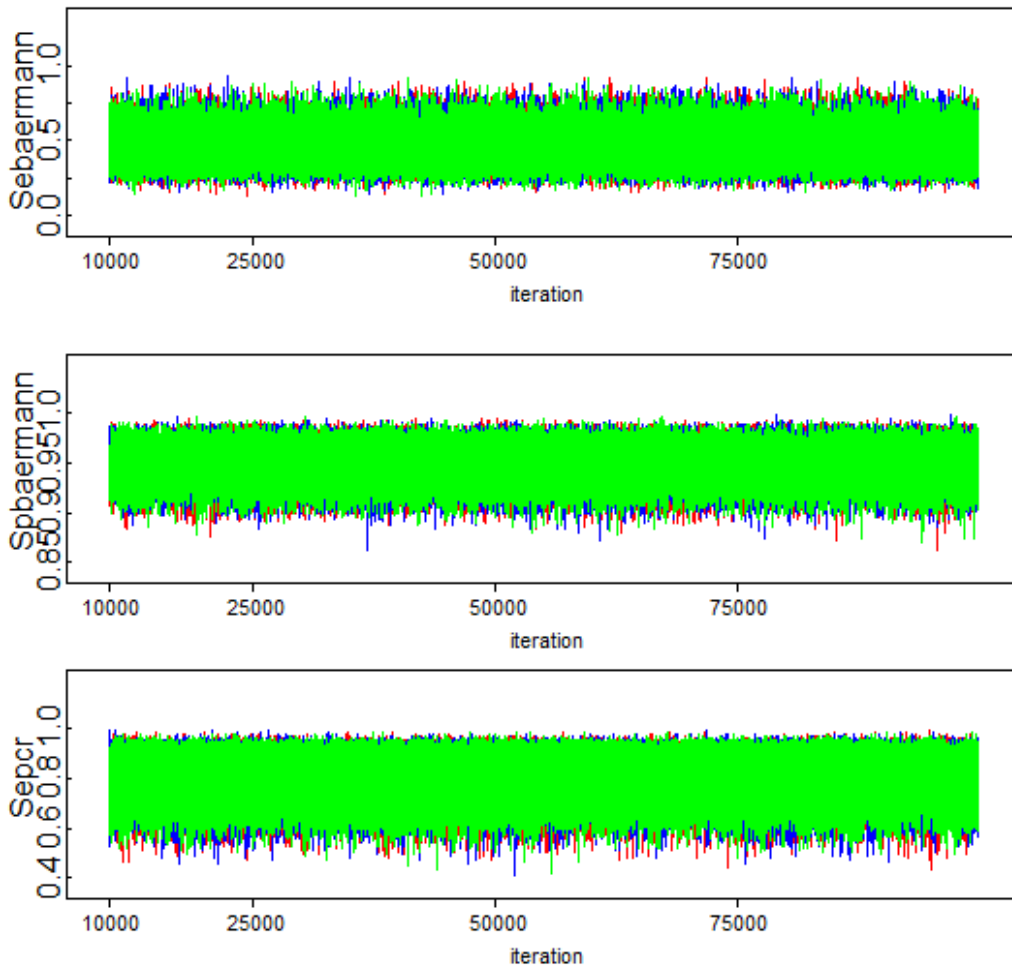
**Tabla 12.** Valores  $\alpha$  y  $\beta$  de la distribución a priori y la media más desviación estándar (D.E.) de la sensibilidad de la PCR y del Baermann, la especificidad de la PCR y Baermann y prevalencias de población de niños descalzos (1) y de niños calzados (2) de la distribución a posteriori.

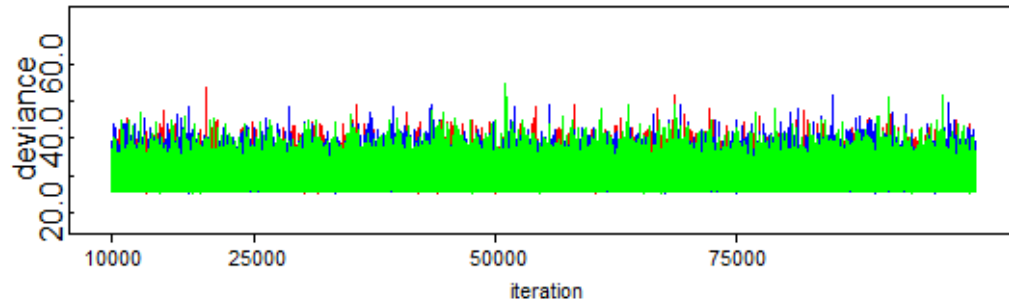
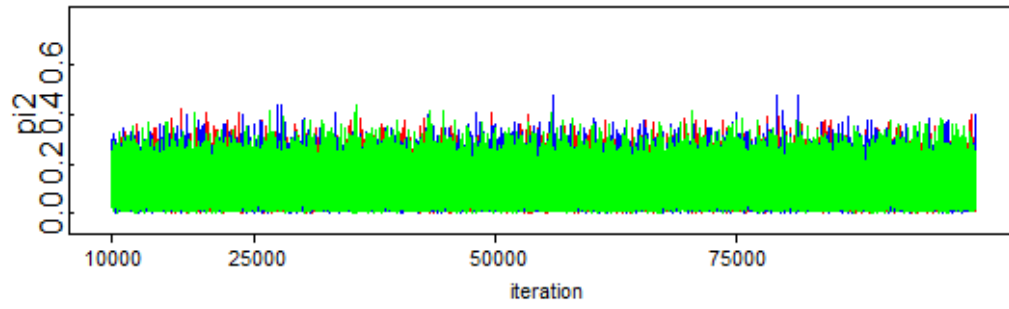
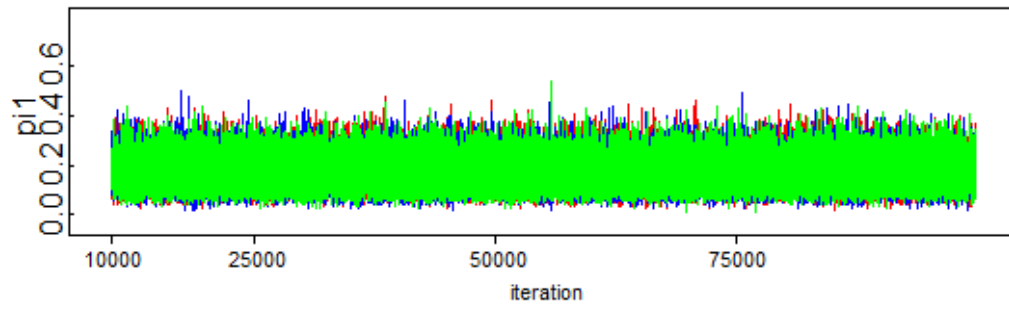
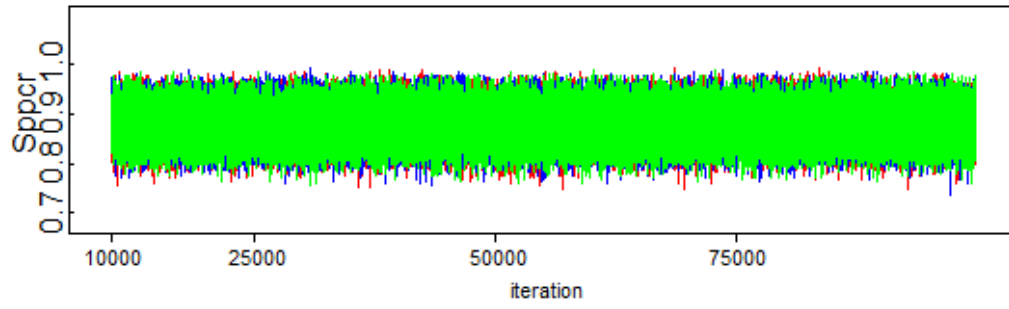
### 4.3. Estudio de convergencia para los tres modelos.

#### 4.3.1. Estudio de convergencia: Trazas

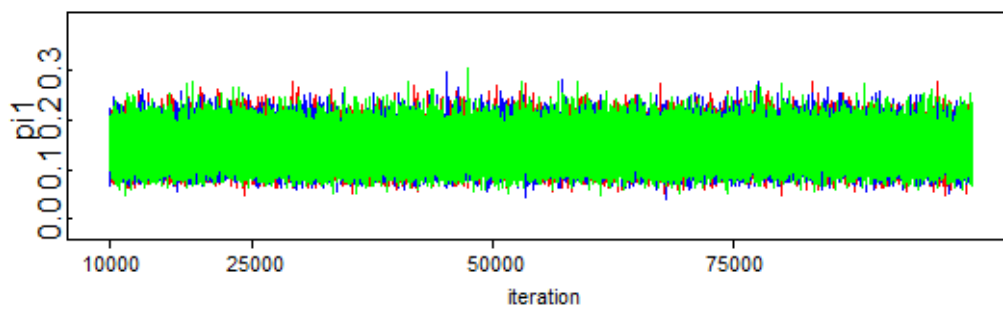
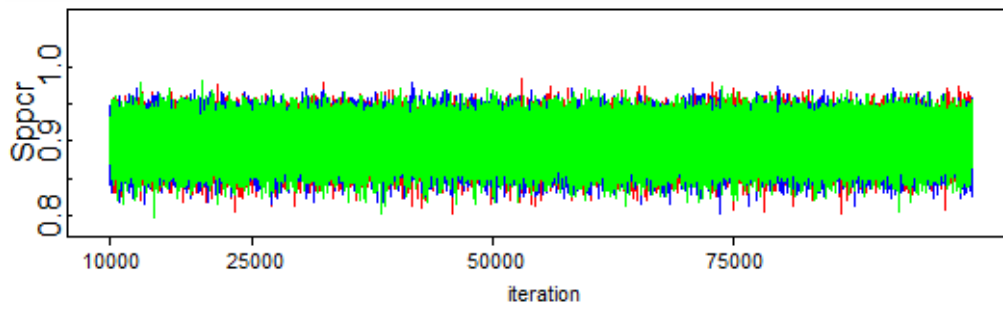
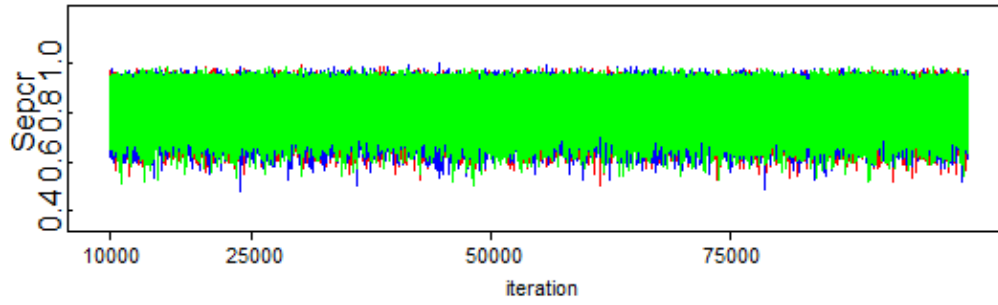
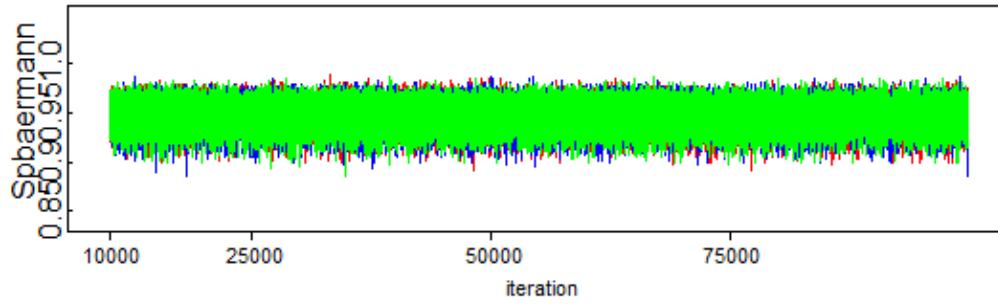
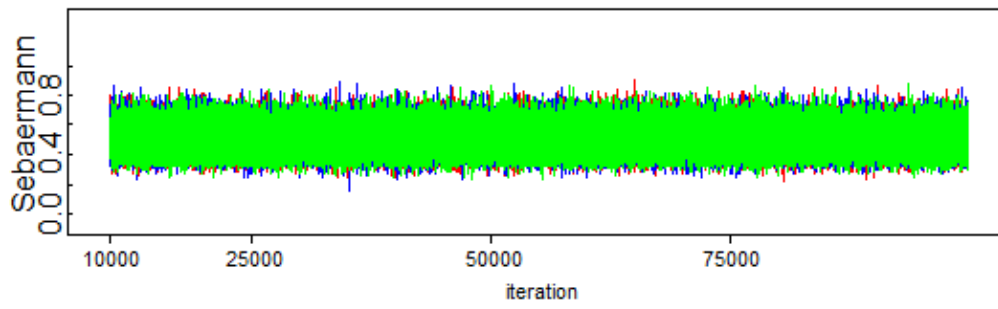
En todas las trazas se observa una buena convergencia en cada cadena para todos los parámetros analizados observándose una estabilidad tras el proceso de *burn-in* de las primeras iteraciones. Por consiguiente, gráficamente se puede concluir que todas las cadenas generadas convergen, para los tres modelos planteados en el trabajo, tal y como se muestra en los siguientes apartados.

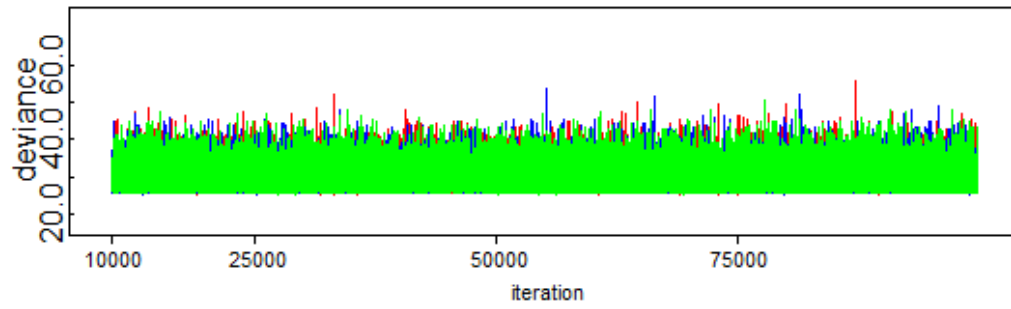
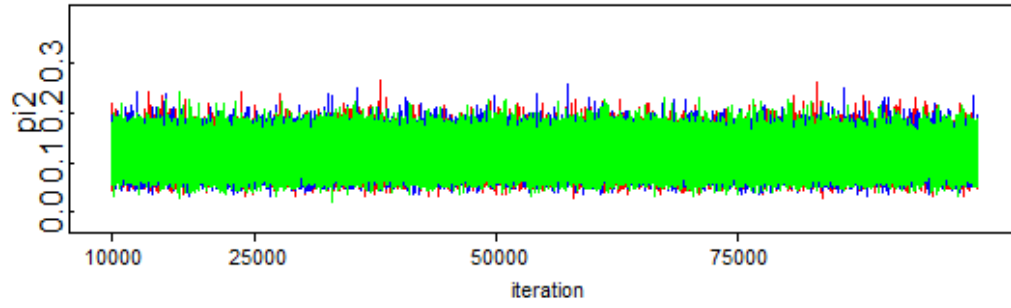
##### 4.3.1.1. Trazas del modelo 1



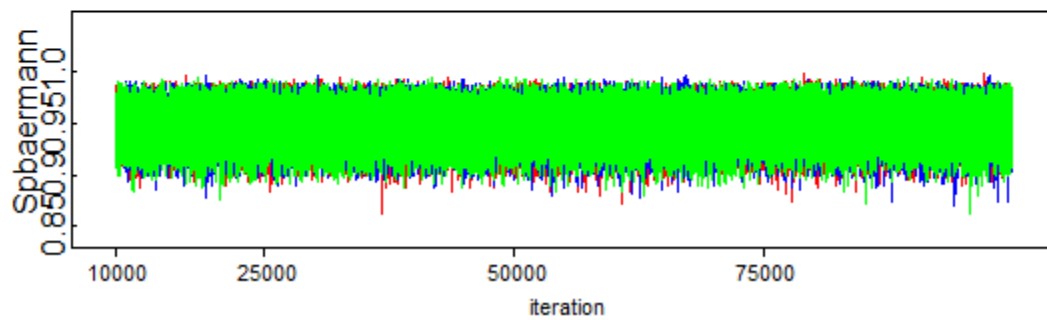
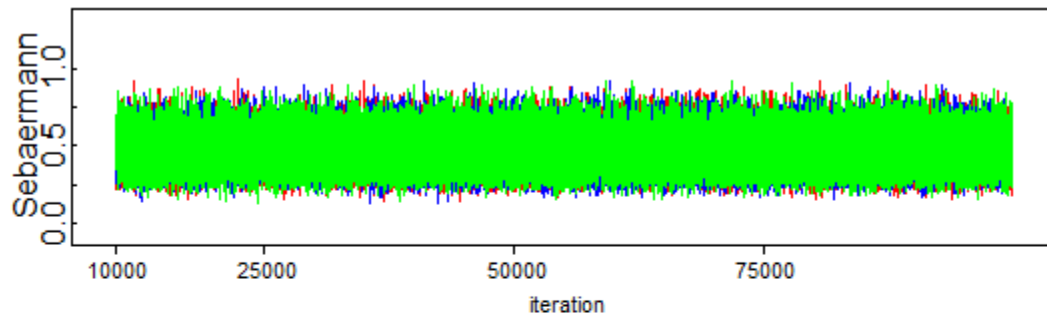


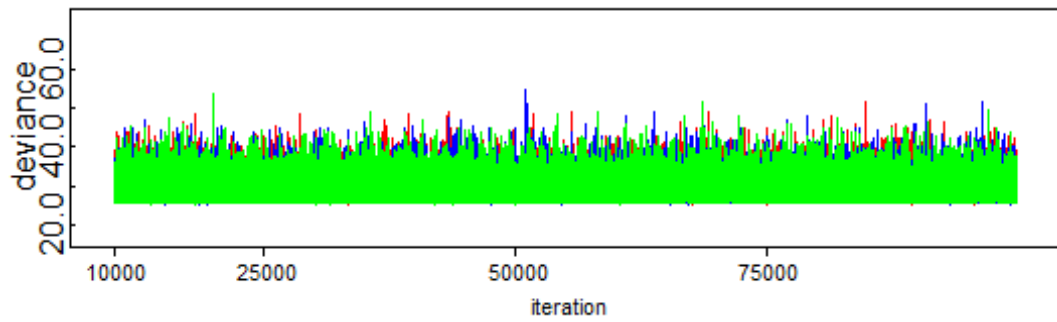
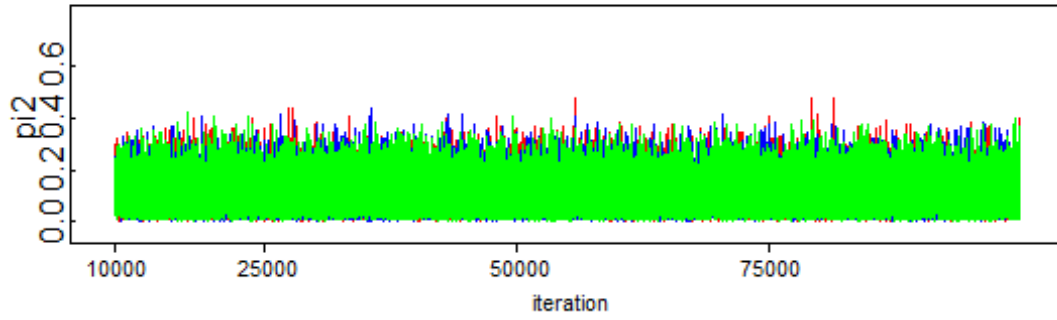
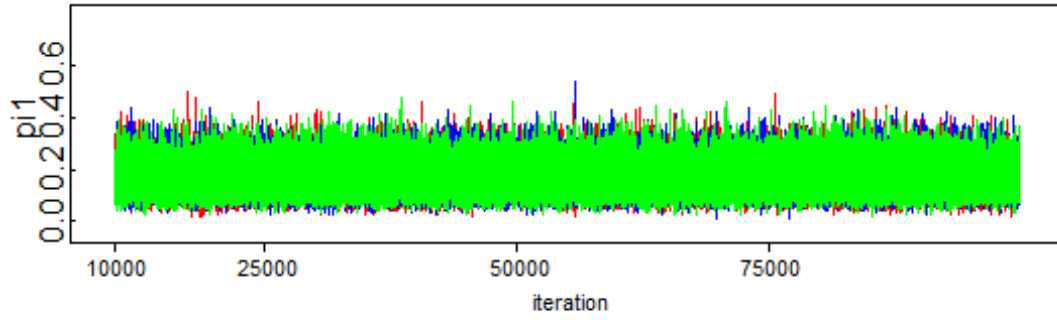
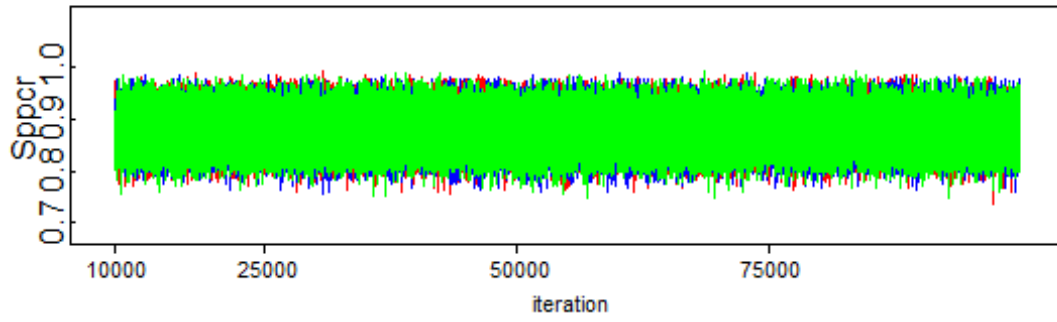
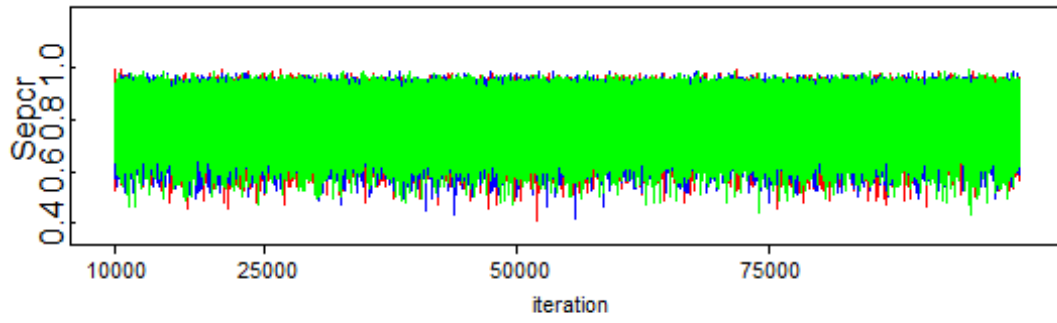
4.3.1.2. Trazas del modelo 2





#### 4.3.1.3. Trazas del modelo 3





### 4.3.2. Estudio de convergencia: criterio Geweke y PSRF

#### 4.3.2.1. Criterio Geweke y PSRF del modelo 1

##### Criterio Geweke

Se observa convergencia entre las cadenas ya que en todos los casos el criterio de Geweke toma valores en el intervalo  $(-1,96; 1,96)$  salvo para la primera cadena en la que los parámetros Sepcr y pi2 no muestran dicha convergencia. Para determinar si finalmente las cadenas convergen de forma conjunta hay que determinar el valor del criterio PSRF que se muestra a continuación.

##### Cadena 1

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
0,715	2,1565	-1,2523	-1,0989	-1,3399	-17.574

##### Cadena 2

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
0,06478	-0,52196	1,04988	0,62762	0,79494	1,10623

##### Cadena 3

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
0,1112	0,4411	0,3944	-0,6665	-0,2177	-1,2135

**Tabla 13.** Valores del criterio Geweke para los distintos parámetros en las tres cadenas del modelo 1.

##### Criterio PSRF

	Point est.	Upper C.I.
Sebaermann	1	1
Sepcr	1	1
Spbaermann	1	1
Sppcr	1	1
pi1	1	1
pi2	1	1

Multivariate psrf	1
-------------------	---

**Tabla 14.** Valores del criterio PSRF para los distintos parámetros del modelo 1.

De forma global se puede concluir que las cadenas convergen ya que con el criterio PSRF se obtiene un valor 1 para el todos los parámetros individuales así como para el multivariado.

#### 4.3.2.2. Criterio Geweke y PSRF del modelo 2

##### Criterio Geweke

Se observa convergencia entre las cadenas ya que en todos los casos el criterio de Geweke toma valores en el intervalo (-1,96 - 1,96) excepto la Sepcr en la cadenas 1 y 3 donde no se encuentra dicha convergencia. Es necesario evaluar el criterio PSRF en estos parámetros para poder concluir si finalmente convergen de forma conjunta.

##### Cadena 1

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
-0,3444	2,3707	-0,8213	-0,5627	-1,4649	0,4163

##### Cadena 2

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
0,49168	1,60747	0.23958	0,01163	-0,7286	0,11265

##### Cadena 3

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
0,10500	2,02722	-1,57659	0,60125	-0,03736	0,07982

**Tabla 15.** Valores del criterio Geweke para los distintos parámetros en las tres cadenas del modelo 2.

##### Criterio PSRF

	Point est.	Upper C.I.
<b>Sebaermann</b>	1	1
<b>Sepcr</b>	1	1
<b>Spbaermann</b>	1	1
<b>Sppcr</b>	1	1
<b>pi1</b>	1	1
<b>pi2</b>	1	1
<b>Multivariate psrf</b>	1	

**Tabla 16.** Valores del criterio PSRF para los distintos parámetros del modelo 2.

El criterio PSRF nos indica que de forma global el parámetro Sepcr converge. Asimismo, las cadenas mediante las que se generan las muestras de valores a posteriori de los parámetros convergen.

#### 4.3.2.3. Criterio Geweke y PSRF del modelo 3

Se observa convergencia entre las cadenas ya que en todos los casos el criterio de Geweke toma valores en el intervalo (-1,96 - 1,96) excepto la Sepcr en la cadenas 1 y la Spbaermann en la cadena 2 donde no se encuentra dicha convergencia.

#### Criterio Geweke

Cadena 1

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
-0,3429	2,3149	-1,4792	1,0264	-0,8465	-1,3286

Cadena 2

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
0,7359	0,8000	3,7329	0,9674	0,7777	1,4219

Cadena 3

Sebaermann	Sepcr	Spbaermann	Sppcr	pi1	pi2
-0,6033	0,8279	-0,3337	1,1309	0,1237	-0,6237

**Tabla 17.** Valores del criterio Geweke para los distintos parámetros en las tres cadenas del modelo 3.

#### Criterio PSRF

	Point est.	Upper C.I.
<b>Sebaermann</b>	1	1
<b>Sepcr</b>	1	1
<b>Spbaermann</b>	1	1
<b>Sppcr</b>	1	1
<b>pi1</b>	1	1
<b>pi2</b>	1	1

<b>Multivariate psrf</b>	1
--------------------------	---

**Tabla 18.** Valores del criterio PSRF para los distintos parámetros del modelo 3.

Con el criterio PSRF podemos concluir que las cadenas generadas convergen para todos los parámetros, incluso para Sepcr y Spbaermann, ya que su valor es 1 en todos los casos.

#### 4.4. Comparación de modelos

Los resultados de los tres modelos son muy similares tanto de los valores de los parámetros que aparecen en las tablas como en el valor de la deviance. El modelo 1, que es el que menor información incorporaba es el que mayor valor de DIC tiene, por lo que lo descartamos. En cuanto al modelo 2 y 3 son los que presentan resultados muy similares siendo el menor valor del DIC para el modelo 3. Sin embargo, al evaluar la precisión de los intervalos de probabilidad observamos que éstos son más estrechos en el modelo 2, por lo que se escogió este modelo como el modelo que mejor se ajusta a nuestros datos.

#### 4.5. Conclusiones:

El modelo 2, en donde la elicitación se realiza basándose en un modelo “intermedio” con los datos de la población en Etiopía, es el que se ajusta mejor a nuestros datos según la combinación del criterio DIC y precisión de los intervalos de probabilidad; este hecho es coherente ya que para el modelo “intermedio” se utiliza como información a priori la información bibliográfica y por tanto, el modelo 2 incorpora toda la información disponible, es decir, la información bibliográfica y la de los datos de la población de Etiopía.

La sensibilidad para la técnica Baermann y su intervalo de probabilidad al 95% fue 51.54% (35.72%-69.14%), mientras que para PCRrt fue de 81.75% (67.54%-92.85%). La especificidad para Baermann fue de 94.66% (92.27%-96.73%) y para PCRrt 89.90% (85.70%-93.91%). Las prevalencias estimadas fueron 14.15% (9.02%-20.00%) para la población que camina descalza y 11.10% (6.48%-16.71%) para la que camina calzada, siendo este dato coherente con el mecanismo de transmisión de esta parasitosis. Estos resultados están en consonancia con las distintas publicaciones que abordan esta misma problemática, sin embargo, con la Estadística Bayesiana se ha podido estimar con mayor precisión tanto la sensibilidad y especificidad de ambas técnicas, como la prevalencia de la enfermedad en una zona donde los estudios eran escasos y la situación totalmente desconocida. Esto supone un reto ya que a pesar de que existen estudios similares que tratan de estimar la prevalencia de la estrongiloidiasis, en algunos casos los

intervalos resultantes eran menos precisos (Morawski BN et al., 2017) o similares (Amor et al., 2016).

En este trabajo, la Estadística Bayesiana ha permitido incorporar la información de otros estudios y conjugarlo con los datos recogidos inicialmente en población infantil de Etiopía y posteriormente en población infantil de Cubal (Angola). Este procedimiento permite ajustar con mayor precisión todos los parámetros de interés, con especial atención a la prevalencia de la estrongiloidiasis en ausencia de *gold standard*.

## Anexo 1

### Programa 1. Elicitación de los parámetros mediante bibliografía

```
#Elicitación de los parámetros alfa y beta de la sensibilidad de la
técnica molecular

#utilizando dos cuantiles

#lo primero es instalar el paquete LearnBayes

cuantil1=list(p=0.1,x=0.7)

cuantil2=list(p=0.90,x=0.9)

beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución

grill <- seq(0,1,length = 100)

# valores de la función de densidad de la beta(a,b)

valores<- dbeta(grill,19.87,4.80)

plot(grill, valores,

      main="Beta (19.87,4.80)",type="l",xlab="p",ylab="Distribución Beta
Sensibilidad PCR",lwd=2)

#Elicitación de los parámetros alfa y beta de la especificidad de la
técnica molecular

cuantil1=list(p=0.05,x=0.7)

cuantil2=list(p=0.85,x=0.9)

beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución

grill <- seq(0,1,length = 100)

# valores de la función de densidad de la beta(a,b)

valores<- dbeta(grill,23.01,4.78)

plot(grill, valores,

      main="Beta(23.01,4.78)",type="l",xlab="p",ylab="Distribución Beta
Especificidad PCR",lwd=2)
```

```

#Elicitación de los parámetros alfa y beta de la sensibilidad de la
técnica parasitológica

cuantil1=list(p=0.15,x=0.3)
cuantil2=list(p=0.85,x=0.6)
beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)

# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,5.28,6.47)
plot(grill, valores,
      main="Beta(5.28,6.47)",type="l",xlab="p",ylab="Distribución Beta
Sensibilidad Baermann",lwd=2)

#Elicitación de los parámetros alfa y beta de la especificidad de la
técnica parasitológica

cuantil1=list(p=0.05,x=0.9)
cuantil2=list(p=0.95,x=0.98)
beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)

# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,77.32,4.40)
plot(grill, valores,
      main="Beta(77.32,4.40)",type="l",xlab="p",ylab="Distribución Beta
Especificidad Baermann",lwd=2)

```

## **Programa 2. Elicitación de los parámetros mediante los datos obtenidos de la a posteriori de Etiopía**

### **Primero programa 2.1. Modelo “intermedio” (datos de Etopía)**

```

cat("
      model

```

```

{
y1[1:Q, 1:Q] ~ dmulti(p1[1:Q, 1:Q], n1)
p1[1,1] <- pi*Sepcr*Sebaermann + (1-pi)*(1-Sppcr)*(1-Spbaermann)
p1[1,2] <- pi*Sepcr*(1-Sebaermann) + (1-pi)*(1-Sppcr)*Spbaermann
p1[2,1] <- pi*(1-Sepcr)*Sebaermann + (1-pi)*Sppcr*(1-Spbaermann)
p1[2,2] <- 1-p1[1,1]- p1[1,2]-p1[2,1]
pi ~ dbeta(1, 1)
Sepcr ~ dbeta(19.87,4.80)
Sppcr ~ dbeta(23.01,4.78)
Sebaermann ~ dbeta(5.28,6.47)
Spbaermann ~ dbeta(77.32,4.40)
}
",
file = "modelo.txt")

#Datos
Q=2
n1=393
y1=structure(.Data=c(21,26,32,314),.Dim=c(2,2))
datos <- list ("Q","n1","y1")
iniciales <- function()
{list(pi=runif(1),Sepcr=runif(1),Sppcr=runif(1),Sebaermann=runif(1),Spbaermann=runif(1))}
parametros <- c("pi","Sepcr","Sppcr","Sebaermann","Spbaermann")
library(R2OpenBUGS)
modelo <- bugs(data = datos, inits = iniciales,
               parameters.to.save = parametros,
               model.file = "modelo.txt", n.chains = 3,
               n.iter=100000, n.burnin = 10000,

```

```

working.directory = Taruta,coda=TRUE,
clearWD = TRUE, debug = TRUE)

#Utilizando las a posteriori de este modelo

#Elicitación de los parámetros alfa y beta de la sensibilidad de la
técnica molecular

cuantil1=list(p=0.025,x=0.61)
cuantil2=list(p=0.975,x=0.92)
beta.select(cuantil1,cuantil2)

# Valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)

# Valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,19.69 ,5.34)

plot(grill, valores,
      main="Beta (19.69, 5.34)",type="l",xlab="p",ylab="Distribución
Beta Sensibilidad PCR",lwd=2)

#Elicitación de los parámetros alfa y beta de la especificidad de la
técnica molecular

cuantil1=list(p=0.025,x=0.89)
cuantil2=list(p=0.975,x=0.97)
beta.select(cuantil1,cuantil2)

# Valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)

# Valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,132.22, 9.08)

plot(grill, valores,
      main="Beta(132.22, 9.08)",type="l",xlab="p",ylab="Distribución
Beta Especificidad PCR",lwd=2)

```

```

#Elicitación de los parámetros alfa y beta de la sensibilidad de la
técnica parasitológica

#utilizando dos cuantiles

cuantil1=list(p=0.025,x=0.38)
cuantil2=list(p=0.975,x=0.78)
beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)

# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,12.83,9.02)

plot(grill, valores,
      main="Beta(12.83,9.02)",type="l",xlab="p",ylab="Distribución Beta
Sensibilidad Baermann",lwd=2)

#Elicitación de los parámetros alfa y beta de la especificidad de la
técnica parasitologica

cuantil1=list(p=0.025,x=0.91)
cuantil2=list(p=0.975,x=0.97)
beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)

# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,211.0,12.6)

plot(grill, valores,
      main="Beta(211.0,12.6)",type="l",xlab="p",ylab="Distribución Beta
Especificidad Baermann",lwd=2)

#Elicitación de los parámetros alfa y beta de la prevalencia enfermedad

cuantil1=list(p=0.025,x=0.06)
cuantil2=list(p=0.975,x=0.18)
beta.select(cuantil1,cuantil2)

# valores de p donde se va a calcular la distribución

```

```

grill <- seq(0,1,length = 100)
# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,11.82, 92.89)
plot(grill, valores,
      main="Beta(11.82, 92.89)",type="l",xlab="p",ylab="Distribución
Beta Prevalencia Estrongiloidiasis",lwd=2)

```

### **Programa 3. Elicitación de los parámetros mediante bibliografía para la sensibilidad y especificidad y mediante datos de Etiopía.**

#### **Estos datos han sido elicitados en el programa 1 y programa 2.**

```

#Elicitación de los parámetros alfa y beta de la sensibilidad de la
técnica molecular
#utilizando dos cuantiles
#lo primero es instalar el paquete LearnBayes
cuantil1=list(p=0.1,x=0.7)
cuantil2=list(p=0.90,x=0.9)
beta.select(cuantil1,cuantil2)
# valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)
# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,19.87,4.80)
plot(grill, valores,
      main="Beta (19.87,4.80)",type="l",xlab="p",ylab="Distribución Beta
Sensibilidad PCR",lwd=2)
#Elicitación de los parámetros alfa y beta de la especificidad de la
técnica molecular
cuantil1=list(p=0.05,x=0.7)
cuantil2=list(p=0.85,x=0.9)
beta.select(cuantil1,cuantil2)
# valores de p donde se va a calcular la distribución

```

```

grill <- seq(0,1,length = 100)
# Valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,23.01,4.78)
plot(grill, valores,
      main="Beta(23.01,4.78)",type="l",xlab="p",ylab="Distribución Beta
Especificidad PCR",lwd=2)

```

#Elicitación de los parámetros alfa y beta de la sensibilidad de la técnica parasitológica

```
cuantil1=list(p=0.15,x=0.3)
```

```
cuantil2=list(p=0.85,x=0.6)
```

```
beta.select(cuantil1,cuantil2)
```

# Valores de p donde se va a calcular la distribución

```
grill <- seq(0,1,length = 100)
```

```
# Valores de la función de densidad de la beta(a,b)
```

```
valores<- dbeta(grill,5.28,6.47)
```

```
plot(grill, valores,
```

```
      main="Beta(5.28,6.47)",type="l",xlab="p",ylab="Distribución Beta
Sensibilidad Baermann",lwd=2)
```

#Elicitación de los parámetros alfa y beta de la especificidad de la técnica parasitológica

```
cuantil1=list(p=0.05,x=0.9)
```

```
cuantil2=list(p=0.95,x=0.98)
```

```
beta.select(cuantil1,cuantil2)
```

# Valores de p donde se va a calcular la distribución

```
grill <- seq(0,1,length = 100)
```

```
# Valores de la función de densidad de la beta(a,b)
```

```
valores<- dbeta(grill,77.32,4.40)
```

```

plot(grill, valores,
      main="Beta(77.32,4.40)",type="l",xlab="p",ylab="Distribución Beta
Especificidad Baermann",lwd=2)
#Elicitación de los parámetros alfa y beta de la prevalencia enfermedad
cuantil1=list(p=0.025,x=0.06)
cuantil2=list(p=0.975,x=0.18)
beta.select(cuantil1,cuantil2)
# valores de p donde se va a calcular la distribución
grill <- seq(0,1,length = 100)
# valores de la función de densidad de la beta(a,b)
valores<- dbeta(grill,11.82, 92.89)
plot(grill, valores,
      main="Beta(11.82, 92.89)",type="l",xlab="p",ylab="Distribución
Beta Prevalencia Estrongiloidiasis",lwd=2)

```

## Tablas de contingencia

### Angola

```

table(ANGOLA$PCR_Strongy, ANGOLA$`Prueba Baermann`)
D<-ANGOLA$`CAMINA DESCALZO`==1
Z<-ANGOLA$`CAMINA DESCALZO`==0
y<-table(ANGOLA$PCR_Strongy, ANGOLA$`Prueba Baermann`, data=D)
z<-table(ANGOLA$PCR_Strongy, ANGOLA$`Prueba Baermann`, data=Z)

```

### Etiopía

```

table(Etiopia_Datos$P, Etiopia_Datos$B)

```

## Modelo 1 Angola

```

#Prevalencias no informativas, sensibilidad y especificidad
bibliografía

```

```

cat("
  model
  {
y[1:Q, 1:Q] ~ dmulti(p1[1:Q, 1:Q], n1)
z[1:Q, 1:Q] ~ dmulti(p2[1:Q, 1:Q], n2)
p1[1,1] <- pi1*Sepcr*Sebaermann + (1-pi1)*(1-Sppcr)*(1-Spbaermann)
p1[1,2] <- pi1*Sepcr*(1-Sebaermann) + (1-pi1)*(1-Sppcr)*Spbaermann
p1[2,1] <- pi1*(1-Sepcr)*Sebaermann + (1-pi1)*Sppcr*(1-Spbaermann)
p1[2,2] <- pi1*(1-Sepcr)*(1-Sebaermann) + (1-pi1)*Sppcr*Spbaermann
p2[1,1] <- pi2*Sepcr*Sebaermann + (1-pi2)*(1-Sppcr)*(1-Spbaermann)
p2[1,2] <- pi2*Sepcr*(1-Sebaermann) + (1-pi2)*(1-Sppcr)*Spbaermann
p2[2,1] <- pi2*(1-Sepcr)*Sebaermann + (1-pi2)*Sppcr*(1-Spbaermann)
p2[2,2] <- pi2*(1-Sepcr)*(1-Sebaermann) + (1-pi2)*Sppcr*Spbaermann
pi1 ~ dbeta(1, 1)
Sepcr ~ dbeta(19.87, 4.80)
Sppcr ~ dbeta(23.01, 4.78)
pi2 ~ dbeta(1, 1)
Sebaermann ~ dbeta(5.28, 6.47)
Spbaermann ~ dbeta(77.32, 4.40)
}
",
  file = "modelo1.txt")

#Datos
Q=2
n1=206
n2=118
y=structure(.Data=c(14,10,35,147), .Dim=c(2,2))
z=structure(.Data=c(5,7,17,89), .Dim=c(2,2))

```

```

#FNbaermann <- 1-Sebaermann

datos <- list ("Q","n1","n2","y","z")

iniciales <- function()
{list(pi1=runif(1),Sepcr=runif(1),Sppcr=runif(1),pi2=runif(1),Sebaerman
n=runif(1),Spbaermann=runif(1))}

parametros <- c("pi1","Sepcr","Sppcr","pi2","Sebaermann","Spbaermann")

library(R2OpenBUGS)

modelo1 <- bugs(data = datos, inits = iniciales,
               parameters.to.save = parametros,
               model.file = "modelo1.txt", n.chains = 3,
               n.iter=100000, n.burnin = 10000,
               working.directory = laruta, coda = TRUE,
               clearWD = TRUE, debug = TRUE)

library(coda)

cosacoda<-
read.bugs(c("CODAchain1.txt","CODAchain2.txt","CODAchain3.txt"))

geweke.diag(cosacoda, frac1=0.1, frac2=0.5)

gelman.diag(cosacoda, confidence=0.95, transform=FALSE,
autoburnin=FALSE)

```

## Modelo 2

```

####con las a priori obtenidas de Etiopia

laruta <- "D:/Usuarios/edacal/Desktop/Trabajo Fin Master Estadistica"

setwd(laruta)

cat("
  model
  {
    y[1:Q, 1:Q] ~ dmulti(p1[1:Q, 1:Q], n1)

```

```

z[1:Q, 1:Q] ~ dmulti(p2[1:Q, 1:Q], n2)
p1[1,1] <- pi1*Sepcr*Sebaermann + (1-pi1)*(1-Sppcr)*(1-Spbaermann)
p1[1,2] <- pi1*Sepcr*(1-Sebaermann) + (1-pi1)*(1-Sppcr)*Spbaermann
p1[2,1] <- pi1*(1-Sepcr)*Sebaermann + (1-pi1)*Sppcr*(1-Spbaermann)
p1[2,2] <- pi1*(1-Sepcr)*(1-Sebaermann) + (1-pi1)*Sppcr*Spbaermann
p2[1,1] <- pi2*Sepcr*Sebaermann + (1-pi2)*(1-Sppcr)*(1-Spbaermann)
p2[1,2] <- pi2*Sepcr*(1-Sebaermann) + (1-pi2)*(1-Sppcr)*Spbaermann
p2[2,1] <- pi2*(1-Sepcr)*Sebaermann + (1-pi2)*Sppcr*(1-Spbaermann)
p2[2,2] <- pi2*(1-Sepcr)*(1-Sebaermann) + (1-pi2)*Sppcr*Spbaermann

pi1 ~ dbeta(11.82, 92.89)
Sepcr ~ dbeta(19.69, 5.34)
Sppcr ~ dbeta(132.22, 9.08)
pi2 ~ dbeta(11.82, 92.89)
Sebaermann ~ dbeta(12.83, 9.02)
Spbaermann ~ dbeta(211, 12.06)

}

",

file = "modelo2.txt")

```

```

#Datos
Q=2
n1=206
n2=118
#y=structure(.Data=c(147,35,10,14),.Dim=c(2,2))
#z=structure(.Data=c(89,17,7,5),.Dim=c(2,2))
y=structure(.Data=c(14,10,35,147),.Dim=c(2,2))
z=structure(.Data=c(5,7,17,89),.Dim=c(2,2))

```

```

#FNbaermann <- 1-Sebaermann

datos <- list ("Q","n1","n2","y","z")

iniciales <- function()
{list(pi1=runif(1),Sepcr=runif(1),Sppcr=runif(1),pi2=runif(1),Sebaerman
n=runif(1),Spbaermann=runif(1))}

parametros <- c("pi1","Sepcr","Sppcr","pi2","Sebaermann","Spbaermann")

library(R2OpenBUGS)

modelo2 <- bugs(data = datos, inits = iniciales,
               parameters.to.save = parametros,
               model.file = "modelo2.txt", n.chains = 3,
               n.iter=100000, n.burnin = 10000,
               working.directory = laruta, coda = TRUE,
               clearWD = TRUE, debug = TRUE)

library(coda)

cosacoda<-
read.bugs(c("CODAchain1.txt","CODAchain2.txt","CODAchain3.txt"))

geweke.diag(cosacoda, frac1=0.1, frac2=0.5)

gelman.diag(cosacoda, confidence=0.95, transform=FALSE,
autoburnin=FALSE)

```

### Modelo 3

```

#modelo 3 prevalencia a priori obtenida de la a posteriori Etiopía +
sens y esp bibliografía

laruta <- "D:/Usuarios/edacal/Desktop/Trabajo Fin Master Estadística"

setwd(laruta)

cat("

    model

```

```

{
y[1:Q, 1:Q] ~ dmulti(p1[1:Q, 1:Q], n1)
z[1:Q, 1:Q] ~ dmulti(p2[1:Q, 1:Q], n2)

p1[1,1] <- pi1*Sepcr*Sebaermann + (1-pi1)*(1-Sppcr)*(1-Spbaermann)
p1[1,2] <- pi1*Sepcr*(1-Sebaermann) + (1-pi1)*(1-Sppcr)*Spbaermann
p1[2,1] <- pi1*(1-Sepcr)*Sebaermann + (1-pi1)*Sppcr*(1-Spbaermann)
p1[2,2] <- pi1*(1-Sepcr)*(1-Sebaermann) + (1-pi1)*Sppcr*Spbaermann
p2[1,1] <- pi2*Sepcr*Sebaermann + (1-pi2)*(1-Sppcr)*(1-Spbaermann)
p2[1,2] <- pi2*Sepcr*(1-Sebaermann) + (1-pi2)*(1-Sppcr)*Spbaermann
p2[2,1] <- pi2*(1-Sepcr)*Sebaermann + (1-pi2)*Sppcr*(1-Spbaermann)
p2[2,2] <- pi2*(1-Sepcr)*(1-Sebaermann) + (1-pi2)*Sppcr*Spbaermann

pi1 ~ dbeta(11.82, 92.89)
sepcr ~ dbeta(19.87, 4.80)
sppcr ~ dbeta(23.01, 4.78)
pi2 ~ dbeta(11.82, 92.89)
Sebaermann ~ dbeta(5.28, 6.47)
Spbaermann ~ dbeta(77.32, 4.40)
}

",
file = "modelo3.txt")

```

```

#Datos
Q=2
n1=206
n2=118
#y=structure(.Data=c(147,35,10,14),.Dim=c(2,2))
#z=structure(.Data=c(89,17,7,5),.Dim=c(2,2))
y=structure(.Data=c(14,10,35,147),.Dim=c(2,2))

```

```

z=structure(.Data=c(5,7,17,89),.Dim=c(2,2))

#FNbaermann <- 1-Sebaermann

datos <- list ("Q","n1","n2","y","z")

iniciales <- function()
{list(pi1=runif(1),Sepcr=runif(1),Sppcr=runif(1),pi2=runif(1),Sebaerman
n=runif(1),Spbaermann=runif(1))}

parametros <- c("pi1","Sepcr","Sppcr","pi2","Sebaermann","Spbaermann")

library(R2OpenBUGS)

modelo3 <- bugs(data = datos, inits = iniciales,
               parameters.to.save = parametros,
               model.file = "modelo1.txt", n.chains = 3,
               n.iter=100000, n.burnin = 10000,
               working.directory = "aruta", coda = TRUE,
               clearWD = TRUE, debug = TRUE)

library(coda)

cosacoda<-
read.bugs(c("CODAchain1.txt","CODAchain2.txt","CODAchain3.txt"))

geweke.diag(cosacoda, frac1=0.1, frac2=0.5)

gelman.diag(cosacoda, confidence=0.95, transform=FALSE,
autoburnin=FALSE)

```

## Bibliografía

Amor A, Rodriguez E, Saugar JM, Arroyo A, López-Quintana B, Abera B, Yimer M, Yizengaw E, Zewdie D, Ayehubizu Z, Hailu T, Mulu W, Echazú A, Krolewski AJ, Aparicio P, Herrador Z, Anegagrie M, Benito A. High prevalence of *Strongyloides stercoralis* in school-aged children in a rural highland of north-western Ethiopia: the role of intensive diagnostic work-up. *Parasit Vectors*. 2016; 1;9(1):617.

Assefa T, Woldemichael T, Seyoum T. Evaluation of the modified Baermann's method in the laboratory diagnosis of *Strongyloides stercoralis*. *Ethiop Med J*. 1991; 29(4):193-8.

Becker SL, Piraisoody N, Kramme S, Marti H, Silué KD, Panning M, Nickel B, Kern WV, Herrmann M, Hatz CF, N'Goran EK, Utzinger J, von Müller L. Real-time PCR for detection of *Strongyloides stercoralis* in human stool samples from Côte d'Ivoire: diagnostic accuracy, inter-laboratory comparison and patterns of hookworm co-infection. *Acta Trop*. 2015; 150:210-7.

Bird C, Ame S, Albonico M, Bickle Q. Do shoes reduce hookworm infection in school-aged children on Pemba Island, Zanzibar? A pragmatic trial. *Trans R Soc Trop Med Hyg*. 2014;108(5):297-304.

Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med*. 2005; 10;68(2-4):145-63.

Campbell SJ, Stothard JR, O'Halloran F, Sankey D, Durant T, Ombede DE, Chuinteu GD, Webster BL, Cunningham L, LaCourse EJ, Tchuem-Tchuente LA. Urogenital schistosomiasis and soil-transmitted helminthiasis (STH) in Cameroon: An epidemiological update at Barombi Mbo and Barombi Kotto crater lakes assessing prospects for intensified control interventions. *Infect Dis Poverty*. 2017; 27;6(1):49.

Costa-Cruz JM, Madalena J, Silva DA, Sopelete MC, Campos DM, Taketomi EA. Heterologous antigen extract in ELISA for the detection of human IgE anti-*Strongyloides stercoralis*. *Rev Inst Med Trop Sao Paulo*. 2003; 45: 265-68.

Dendukuri N, Bélisle P, Joseph L. Bayesian sample size for diagnostic test studies in the absence of a gold standard: Comparing identifiable with non-identifiable models. *Stat Med*. 2010; 20;29(26):2688-97.

- Depaoli, S, & van de Schoot, R. Improving transparency and replication in Bayesian Statistics: The WAMBS-Checklist. *Psychol Methods*. 2017; 22(2):240-261
- Dreyer G, Fernandes-Silva E, Alvez S, Rocha A, Albuquerque R, Addis D. Patterns of detection of *Strongyloides stercoralis* in stool specimens: implications for diagnosis and clinical trials. *J Clin Microbiol*. 1996; 34: 2569-71.
- Easton AV, Oliveira RG, O'Connell EM, Kepha S, Mwandawiro CS, Njenga SM, Kihara JH, Mwatele C, Odiero MR, Brooker SJ, Webster JP, Anderson RM, Nutman TB. Multi-parallel qPCR provides increased sensitivity and diagnostic breadth for gastrointestinal parasites of humans: field-based inferences on the impact of mass deworming. *Parasit Vectors*. 2016; 27;9:38.
- Enøe C, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev Vet Med*. 2000; 30;45(1-2):61-81.
- García LS. *Parasitic Infections in the Compromised Host. Diagnostic Medical Parasitology*. 540. ASM PRESS. Washington DC, EEUU. 2007.
- Globe DI. *A History of Human Helminthology*. CAB international, Wallingford, Reino Unido. 1990; 543-70.
- Hasegawa H, Hayashida S, Ikeda Y, Sato H. Hyper-variable regions in 18S rDNA of *Strongyloides* spp. as markers for species-specific diagnosis. *Parasitol Res*. 2009;104: 869-874.
- Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics*. 1980; 36(1):167-71.
- Humphries D, Simms BT, Davey D, Otchere J, Quagraine J, Terryah S, Newton S, Berg E, Harrison LM, Boakye D, Wilson M, Cappello M. Hookworm infection among school age children in Kintampo north municipality, Ghana: nutritional risk factors and response to albendazole treatment. *Am J Trop Med Hyg*. 2013;89(3):540-8.
- 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.

Lau C, Samalvides F, Terashima A. Evaluación de técnicas parasitológicas en el diagnóstico de strongiloidiasis por *Strongyloides stercoralis*. Rev Med Hered. 2005;16:11-18.

Levenhagen MA y Costa-Cruz JM. Update on immunologic and molecular diagnosis of human strongyloidiasis. Acta trop.2014; 135: 33-43.

Lindo JF, Conway DJ, Atkins NS, Bianco AE, Robinson RD, Bundy DAP. Prospective evaluation of enzyme-linked immunosorbent assay and immunoblot methods for the diagnosis of endemic *Strongyloides stercoralis* infection. Am J Trop Med Hyg. 1994; 51: 175-79.

Madinga J, Polman K, Kanobana K, van Lieshout L, Brienen E, Praet N, Kabwe C, Gabriël S, Dorny P, Lutumba P, Speybroeck N. Epidemiology of polyparasitism with *Taenia solium*, schistosomes and soil-transmitted helminths in the co-endemic village of Malanga, Democratic Republic of Congo. Acta Trop. 2017;171:186-193.

Meurs L, Polderman AM, Vinkeles Melchers NV, Brienen EA, Verweij JJ, Groosjohan B, et al. Diagnosing polyparasitism in a high-prevalence setting in Beira, Mozambique: Detection of intestinal parasites in fecal samples by microscopy and real-time PCR. PLoS Negl Trop Dis. 2017;11:e0005310.

Moghaddassani H, Mirhendi H, Hosseini M, Rokni M, Mowlavi G, Kia E. Molecular Diagnosis of *Strongyloides stercoralis* Infection by PCR Detection of Specific DNA in Human Stool Samples. Iran J Parasitol 2011; 6: 23-30.

Morawski BM, Yunus M, Kerukadho E, Turyasingura G, Barbra L, Ojok AM, DiNardo AR, Sowinski S, Boulware DR, Mejia R. Hookworm infection is associated with decreased CD4+ T cell counts in HIV-infected adult Ugandans. PLoS Negl Trop Dis. 2017; 25;11(5):e0005634.

Oliveira D, Ferreira FS, Atouguia J, Fortes F, Guerra A, Centeno-Lima S. Infection by Intestinal Parasites, Stunting and Anemia in School-Aged Children from Southern Angola. PLoS One. 2015; 15;10(9):e0137327.

Paula FM, Malta Fde M, Marques PD, Sitta RB, Pinho JR, Gryscek RC, Chieffi PP. Molecular diagnosis of strongyloidiasis in tropical areas: a comparison of conventional and

real-time polymerase chain reaction with parasitological methods. Mem Inst Oswaldo Cruz. 2015 Apr;110(2):272-4. doi: 10.1590/0074-02760140371. Epub 2015 Mar 6.

Requena-Méndez A, Chiodini P, Bisoffi Z, Buonfrate D, Gotuzzo E, Muñoz J. The Laboratory Diagnosis and Follow Up of Strongyloidiasis: A Systematic Review. PLoS Negl Trop Dis. 2013; 7: issue 1/e 2002.

Ruano A, Martín T, Pardo J, López-Abán J, Muro A. Avances en el estudio sobre la strongiloidosis. Enfermedades Emergentes. 2005; 7: 102-109.

Saugar JM, Merino FJ, Martín-Rabadan P, Fernández-Soto P, Ortega S, Garate T, Rodríguez E. Application of real-time PCR for the detection of *Strongyloides* spp. in clinical samples in a reference center in Spain. Acta trop 2015; 142: 20-25.

Schar F, Odermatt P, Khieu V, Panning M, Duong S, Muth S, Marti H & Kramme S, Evaluation of real-time PCR for *Strongyloides stercoralis* and hookworm as diagnostic tool in asymptomatic schoolchildren in Cambodia. Acta tropica 2013; 126: 89-92.

Sharifdini M, Mirhendi H, Ashrafi K, Hosseini M, Mohebbi M, Khodadadi H, Kia EB. Comparison of Nested Polymerase Chain Reaction and Real-Time Polymerase Chain Reaction with Parasitological Methods for Detection of *Strongyloides stercoralis* in Human Fecal Samples. Am J Trop Med Hyg. 2015;93(6):1285-91

Shiferaw MB, Mengistu AD. Helminthiasis: Hookworm Infection Remains a Public Health Problem in Dera District, South Gondar, Ethiopia. PLoS One. 2015; 10;10(12):e0144588.

Silva LP, Barcelos IS, Passos-Lima AB, Espindola FS, Campos DM, Costa-Cruz JM. Western Blotting Using *Strongyloides ratti* Antigen for the Detection of IgG Antibodies as Confirmatory Test in Human *Strongyloides*. Mem Inst Oswaldo Cruz. 2003; 98(5): 687-91.

Silva, LC & Benavides, A. El enfoque bayesiano: otra manera de inferir. Gac Sanit 2001; 15 (4): 341-346.

Sudré AP, Siquira RC, Barreto MGM, Peralta RHS, Macedo HS, Peralta JM. Identification of a 26-kDa protein fraction as an important antigen for application in the immunodiagnosis of strongyloidiasis. Parasitol Res. 2007; 101:1117-23.

Sultana Y, Jeffreys N, Watts MR, Gilbert GL, Lee R. Real-time polymerase chain reaction for detection of *Strongyloides stercoralis* in stool. *Am J Trop Med Hyg.* 2013; 88: 1048-1051.

Vadlamudi RS, Chi DS, Krishnaswamy G. Intestinal strongyloidiasis and hyperinfection syndrome. *Clin Mol Allergy.* 2006; 30: 4-8.

Van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, van Aken MA. A gentle introduction to bayesian analysis: applications to developmental research. *Child Dev.* 2014; 85(3):842-60.

Verweij, J.J., M. Canales, K. Polman, J. Ziem, E.A. Brienen, A.M. Polderman & L. van Lieshout, Molecular diagnosis of *Strongyloides stercoralis* in faecal samples using real-time PCR. *Trans R Soc Trop Med Hyg.* 2009; 103: 342-346.