# Problemas de Inferencia Estadística en los Tests de Cribaje de una Enfermedad en Varias Fases con Distintos Grados de Verificación con Gold Estandar


UNIVERSIDAD DE GRANADA

**Jorge Mario Estrada Álvarez**

Programa de Doctorado en Estadística Matemática y Aplicada

Universidad de Granada

Tesis presentada para optar al título de

*Doctor en Estadística Matemática y Aplicada*

Departamento de Estadística e Investigación Operativa

2025

Editor: Universidad de Granada. Tesis Doctorales
Autor: Jorge Mario Estrada Álvarez
ISBN: 978-84-1195-971-1
URI: https://hdl.handle.net/10481/110581

**Director y Tutor**

Dr. Juan de Dios Luna del Castillo

Profesor Emerito de la Universidad de Granada

**Fecha**

Mayo, 2025.

*Quiero dedicar esta tesis a mi Madre* ***Maria Beatriz Alvarez Zambrano*** *(QEPD)*
*. . . Se que desde el cielo, como en todos mis logros estaras orgullosa de mi.*

*A toda mi familia . . .*

# Agradecimientos

Quiero agradecer muy especialmente a las siguientes personas e instituciones que de alguna manera u otra contribuyeron a este logro sustancial en mi vida profesional.

Gracias !

Esta investigación fue desarrollada con financiamiento de **PCT Therapeutics** y **La Secretaria de Salud Pública y Seguridad Social de Pereira**, Colombia.

Me gustaría agradecer también al: **Dr. Juan de Dios Luna del Castillo**, más que director de tesis, un mentor.

Al tribunal de evaluación de tesis por sus comentarios y sugerencias que permitieron mejorar este documento.

Finalmente, agradecer a:

Dra. Gloria Liliana Porras Hurtado, Dr. Hernán Felipe García Arias, a mi equipo de trabajo de la Secretaría de Salud Pública (Maryluz H. y Ángela M.), a todas las compañeras del Centro de Investigación en Salud Comfamiliar, a mis compañeros del FETP Colombia (Juan Sebastián, Camilo y Yesith).

# Resumen

La estimación precisa de la prevalencia de enfermedades constituye un componente esencial en epidemiología y salud pública, especialmente en condiciones de baja frecuencia donde los errores metodológicos pueden conducir a sesgos significativos. Este trabajo doctoral se centra en comparar distintos enfoques de estimación de prevalencias en contextos con pruebas imperfectas y bajas prevalencias, evaluando métodos frecuentistas, correcciones clásicas como Rogan-Gladen, y enfoques Bayesianos, bajo diferentes esquemas de muestreo, incluyendo una corrección propuesta por post-estratificación.

Se desarrolló un estudio basado en tres componentes metodológicos. Primero, se evaluaron estimadores clásicos aplicados sobre la identificación genética de casos de quilomicronemia familiar a partir de un esquema de tamizaje en dos fases. Posteriormente, se implementaron análisis aplicados para estimar prevalencias ajustadas mediante post-estratificación en escenarios de verificación parcial, tanto con gold estándar como con pruebas de sensibilidad y especificidad conocidas. Finalmente, se aplicaron modelos de regresión logística corregidos por clasificación errónea en datos reales de prevalencia de infecciones de transmisión sexual (VIH y sífilis) en una muestra de 11452 individuos, comparando el desempeño de modelos frecuentistas, ajustes simultáneos de parámetros y modelos Bayesianos.

Los resultados mostraron que en estimaciones directas utilizando gold estándar, los métodos frecuentistas como Agresti-Coull y Jeffreys ofrecieron resultados consistentes con baja variabilidad. No obstante, al integrar el efecto de la clasificación errónea, métodos como el de Lang basados en Rogan-Gladen subestimaron notablemente la prevalencia en escenarios de baja prevalencia, mientras que los enfoques Bayesianos ofrecieron estimaciones más estables y acordes con la realidad epidemiológica. La corrección por post-estratificación, y los métodos Bayesianos mejoraron la cobertura y representatividad poblacional, particularmente en esquemas de verificación parcial, ampliando los intervalos de incertidumbre para reflejar adecuadamente la variabilidad introducida. En el estudio de infecciones de transmisión sexual, el modelo Bayesiano con corrección implícita equilibró precisión y ajuste, produciendo estimaciones intermedias entre los

modelos tradicionales y complejos como el de Liu, pero con menores errores estándar y mejores propiedades de cobertura.

Se concluye que el uso de métodos que incorporan explícitamente la clasificación errónea, como los modelos Bayesianos corregidos, es fundamental para obtener estimaciones de prevalencia válidas y precisas en contextos de baja prevalencia y pruebas diagnósticas imperfectas. Asimismo, la post-estratificación constituye una herramienta crítica para corregir el sesgo en muestras no probabilísticas, aumentando la validez externa de las inferencias. Los métodos frecuentistas, aunque útiles en condiciones ideales, pueden ser inadecuados en contextos reales complejos. Finalmente, los modelos Bayesianos ofrecen ventajas notables en términos de flexibilidad, incorporación de información previa y capacidad de modelar la incertidumbre, posicionándose como alternativas metodológicas robustas para estudios epidemiológicos y aplicaciones en salud pública.

# Índice

# Lista de Figuras

# Lista de Tablas

# Capítulo 1

# Introducción

La estimación precisa de proporciones y prevalencias es un componente esencial en epidemiología, salud pública y diversas áreas clínicas. Desde una perspectiva metodológica, el problema fundamental consiste en inferir el valor de un parámetro poblacional desconocido a partir de una muestra de datos, incorporando de manera adecuada la incertidumbre inherente al proceso de muestreo y medición. Tradicionalmente, la inferencia sobre proporciones se ha basado en métodos frecuentistas que asumen muestras aleatorias y pruebas diagnósticas perfectas (1). Sin embargo, en la práctica, prevalencias bajas, errores de clasificación diagnóstica y limitaciones logísticas en el diseño de muestreo desafían los supuestos clásicos, exigiendo metodologías robustas de corrección y ajuste.

La tesis aborda el problema fundamental de la estimación de prevalencias mediante diversos intervalos de confianza, con énfasis en situaciones de baja prevalencia. Se analizan críticamente métodos como los intervalos de Wald, Agresti-Coull, Jeffreys y Clopper-Pearson. Se destaca que en contextos de prevalencias bajas, los métodos tradicionales tienden a subestimar la cobertura nominal, mientras que enfoques corregidos como el de Agresti-Coull y el de Jeffrey muestran un desempeño superior.

En la práctica, uno de los principales desafíos que enfrenta la estimación de prevalencias es el uso de pruebas diagnósticas imperfectas, de uso más generalizado en epidemiología, debido principalmente a la facilidad de su implementación en estudios que requieren trabajo de campo y suponen una dificultad la utilización de un gold estándar en dichos escenarios, lo que representa desafíos prácticos importantes a la hora de la estimación. Algunas alternativas frecuentistas clásicas, destacándose la corrección de Rogan y Gladen. Más recientemente, los enfoques Bayesianos han ganado terreno como alternativas robustas para modelar la incertidumbre diagnóstica, permitiendo tratar la

sensibilidad y especificidad como variables aleatorias y construir distribuciones posteriores que reflejan de manera coherente todas las fuentes de variabilidad. Por otro lado, el problema de las estimaciones en escenarios prácticos en los que son utilizadas estrategias de muestreo en fases ha sido bien descrito en el contexto de evaluación de condiciones mórbidas de baja prevalencia, pero que a su vez requieren el uso de pruebas con sensibilidad y especificidad imperfectas, llevando también a la necesidad de métodos que implementen correcciones y ajustes en la medida estimada.

Diferentes métodos y escenarios descritos ya presentan soluciones a la estimación de prevalencia bajo correcciones por imperfección diagnóstica, pero surge la necesidad de abordar el problema de inferencia bajo muestras no probabilísticas, lo que introduce un desafío adicional en los procesos de estimación, lo que abre las puertas a la aproximación de correcciones novedosas en este campo a través de técnicas clásicas como la post-estratificación y, como novedad en este trabajo, se propone y se evalúa una corrección basada en post-estratificación. Esta integración metodológica ofrece una vía para aprovechar la información recolectada en escenarios de práctica real, donde las condiciones ideales de muestreo raramente se cumplen.

Otro punto importante en la inferencia de la prevalencia se da cuando el investigador requiere no solo una estimación corregida por el uso de test imperfectos, sino también el deseo de obtener medidas ajustadas por covariables que pueden jugar el papel de confusores o generar sesgos por confusión en la prevalencia estudiada. Entonces, técnicas como la regresión logística aportan una solución. Sin embargo, la multiplicidad de métodos en este campo hace necesaria la evaluación en el desempeño de las mismas entre sí, que permitan al investigador actual tomar las mejores decisiones en cuanto a los métodos estadísticos disponibles.

Este trabajo integra datos de dos estudios inéditos de autoría propia en el marco del cumplimiento de los objetivos de esta tesis doctoral: El primero es un estudio de corte transversal desarrollado durante los años 2020 y 2021, con el fin de estimar la prevalencia de quilomicronemia familiar, una enfermedad huérfana que consiste en un trastorno genético que se manifiesta por un inadecuado metabolismo de los lípidos, trabajo que fue desarrollado de base hospitalaria en la Ciudad de Pereira, Colombia. El segundo corresponde a los datos de un programa de tamizaje para enfermedades de transmisión sexual, que tenía como objetivo estimar la prevalencia de la infección por Virus de la Inmunodeficiencia Humana (VIH), infección por sífilis, Hepatitis B, Hepatitis C, en población general y población clave en riesgo de infección, de la Ciudad de Pereira, Colombia. Estos dos trabajos corresponden con el problema principal de esta tesis, ya

que contenían los desafíos metodológicos de los que trata esta tesis: uso de pruebas diagnósticas imperfectas, estimaciones de prevalencia baja, muestras no probabilísticas y estimación bajo el ajuste de covariables de interés.

La tesis concluye proponiendo líneas futuras de investigación orientadas a extender los métodos Bayesianos en presencia de datos jerárquicos, covariables y estructuras de dependencia, consolidando así una plataforma metodológica sólida y flexible para la estimación de prevalencias, que sea aplicable tanto en condiciones ideales como en los escenarios complejos y heterogéneos que caracterizan la práctica empírica actual.

## 1.1 Objetivos

### 1.1.1 Objetivo general

Analizar el problema de la estimación de la prevalencia de una enfermedad con test imperfectos y prueba gold estándar

### 1.1.2 Objetivos Específicos

- Comparar deferentes intervalos para la estimación de prevalencias bajo estudios de una o mas fases, con un test imperfecto y Gold estándar.
- Evaluar los intervalos para la estimación de prevalencias bajo diferentes escenarios de aplicación en muestras no probabilísticas
- Comparar diferentes métodos de estimación de prevalencia mediante enfoque de regresión logística con corrección por mala clasificación.

## 1.2 Estructura de la Tesis

Esta tesis doctoral se organiza en tres capítulos principales:

- **Capitulo 2:** Se desarrolla el problema clásico de la estimación de prevalencias bajo el supuesto uso de gold estandar, se revisan métodos de corrección frecuentistas como el estimador de Rogan-Gladen y se introduce un modelo Bayesiano que incorpora la incertidumbre en sensibilidad y especificidad, proporcionando inferencias más robustas y coherentes.

- **Capítulo 3:** Se aborda la estimación de prevalencia en presencia de test diagnósticos imperfectos, esquema de selección de sujetos en fases con la combinación de pruebas imperfectas y gold estándar, ademas de una propuesta de corrección por post-estratificación para muestras no probabilísticas en estimaciones puntuales y por intervalo previamente descritos

- **Capítulo 4**: Se aborda mediante una revisión sistemática de literatura la identificación de métodos de regresión logistica para la estimación de prevalencia y ajuste por covariables incluyendo la corrección por mala clasificación introducida por el test diagnostico. Se revisa el enfoque frecuentista y Bayesiano y se comparan en términos de medidas de desempeño.

Cada capítulo incluye una sección de resultados empíricos que ilustra la aplicación práctica de los métodos propuestos y una discusión crítica sobre sus ventajas y limitaciones. Finalmente, se presenta una sección de conclusiones e investigaciones futuras.

## 1.3 Publicaciones Asociadas

Los siguientes artículos (Apéndice D.1) avalan esta tesis doctoral:

- Rodriguez* FH, Estrada* JM, Quintero HMA, Nogueira JP, Porras-Hurtado GL. Analyses of familial chylomicronemia syndrome in Pereira, Colombia 2010–2020: a cross-sectional study. Lipids in Health and Disease. 2023;22:43. Available from: https://doi.org/10.1186/s12944-022-01768-x Apendice
  * *Se comparte primera autoria*
  **Año SJR**: 2023
  **Factor de impacto**: 3.9
  **Quartil**: Q1

- Estrada Alvarez JM, Luna del Castillo JdD, Montero-Alonso M Point and Interval Estimation of Population Prevalence Using a Fallible Test and a Non-Probabilistic Sample: Post-Stratification Correction. Mathematics. 2025;13(5). Available from: https://www.mdpi.com/2227-7390/13/5/805
  **Año JCR**: 2023
  **Factor de impacto**: 2.3
  **Quartil**: Q1

- Alvarez JME, Garcia HF, Ángel Montero-Alonso M, de Dios Luna del Castillo J. Prevalence estimation in infectious diseases with imperfect tests: A comparison of Frequentist and Bayesian Logistic Regression methods with misclassification correction; 2025. Available from: https://arxiv.org/abs/2504.15150

# Capítulo 2

# Estimación de la prevalencia de una enfermedad con test imperfectos

## 2.1 Introducción

La estimación precisa de prevalencias es un objetivo central en epidemiología y salud pública, dado que constituye la base para la toma de decisiones clínicas, el diseño de intervenciones preventivas y la evaluación de programas de control de enfermedades. En este contexto, los métodos de estimación puntual y por intervalo de confianza han sido ampliamente desarrollados, proporcionando herramientas estadísticas fundamentales para cuantificar la incertidumbre inherente a las estimaciones muestrales.

Sin embargo, cuando se trata de estimar prevalencias bajas, situaciones frecuentes en enfermedades raras o en enfermedades infecciosas emergentes, surgen desafíos metodológicos particulares. La baja frecuencia de eventos implica que los métodos tradicionales de construcción de intervalos, como el intervalo de Wald clásico, presenten un desempeño inadecuado, con cobertura inferior al nivel nominal y alta inestabilidad en las estimaciones. Esta problemática ha motivado el desarrollo y recomendación de alternativas como los intervalos de Agresti-Coull y aquellos basados en la prior de Jeffreys, los cuales han demostrado un mejor comportamiento en términos de cobertura y amplitud bajo condiciones de baja prevalencia según lo revisado por Brown et al. (5).

Adicionalmente, en la práctica real, la evaluación de la presencia o no de la enfermedad en el individuo no siempre se realiza mediante procedimientos diagnósticos perfectos. La mayoría de las pruebas diagnósticas disponibles presentan sensibilidades y especificidades inferiores a 1.0, introduciendo errores de clasificación que sesgan las estimaciones de prevalencia. Para abordar esta situación, la literatura ha propuesto

diferentes estrategias de corrección tanto bajo enfoques frecuentistas como Bayesianos.

Entre las correcciones frecuentistas, destaca el estimador de Rogan y Gladen (6), que constituye un referente clásico en este campo. Este estimador ajusta la prevalencia observada tomando en cuenta la sensibilidad y especificidad del test, proporcionando una corrección directa y sencilla. No obstante, su aplicación requiere cautela, dado que puede generar estimaciones fuera del rango $[0, 1]$ en ciertos escenarios, especialmente cuando la sensibilidad y especificidad no son suficientemente altas.

Desde el enfoque Bayesiano, se propone modelar la sensibilidad, especificidad y prevalencia como parámetros aleatorios, lo que permite incorporar explícitamente la incertidumbre inherente al desempeño diagnóstico y realizar inferencias más robustas, particularmente en situaciones de baja prevalencia o tamaños muestrales reducidos.

Otro aspecto crítico a considerar en la estimación de prevalencias es el diseño de muestreo utilizado. Si bien la teoría clásica asume la existencia de una muestra aleatoria de la población objetivo, en la práctica esto rara vez se cumple, especialmente en estudios de tamizaje masivo o en contextos de acceso desigual a servicios de salud. En tales situaciones, surge la necesidad de implementar estrategias de optimización de recursos, tales como el muestreo en fases, donde se utiliza un test inicial de tamizaje seguido de la aplicación de una prueba gold estándar solo en una submuestra seleccionada. Esta estrategia, además de ser más costo-efectiva y éticamente viable, plantea retos metodológicos adicionales que requieren métodos estadísticos específicos para ajustar adecuadamente las estimaciones.

A partir de estas consideraciones, el presente capítulo revisa los principales métodos de estimación de prevalencias bajo la presencia de test imperfectos y estrategias de muestreo complejas, abordando tanto alternativas frecuentistas como Bayesianas, y discutiendo sus fortalezas y limitaciones en distintos contextos epidemiológicos.

## 2.2 Métodos

### 2.2.1 Estimación puntual y por intervalo de la prevalencia de una enfermedad bajo un gold estándar

El problema clásico en relación con la estimación del valor del parámetro $p$ radica en la existencia de una población en la cual se requiere conocer dicho valor. Sin embargo, debido al tamaño de la población, no es posible evaluar a todos los individuos, por lo que se recurre a técnicas de muestreo para obtener una estimación.

Para este propósito, se extrae una muestra probabilística de tamaño $n$, en la cual todos los individuos son evaluados mediante una prueba diagnóstica considerada como *gold estándar*. Esta prueba tiene únicamente dos posibles resultados: positivo o negativo. Un resultado positivo indica la presencia de la enfermedad o condición en estudio, mientras que un resultado negativo se interpreta como ausencia de la misma.

Desde el punto de vista estadístico, la estimación puntual de la prevalencia se reduce a la estimación de una proporción $p$, utilizando el método de máxima verosimilitud. Para ello se asume que:

- $x$: representa una variable aleatoria binomial, que indica el número de *éxitos*, es decir, el número de individuos clasificados como positivos por la prueba.
- $n$: número total de individuos evaluados en la muestra.
- $p$: parámetro de interés que representa la prevalencia poblacional de la enfermedad, es decir, la probabilidad de que un individuo seleccionado al azar tenga la condición.

Bajo estos supuestos, la variable aleatoria $x$ sigue una distribución binomial, y su función de probabilidad está dada por:

$$Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{para } x = 0, 1, \ldots, n. \tag{2.1}$$

El estimador de máxima verosimilitud (MLE) para $p$ se obtiene al maximizar esta función de verosimilitud con respecto a $p$, lo que conduce al estimador puntual clásico:

$$\hat{p} = \frac{x}{n} \tag{2.2}$$

Este estimador es insesgado y consistente bajo el supuesto de que la prueba utilizada clasifica perfectamente a los individuos, es decir, sin errores de sensibilidad o especificidad. Diferentes métodos se han desarrollado para estimación por intervalo de confianza (IC) para la prevalencia poblacional $p$, para lo cual se revisan a continuación los más conocidos y reportados en la literatura científica.

### Intervalo de Wald

El intervalo de confianza de Wald para una proporción $p$ se define como:

$x =$ número de éxitos en $n$ realizaciones independientes, $\hat{p} = \frac{X}{n}$, $\kappa = z_{\alpha/2} = \Phi^{-1}(1-\frac{\alpha}{2})$ y $\hat{q} = 1 - \hat{p}$, $\Phi(z)$ es la función de distribución de una normal estándar. En ocasiones, el intervalo de Wald es también llamado **intervalo estándar**

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{2.3}$$

Este intervalo se obtiene al invertir la prueba de hipótesis $H_0 : p = p_0$ usando el estadístico de prueba $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ y la región de no rechazo $|Z| \leq z_{\alpha/2}$.

Aunque el intervalo basado en la aproximación normal, también conocido como intervalo de Wald, ha sido ampliamente utilizado debido a su simplicidad, múltiples estudios han demostrado sus importantes limitaciones en términos de cobertura real (Figura 2.1).

Una crítica fundamental al intervalo de Wald proviene de Brown et al. (5), quienes realizaron una evaluación detallada de la cobertura empírica del intervalo bajo distintas combinaciones de $p$ y $n$. Sus simulaciones revelaron que la cobertura del intervalo Wald oscila de manera no trivial según los valores de la proporción y el tamaño de muestra. En particular, se observan combinaciones $(p, n)$ para las cuales la cobertura se aproxima al nivel nominal deseado (por ejemplo, 95%), pero también otras combinaciones en las que dicha cobertura cae de manera significativa, incluso para tamaños muéstrales relativamente grandes.

Por ejemplo, para una proporción fija de $p = 0.2$ y variando el tamaño de muestra entre $n = 25$ y $n = 100$, encontraron lo siguiente:

- Para el par $(p = 0.2, n = 30)$, la cobertura empírica observada fue aproximadamente 0.946, cercana al valor nominal del 95%.
- Sin embargo, para el par $(p = 0.2, n = 98)$, la cobertura disminuyó a 0.928, lo cual resulta contra intuitivo, dado que se esperaría una mejora de la cobertura con el incremento del tamaño muestral.

Estos hallazgos sugieren que el intervalo de Wald puede comportarse de forma errática en ciertos rangos de valores de $p$ y $n$, generando una falsa sensación de precisión. Como consecuencia, en la literatura actual se recomienda recurrir a métodos alternativos como

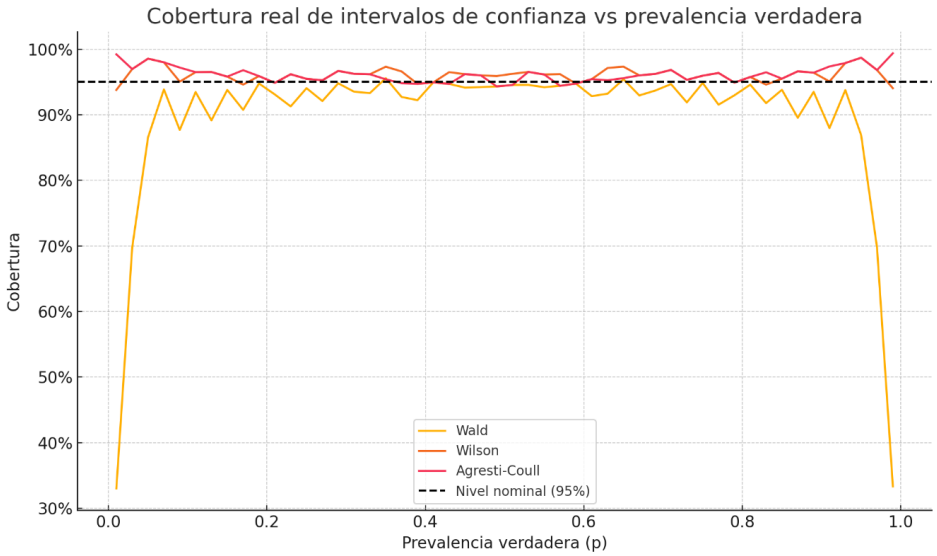


Figura 2.1: Cobertura real de intervalos de confianza para distintos valores de la prevalencia

*La Figura 2.1 muestra la cobertura empírica de los intervalos de confianza construidos por los métodos de Wald, Wilson y Agresti-Coull, simulada para distintos valores de la prevalencia verdadera $p$ con tamaño muestral $n = 40$. Se observa que el método de Wald tiende a infraestimar la cobertura cuando $p$ es extremo (muy bajo o muy alto), mientras que otros métodos mantienen niveles de cobertura más estables y cercanos al 95%.*

el intervalo de Wilson o enfoques Bayesianos, los cuales muestran una cobertura más estable y cercana al valor nominal bajo una amplia gama de condiciones.

Este intervalo tiene una cobertura de probabilidad muy por debajo del nivel de confianza nominal especificado, especialmente cuando $n$ es pequeño y cuando $p$ es cercano a 0 ó 1. Sin embargo, como lo demostró Agresti et al. (7) y se detallará más adelante, incluso en situaciones de tamaños de muestra grandes su cobertura de probabilidad no es adecuada.

Vollset (8) propone una modificación al clásico intervalo tipo Wald mediante una corrección por continuidad, con el objetivo de mejorar su aproximación a los límites exactos. Asumiendo que la variable aleatoria $X$ sigue una distribución binomial $(n, p)$, la proporción estimada se define como $\hat{p} = x/n$, con $\hat{q} = 1 - \hat{p}$, ajusta la estimación de la proporción al incorporar un término de corrección al numerador. El intervalo corregido se expresa como:

$$\frac{x + 0.5}{n} \pm z_{\alpha/2}\sqrt{\frac{(x + 0.5)(1 - \frac{x+0.5}{n})}{n}}. \tag{2.4}$$

donde $z_{\alpha/2}$ es el cuántil $1 - \alpha/2$ de la distribución normal estándar.

### Intervalo de Wilson

Este intervalo está basado en el score test como lo expuso Agresti (9), tomando el valor del estadístico $Z$, pero evaluando el valor nulo de la hipótesis $Ho : p = p_0$ (especialmente, el error estándar es calculado con el valor nulo $p_0$)

Suponga que$X_1, X_2, ..., X_n$ es una muestra aleatoria de una población Bin$(n, p)$, se considera probar $H_0 : p = p_0$ vs $H_1 : p \neq p_0$, donde $p_0$ es un valor fijo, $0 < p_0 < 1$. Cuando:

$$n \to \infty \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \to N(0, 1)$$

Entonces una prueba aproximada está basada en este estadístico, por lo cual la región de rechazo es

$$x : \left| \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \right| > z_{\alpha/2}$$

.

De donde, $H_0$ no se rechaza para puntos muestrales que satisfacen

$$\left| \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \right| \leq z_{\alpha/2}$$

Al invertir la región de no rechazo se obtendrá un intervalo de confianza para $p_0$ (Ecuación 2.5, se obtiene que:

$$|X - p_0 n| \leq z_{\alpha/2}\sqrt{np_0\left(1-p_0\right)},$$

$$X^2 - 2p_0 nX \leq -n\left(p_0^2\left(n + z_{\alpha/2}^2\right) - p_0 z_{\alpha/2}^2\right)$$

$$X^2 - 2p_0 nX \leq -n\frac{p_0^2\left(n + z_{\alpha/2}^2\right)^2 - 2p_0^2\left(n + z_{\alpha/2}^2\right)\frac{z_{\alpha/2}^2}{2p_0} + \frac{z_{\alpha/2}^4}{4} - \frac{z_{\alpha/2}^4}{4}}{n + z_{\alpha/2}^2}$$

$$X^2 - 2p_0 nX \leq -n\frac{p_0^2\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)^2 - \frac{z_{\alpha/2}^4}{4}}{n + z_{\alpha/2}^2}$$

$$X^2 - \frac{2p_0\left(n + z_{\alpha/2}^2\right)}{z_{\alpha/2}^2 n}\left(\frac{z_{\alpha/2}^2 n^2}{n + z_{\alpha/2}^2}\right)X \leq \frac{z_{\alpha/2}^2 n^2}{n + z_{\alpha/2}^2}\left(-\frac{p_0^2\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)^2 - \frac{z_{\alpha/2}^4}{4}}{z_{\alpha/2}^2 n}\right)$$

$$X^2\left(\frac{n + z_{\alpha/2}^2}{z_{\alpha/2}^2 n^2}\right) - \frac{2X_0\left(n + z_{\alpha/2}^2\right)}{z_{\alpha/2}^2 n} \leq -\frac{p_0^2\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)^2 - \frac{z_{\alpha/2}^4}{4}}{z_{\alpha/2}^2 n}$$

$$\frac{X^2}{z_{\alpha/2}^2 n} + \frac{X^2}{n^2} - \frac{2X^{z_{\alpha/2}^2 n}}{p_0\left(n + z_{\alpha/2}^2\right)} \leq -\frac{p_0^2\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)^2 - \frac{z_{\alpha/2}^4}{4}}{z_{\alpha/2}^2 n}$$

$$\frac{X^2 + Xz_{\alpha/2}^2 - 2XP_{p0}\left(n + z_{\alpha/2}^2\right)}{z_{\alpha/2}^2 n} \leq \frac{X}{n} - \frac{X^2}{n^2} + \frac{z_{\alpha/2}^2}{4n} - \frac{p_0^2\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)^2}{z_{\alpha/2}^2 n}$$

$$\frac{X^2 - 2X_{p0}\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)}{z_{\alpha/2}^2 n} \leq \frac{X}{n} - \frac{X^2}{n^2} + \frac{z_{\alpha/2}^2}{4n} - \frac{p_0^2\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right)^2}{z_{\alpha/2}^2 n}$$

$$\frac{\left(p_0\left(n + z_{\alpha/2}^2 - \frac{z_{\alpha/2}^2}{2p_0}\right) - X\right)^2}{z_{\alpha/2}^2 n} \leq \hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}$$

$$\frac{\left|p_0 n + p_0 z_{\alpha/2}^2 - X - \frac{z_{\alpha/2}^2}{2}\right|}{z_{\alpha/2} n^{1/2}} \leq \left(\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}\right)^{1/2}$$

$$\left|p_0 - \frac{X + \frac{z_{\alpha/2}^2}{2}}{n + z_{\alpha/2}^2}\right| \leq \frac{z_{\alpha/2} n^{1/2}}{n + z_{\alpha/2}^2}\left(\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}\right)^{1/2}$$

$$\frac{\left|p_0 n + p_0 z_{\alpha/2}^2 - X - \frac{z_{\alpha/2}^2}{2}\right|}{z_{\alpha/2} n^{1/2}} \leq \left(\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}\right)^{1/2}$$

$$\left|p_0 - \frac{X + \frac{z_{\alpha/2}^2}{2}}{n + z_{\alpha/2}^2}\right| \leq \frac{z_{\alpha/2} n^{1/2}}{n + z_{\alpha/2}^2}\left(\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}\right)^{1/2}$$

$$\frac{X + \frac{z_{\alpha/2}^2}{2}}{n + z_{\alpha/2}^2} - \frac{z_{\alpha/2} n^{1/2}}{n + z_{\alpha/2}^2}\left(\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}\right)^{1/2} \leq p_0 \leq \frac{X + \frac{z_{\alpha/2}^2}{2}}{n + z_{\alpha/2}^2} + \frac{z_{\alpha/2} n^{1/2}}{n + z_{\alpha/2}^2}\left(\hat{p}\hat{q} + \frac{z_{\alpha/2}^2}{4n}\right)^{1/2} \quad (2.5)$$

En resumen, el intervalo propuesto por Wilson corresponde a que a una proporción muestral se le adicione 2 observaciones de cada tipo, es decir 2 éxitos y 2 fallas. Mientras que el término de la derecha que acompaña el $z_{\alpha/2}$ en la ecuación 2.5, corresponde al promedio ponderado de la varianza cuando $p = \hat{p}$ y la varianza cuando $P = 1/2$, aplicando un ajuste sobre el tamaño de muestra, usando $n + z_{\alpha/2}^2$ en lugar de $n$.

El intervalo score o Wilson, acorde a lo estudiado por Brown et al. (5) presenta mejor desempeño, incluso que el intervalo por métodos exactos, siendo altamente cercana su cobertura de probabilidad a la confianza nominal, para casi cualquier tamaño de muestra (Figura 2.1 $n = 40$, Cobertura de 95.5%), probablemente se le atribuyen tales propiedades ya que se basa en que su punto medio del intervalo, como lo previamente mencionado, corrige la asimetría que pueda haber por tamaños de muestra pequeños, ya que este desplazamiento es menor a medida que $n$ incrementa. Este IC muestra superioridad, teniendo las menores amplitudes, en comparación con el IC Wald y otros (Figura 2.1).

Sin embargo, el intervalo Score presenta algunos inconvenientes en zonas límites de valores de $p$, cuando estos son cercanos a 0 o 1. En estos límites, sus coberturas de probabilidad caen dramáticamente. Los límites de estas regiones, acorde a Brown et al. (5) son: para $n = 10$ e intervalos con confianza nominal de 95%, la cobertura es de 0.83 para valores de $p$ hasta de 0.018 y 0.982; mientras que para $n = 100$, un mínimo de cobertura es de 0.838 con $p = 0.002$ y $p = 0.998$, lo que muestra que el nivel de confianza actual no converge al nivel nominal como $n$ incrementa, y esto podría ser problemático. Sin embargo, la región de valores en el espacio del parámetro, en la cual la cobertura de probabilidad del intervalo al 95% nominal cae por debajo de 90, no es más de 0.01 cuando $n >= 20$. Sumado a lo anterior, el intervalo Score o Wilson

muestra coberturas de probabilidad más cercanas a 0.95 en más del 90% del espacio del parámetro comparado con otros intervalos, y esto para varios tamaños de muestra reportados en el estudio ($n = 5$ hasta $n = 100$).

Otras correcciones propuestas para IC de Wilson, para subsanar algunos problemas en sus extremos, contienen una corrección para continuidad propuesta igualmente por Vollset (8):

$$\frac{(x \pm 0.5) + \frac{z^2_{\alpha/2}}{2} \pm z_{\alpha/2}\sqrt{\left[(x \pm 0.5) - \frac{(x \pm 0.5)^2}{n} + \frac{z^2_{\alpha/2}}{4}\right]}}{n + z^2_{\alpha/2}} \tag{2.6}$$

### Intervalo de Agresti-Coull

Agresti y Coull (7; 9) publicaron una comparación entre diferentes intervalos de confianza para una proporción, incluyendo un intervalo propuesto por ellos. Los IC evaluados fueron: el IC de Wald, el IC de Wilson , intervalo de Clopper-Pearson (IC exacto, el cual será discutido más adelante). Básicamente, el intervalo propuesto por Agresti et al. consiste en una versión del intervalo de Wald con un ajuste que corresponde a la adición de dos éxitos y dos fallas en las ecuaciones del intervalo. El sustento en dicha corrección o ajuste proviene de sus evaluaciones de desempeño en los intervalos mencionados.

El presente intervalo usa un nuevo estimador $\tilde{p}$ en lugar de $\hat{p}$. Lo anterior se logra mediante el uso del centro de la región de Wilson. Sea $\tilde{X} = X + \frac{z^2_{\alpha/2}}{2}$, $\tilde{n} = n + z^2_{\alpha/2}$, $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$, donde $\Phi(z)$ tiene una distribución normal estándar. Tomando en cuenta las expresiones anteriores, si $\tilde{p} = \frac{\tilde{X}}{\tilde{n}}$ y $\tilde{q} = 1 - \tilde{p}$, el intervalo de Agresti-Coull es:

$$\tilde{p} - z_{\alpha/2}\tilde{n}^{-1/2}\sqrt{\tilde{p}\tilde{q}} \geq p \leq \tilde{p} + z_{\alpha/2}\tilde{n}^{-1/2}\sqrt{\tilde{p}\tilde{q}} \tag{2.7}$$

Si en la expresión 2.7 en lugar de utilizar $z_{\alpha/2}$ se utiliza 2, el intervalo resultante es el intervalo en Agresti-Coull que añade 2 éxitos y 2 fracasos al intervalo de Wald. El IC propuesto por Agresti (Ecuación 2.7) expuso una mejoría sustancial en sus propiedades de desempeño. En cuanto a su cobertura de probabilidad, muestra tener un mínimo de buena cobertura, es bastante conservador cuando $p$ es cerca a 0 ó 1, en comparación con otros intervalos (Figura 2.1); en cuanto a longitud esperada del intervalo, esta precisión decae notablemente cuando $n$ es pequeño, lo que lo hace con un desempeño menor al resto de intervalos evaluados, incluso a tamaño de muestra de 20 la longitud es mayor

en 0.04 a 0.02 en comparación con los otros intervalos y estas diferencias pueden generar discrepancias prácticas importantes. Lo anterior es esperable cuando en la cobertura de probabilidad es bastante conservador a niveles de $p$ cercanas a 0 ó 1.

### Intervalo exacto de Clopper-Pearson

Los intervalos de confianza descritos por Clopper y Pearson (5; 10; 11), se basan en la inversión de pruebas binomiales exactas de dos colas bajo la hipótesis nula $H_0 : p = p_0$. Este intervalo se denomina exacto debido a que se deriva directamente de la distribución binomial, sin recurrir a aproximaciones normales o asintóticas.

Si se observa un número de éxitos $X = x$, el intervalo exacto de Clopper-Pearson se define como:

$$IC_{CP} = [I_{CP}(x),\, S_{CP}(x)]\,, \tag{2.8}$$

donde los extremos inferior y superior del intervalo se obtienen resolviendo las siguientes ecuaciones para $p$:

- Para el límite inferior:

$$\text{Inf} = \begin{cases} 0 & \text{si } x = 0 \\ \text{única solución de } \displaystyle\sum_{k=x}^{n} \binom{n}{k} \text{inferior}^k (1 - \text{inferior})^{n-k} = \frac{\alpha}{2} & \text{si } x > 0 \end{cases}$$

- Para el límite superior:

$$\text{Sup} = \begin{cases} 1 & \text{si } x = n \\ \text{única solución de } \displaystyle\sum_{k=0}^{x} \binom{n}{k} \text{superior}^k (1 - \text{superior})^{n-k} = \frac{\alpha}{2} & \text{si } x < n \end{cases}$$

Con base en estas definiciones, se derivan soluciones explícitas para los casos extremos:

- Si $x = 0$, entonces lower $= 0$ y upper $= 1 - (\alpha/2)^{1/n}$.
- Si $x = n$, entonces lower $= (\alpha/2)^{1/n}$ y upper $= 1$.

El intervalo propuesto por Clopper y Pearson garantiza que la cobertura de probabilidad sea siempre mayor o igual al nivel de confianza nominal. No obstante, para cualquier valor fijo de $p$, la cobertura real puede ser sustancialmente mayor a $1-\alpha$ si el tamaño de muestra $n$ no es lo suficientemente grande. Esta característica hace que el intervalo sea altamente conservador y, por tanto, una elección subóptima en la práctica cotidiana, pues tiende a generar intervalos más amplios e imprecisos.

En general, la solución de las ecuaciones anteriores para los límites inferior y superior no tiene forma cerrada (salvo en los extremos), por lo que los valores se obtienen comúnmente por métodos iterativos computacionales o mediante tablas especializadas.

Una alternativa eficiente se basa en la relación estrecha entre la distribución binomial y la distribución $F$ de Snedecor, ya que existe una correspondencia entre la función de distribución acumulada binomial y los percentiles de la distribución $F$. Esta relación permite expresar los límites del intervalo de Clopper-Pearson en términos de percentiles de la $F$, lo cual facilita su cálculo.

Así, los límites se pueden aproximar de la siguiente manera:

$$I_{CP} = \frac{x}{x + (n - x + 1)F_{2(n-x+1),\,2x;\,\alpha}}, \tag{2.9}$$

$$S_{CP} = \frac{(x+1)F_{2(x+1),\,2(n-x);\,\alpha}}{(n-x) + (x+1)F_{2(x+1),\,2(n-x);\,\alpha}} \tag{2.10}$$

donde $F_{df_1,df_2;\alpha}$ denota el percentil superior de orden $\alpha$ de la distribución $F$ con $df_1$ y $df_2$ grados de libertad.

En resumen, el intervalo exacto de Clopper-Pearson (Ecuacion 2.10) ofrece coberturas superiores al nivel nominal, incluso para niveles de confianza del 95%, donde se ha observado que la cobertura real puede oscilar entre el 96.5% y el 99% dependiendo del tamaño de la muestra. Si bien esto lo hace robusto frente a errores tipo I, también implica una pérdida considerable de precisión, haciéndolo menos atractivo para aplicaciones rutinarias en epidemiología o estudios de prevalencia.

### Ajuste de cobertura sobre el intervalo exacto de Clopper-Pearson

Thulin (12) propuso una modificación al intervalo exacto de Clopper-Pearson para corregir su excesiva conservadurismo, especialmente en valores extremos de la proporción $p$ cercanos a 0 o 1. Aunque el intervalo exacto garantiza que la cobertura mínima sobre todo el rango de $p$ sea al menos $1-\alpha$, en la práctica esta cobertura es frecuentemente muy

superior al nivel nominal deseado, afectando negativamente la precisión del intervalo.

Para abordar esta limitación, Thulin desarrolló una propuesta que ajusta el nivel de significancia $\alpha$ de modo que se satisfaga un criterio de cobertura media, combinando elementos de inferencia frecuentista y bayesiana. En este enfoque, el intervalo se ajusta para tener una cobertura promedio igual a $1-\alpha$ respecto a una distribución previa o posterior sobre $p$. Esta distribución funciona como función de ponderación en el espacio paramétrico.

El principio fundamental es que si se está dispuesto a aceptar una cobertura menor a $1-\alpha$ para ciertos valores de $p$, es posible mejorar el rendimiento general del intervalo eligiendo un nuevo valor de $\alpha$, denotado como $\alpha' > \alpha$, de forma que la cobertura real esté más próxima al nivel nominal en la mayoría del dominio de $p$.

El intervalo ajustado se define mediante las siguientes ecuaciones:

- Para el límite inferior:

$$\text{Inf} = \begin{cases} 0 & \text{si } x = 0 \\ \text{única solución de } \displaystyle\sum_{k=x}^{n} \binom{n}{k} \text{Inf}^k (1-\text{Inf})^{n-k} = \frac{\alpha'}{2} & \text{si } x > 0 \end{cases}$$

- Para el límite superior:

$$\text{Sup} = \begin{cases} 1 & \text{si } x = n \\ \text{única solución de } \displaystyle\sum_{k=0}^{x} \binom{n}{k} \text{Sup}^k (1-\text{Sup})^{n-k} = \frac{\alpha'}{2} & \text{si } x < n \end{cases}$$

El valor ajustado de $\alpha'$ se determina como la solución de la siguiente ecuación integral, que garantiza la cobertura media deseada:

$$C(\alpha', n) = \int_0^1 Pr(p \in IC_{CP}) \cdot f(p)\, dp \tag{2.11}$$

$$= \int_0^1 \sum_{x=0}^{n} 1(p \in IC_{CP}(x, \alpha')) \binom{n}{x} p^x (1-p)^{n-x} \cdot f(p)\, dp = 1-\alpha \tag{2.12}$$

donde $f(p)$ es una función de ponderación definida sobre el intervalo $(0,1)$, que puede ser interpretada como una densidad a priori sobre $p$, estableciendo así una conexión conceptual con el marco bayesiano.

La elección de la función $f(p)$ tiene un impacto significativo en el comportamiento

del intervalo. Al tratarse de una densidad de probabilidad, puede utilizarse para dar mayor peso a ciertas regiones del espacio de parámetros, dependiendo del interés práctico o del conocimiento previo sobre la prevalencia esperada. Por ejemplo, una función $f(p)$ sesgada hacia valores bajos prioriza la cobertura precisa en escenarios de baja prevalencia.

Este enfoque ofrece una alternativa flexible que conserva el control frecuente sobre la cobertura, pero con un desempeño más equilibrado y eficiente, especialmente útil en aplicaciones donde el exceso de cobertura del intervalo de Clopper-Pearson convencional representa una limitación importante.

La construcción de un intervalo con ajuste de cobertura requiere especificar una función de ponderación o densidad a priori sobre el parámetro poblacional $p$. Esta elección permite incorporar conocimiento previo sobre el posible rango de valores de $p$. En el contexto de este trabajo, enfocado en la estimación de prevalencias bajas, es razonable asumir que existe información previa que sugiere que el valor de $p$ se encuentra por debajo de 0.10, e incluso inferior a 0.01.

Bajo esta premisa, Thulin (12) propone analizar tres escenarios distintos:

1. Cuando $p$ es cercano a 0 o a 1.
2. Cuando $p$ se encuentra entre 1/4 y 3/4.
3. Cuando $p$ está aproximadamente en 0.5.

Dado el enfoque de este trabajo, se selecciona el primer escenario como base de análisis. Para representar este conocimiento previo en el ajuste de cobertura, se considera como función de ponderación una distribución Beta simétrica $f(p) \sim \text{Beta}(r, r)$, con $r < 1$. Esta elección es adecuada cuando se espera que $p$ tome valores extremos (cercanos a 0 o 1), ya que la densidad prior coloca mayor masa de probabilidad en las colas del intervalo $(0, 1)$. La cobertura media exacta del intervalo ajustado se puede calcular numéricamente resolviendo la expresion 2.12, donde $IC_{CP}(x, \alpha')$ representa el intervalo de Clopper-Pearson ajustado y $f(p)$ es la función prior (por ejemplo, $\text{Beta}(r, r)$).

En su estudio comparativo, Thulin evaluó el desempeño de los intervalos de Clopper-Pearson corregidos por cobertura frente a otros intervalos conocidos, como el intervalo de Wilson y el intervalo basado en la prior de Jeffrey. Sus resultados indican que el intervalo corregido ofrece un mejor ajuste, particularmente en los extremos del dominio de $p$ (es decir, cuando $p$ es cercano a 0 o 1). Bajo estos escenarios, el intervalo ajustado presenta:

- Coberturas reales muy cercanas al nivel nominal del 95%.
- Amplitud comparable a la de los intervalos de Wilson.

Estas características hacen del intervalo corregido por cobertura una alternativa, especialmente útil en aplicaciones de epidemiología de enfermedades raras, donde las prevalencias son bajas y se requiere tanto precisión como control sobre la cobertura estadística.

### Intervalo bajo el enfoque Bayesiano

El enfoque bayesiano proporciona una alternativa metodológicamente robusta y flexible para la estimación de proporciones, particularmente valiosa en escenarios de baja prevalencia o tamaños muestrales reducidos. A diferencia del paradigma frecuentista, que considera los parámetros poblacionales como constantes desconocidas, la estadística bayesiana los trata como variables aleatorias que poseen distribuciones de probabilidad. Esta conceptualización permite formalizar de manera coherente la incertidumbre y actualizarla a partir de los datos observados, mediante el uso del Teorema de Bayes como lo expone Gelman et al. (13).

**Fundamento teórico: Teorema de Bayes aplicado** Sea $p$ la proporción o prevalencia poblacional de una enfermedad, considerada como variable aleatoria. Supóngase que se extrae una muestra aleatoria simple de tamaño $n$, en la cual se observan $x$ casos positivos, detectados mediante un gold estándar o prueba de referencia. Bajo estas condiciones, el número de positivos $X$ se modela como una variable aleatoria con distribución binomial:

$$X \mid p \sim \text{Bin}(n, p)$$

La función de verosimilitud correspondiente, que representa la probabilidad de los datos observados para un valor dado de $p$, está dada por:

$$\mathcal{L}(p; x, n) = Pr(X = x \mid p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{2.13}$$

A esta función se le incorpora una distribución previa $\pi(p)$, que refleja el conocimiento o creencias previas sobre el parámetro $p$. La inferencia se basa en la distribución posterior, obtenida mediante el Teorema de Bayes:

$$\pi(p \mid x) = \frac{\mathcal{L}(p; x, n) \cdot \pi(p)}{\int_0^1 \mathcal{L}(p; x, n) \cdot \pi(p)\, dp} \tag{2.14}$$

**Modelo conjugado Beta-Binomial** Una elección común y conveniente para la distribución previa es la distribución Beta:

$$p \sim \text{Beta}(\alpha, \beta), \tag{2.15}$$

cuyos parámetros $\alpha > 0$ y $\beta > 0$ determinan la forma de la densidad. Esta distribución es conjugada respecto a la verosimilitud binomial, lo cual implica que la distribución posterior también es una Beta:

$$p \mid x \sim \text{Beta}(\alpha + x, \beta + n - x) \tag{2.16}$$

Este resultado permite realizar inferencia exacta de manera analítica, facilitando tanto la obtención de estimaciones puntuales como la construcción de intervalos de credibilidad.

**Intervalos de credibilidad Bayesianos** Dada la distribución posterior $p \mid x \sim \text{Beta}(\alpha + x, \beta + n - x)$, un intervalo de credibilidad $(I, S)$ con nivel $1 - \alpha$ se obtiene a partir de los cuantiles de dicha distribución:

$$I = Q_{\alpha/2}, \quad S = Q_{1-\alpha/2} \tag{2.17}$$

donde $Q_q$ representa el cuantil $q$ de la distribución posterior. Por ejemplo, para un intervalo del 95%, se toman los cuantiles 0.025 y 0.975. A diferencia del intervalo de confianza frecuentista, el intervalo Bayesiano permite afirmar que, dado el conocimiento disponible, la probabilidad de que $p$ se encuentre entre $I$ y $S$ es del 95%.

**Elección de la distribución previa** La elección de los hiperparámetros $(\alpha, \beta)$ es una decisión fundamental en el análisis bayesiano:

- **Distribución uniforme** Beta$(1, 1)$: utilizada cuando se desea expresar ausencia de información previa.
- **Distribución de Jeffreys** Beta$(0.5, 0.5)$: prior no informativa recomendada por sus propiedades de invarianza y buena cobertura frecuentista.
- **Distribuciones informativas**, como Beta$(2, 38)$ para representar creencias previas centradas en prevalencias del 5%.

**Ejemplo comparativo** En la Figura 2.2, se ilustran las distribuciones posteriores de $p$ para una muestra de $n = 20$ con $x = 3$ casos positivos, bajo tres elecciones diferentes de priors: uniforme, Jeffreys e informativa. Se observa cómo la forma y dispersión de la distribución posterior varían significativamente según la información previa incorporada.

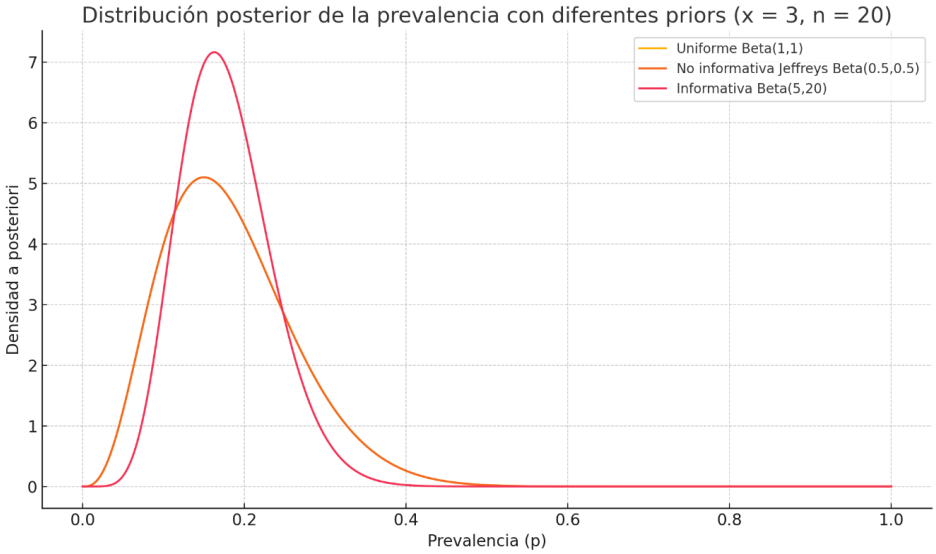


Figura 2.2: Distribución posterior de la prevalencia con distintas distribuciones prior: Beta$(1, 1)$, Beta$(0.5, 0.5)$, y Beta$(5, 20)$

### Intervalo de Jeffreys

El prior de Jeffreys es la distribución Beta$(1/2, 1/2)$, la cual genera una distribución posterior Beta$(x + 1/2, n - x + 1/2)$. El intervalo de Jeffreys se obtiene tomando el intervalo central de esta posterior. Es decir, el intervalo de Jeffreys de nivel $(1-\alpha/2)$ se define mediante los cuantiles $\alpha/2$ y $1-\alpha/2$ de una distribución Beta$(x+1/2, n-x+1/2)$ (Ecuación 2.17).

Este intervalo es invariante bajo reparametrizaciones y posee propiedades frecuentistas razonables en términos de cobertura. Para valores pequeños de $n$, la longitud del intervalo de Jeffreys tiende a ser ligeramente menor que la del intervalo de Wilson.

**Ventajas del enfoque bayesiano** El enfoque bayesiano ofrece diversas ventajas:

- Integra conocimiento previo formalmente, lo que mejora la estimación en muestras pequeñas.
- Proporciona una interpretación probabilística directa del intervalo (credibilidad).
- Permite extensión a modelos jerárquicos, ajustes por covariables y estimación en entornos complejos mediante MCMC.
- Mejora la cobertura efectiva en escenarios de prevalencia baja.

Diversos estudios por Brown et al. (5), Dalitz (14) y Cepeda et al. (15) han mostrado que los intervalos bayesianos, en particular aquellos basados en la prior de Jeffreys, tienden a ofrecer mejor balance entre precisión y cobertura nominal que los métodos clásicos como Wald o Clopper-Pearson, especialmente en el análisis de la prevalencia de enfermedades raras.

### 2.2.2 Estimación puntual y por intervalo de la prevalencia con corrección por Sensibilidad y Especificidad de un test imperfecto

En estudios epidemiológicos, es común encontrarse con la necesidad de estimar la prevalencia de una enfermedad mediante pruebas diagnósticas imperfectas o falibles, es decir, poseen una sensibilidad y especificidad inferiores a 1.0. En este escenario, Rogan et al. (6) denominaron **Prevalencia Aparente** ($AP$) a la prevalencia estimada a través de la proporción de positivos bajo la aplicación de la prueba imperfecta. Para estimar la verdadera prevalencia poblacional ($p$) de la que en la sección 2.2.1 se trató su estimación, se deben tomar en cuenta los errores sistemáticos del test que inducen sesgos importantes, es fundamental introducir correcciones basadas en las propiedades del test, cuando estas son conocidas a partir de estudios de validación independientes. A continuación se presentarán correcciones propuestas tanto en la estimación puntual como en la de intervalo para la corrección de la prevalencia aparente.

### Definición de Sensibilidad y Especificidad de un test diagnostico

El uso de pruebas diagnósticas constituye una herramienta fundamental en múltiples áreas de la salud, con el propósito de establecer la presencia o ausencia de una condición clínica, típicamente una enfermedad. Evaluar el desempeño de dichas pruebas es esencial, ya que este desempeño se relaciona directamente con su capacidad de discriminar entre individuos enfermos y sanos. De forma inherente, toda prueba diagnóstica conlleva cierto grado de error en la clasificación, lo cual introduce el concepto de errores diagnósticos.

Para valorar el rendimiento de un test diagnóstico, se utilizan distintos parámetros estadísticos. Entre ellos, destacan la sensibilidad y la especificidad, así como las razones de verosimilitud positiva y negativa, los cuales representan propiedades intrínsecas de la prueba diagnóstica, es decir, no dependen de la prevalencia de la enfermedad en la población. Por el contrario, los valores predictivos positivo y negativo, aunque dependen directamente de la sensibilidad y especificidad, sí están condicionados por la prevalencia de la condición evaluada (16).

El análisis de estos parámetros se basa tradicionalmente en una tabla de contingencia 2×2, como se muestra en la Tabla 2.1:

- Las filas representan los resultados del test: $T^+$ para resultado positivo y $T^-$ para resultado negativo.
- Las columnas reflejan el estado real de la enfermedad: $D^+$ para sujetos con la condición (enfermos) y $D^-$ para sujetos sin la condición (sanos), definido mediante una prueba de referencia o gold estándar.

Tabla 2.1: Tabla de contingencia 2×2 para los resultados de evaluación de una prueba diagnóstica binaria

| **Resultado del test** | $D^+$ **(enfermo)** | $D^-$ **(sano)** | Total fila |
|---|---|---|---|
| $T^+$ (positivo) | $a$ (VP) | $c$ (FP) | $C_1 = a + c$ |
| $T^-$ (negativo) | $b$ (FN) | $d$ (VN) | $C_2 = b + d$ |
| Total columna | $F_1 = a + b$ | $F_2 = c + d$ | $n = a + b + c + d$ |

A partir de esta tabla, se definen las siguientes categorías:

- **Verdadero positivo (VP):** Un individuo enfermo correctamente clasificado como positivo por el test ($a$).

- **Falso negativo (FN):** Un individuo enfermo clasificado incorrectamente como negativo ($b$).
- **Falso positivo (FP):** Un individuo sano clasificado incorrectamente como positivo ($c$).
- **Verdadero negativo (VN):** Un individuo sano correctamente clasificado como negativo ($d$).

La sensibilidad y la especificidad constituyen dos parámetros fundamentales para caracterizar el desempeño de una prueba diagnóstica binaria . Ambos indicadores reflejan la capacidad del test para discriminar correctamente entre individuos con y sin la condición de interés, y son comúnmente denominados medidas de exactitud diagnóstica, por cuanto evalúan propiedades intrínsecas de la prueba, independientes de la prevalencia de la enfermedad en la población objetivo.

**Sensibilidad** ($Se$) La *sensibilidad* cuantifica la capacidad de la prueba para detectar la enfermedad cuando esta está presente. Se define como la probabilidad condicional de obtener un resultado positivo en un individuo que verdaderamente presenta la condición:

$$Se = P(T^+ \mid D^+) = \frac{a}{F_1} = \frac{a}{a+b} \tag{2.18}$$

Desde el punto de vista operacional, corresponde a la proporción de verdaderos positivos sobre el total de individuos realmente enfermos. Un valor elevado de sensibilidad implica que la prueba tiene baja probabilidad de generar falsos negativos.

**Especificidad** ($Sp$) La *especificidad*, en cambio, mide la capacidad de la prueba para identificar correctamente a los individuos que no presentan la enfermedad. Se expresa como la probabilidad de obtener un resultado negativo dado que el sujeto está sano:

$$Sp = P(T^- \mid D^-) = \frac{d}{F_2} = \frac{d}{c+d} \tag{2.19}$$

Un valor alto de especificidad indica que la prueba produce pocos falsos positivos, es decir, clasifica adecuadamente a los sujetos no enfermos.

**Propiedades generales** Ambos indicadores toman valores en el intervalo $[0, 1]$, donde 1 representa un desempeño perfecto. En la práctica, es raro encontrar pruebas diagnósticas con sensibilidad y especificidad iguales a uno. Sin embargo, se considera que una prueba tiene utilidad diagnóstica razonable si ambos valores son superiores a 0.5.

Luna et al. (17) afirman que una característica clave de la sensibilidad y la especificidad es que no dependen de la prevalencia de la enfermedad en la población estudiada, dado que se condicionan sobre el estado verdadero de la enfermedad ($D^+$ o $D^-$). Por esta razón, se consideran propiedades invariantes del test, y se interpretan como métricas de su validez interna o desempeño diagnóstico intrínseco.

Estas propiedades son especialmente importantes cuando se desea ajustar las estimaciones de prevalencia observadas en poblaciones donde el test diagnóstico empleado no es perfecto, como se desarrollará en las siguientes secciones.

El conocimiento de estas definiciones es crucial cuando se desea ajustar las estimaciones de prevalencia obtenidas mediante un test imperfecto, incorporando la sensibilidad y especificidad como parámetros de corrección, a continuación se revisan algunos métodos de estimación que incorporan la corrección sobre la prevalencia aparente ($AP$).

### Estimador de Rogan-Gladen

Bajo este enfoque, uno de los métodos frecuentistas más utilizados es el propuesto por Rogan y Gladen (6) en el que se puede asumir la estimación de $Se$ y $Sp$ ser conocidas. Desde una tabla de contingencia $2 \times 2$ como la expuesta en 2.1, las siguientes definiciones operacionales:

- Verdaderos positivos (VP): $a$
- Falsos negativos (FN): $b$
- Falsos positivos (FP): $c$
- Verdaderos negativos (VN): $d$
- Tamaño muestral total: $n = a + b + c + d$
- Proporción observada de resultados positivos en la muestra: $\widehat{AP} = \frac{a+c}{n}$

Sea $p$ la prevalencia verdadera de la enfermedad, es decir, la probabilidad de que un individuo de la población esté efectivamente enfermo:

$$p = P(D^+)$$

La proporción observada de resultados positivos en la muestra corresponde a una combinación entre sujetos enfermos correctamente clasificados como positivos (verdaderos positivos) y sujetos sanos mal clasificados como positivos (falsos positivos). Esta mezcla se puede expresar como:

$$AP = P(T^+) = P(T^+ \mid D^+) \cdot P(D^+) + P(T^+ \mid D^-) \cdot P(D^-)$$

Reemplazando los términos, se obtiene:

$$\widehat{AP} = Se \cdot p + (1 - Sp) \cdot (1 - p) \tag{2.20}$$

El objetivo es dejar explicita a $p$ (prevalencia verdadera). Para ello, reorganizamos la expresión:

$$\widehat{AP} = Se \cdot p + (1 - Sp) - (1 - Sp) \cdot p$$

$$\widehat{AP} - (1 - Sp) = p \cdot (Se + Sp - 1)$$

Finalmente, despejando $p$, se obtiene el estimador de Rogan y Gladen:

$$\hat{p}_{\text{RG}} = \frac{\widehat{AP} + Sp - 1}{Se + Sp - 1} \tag{2.21}$$

**Condiciones de validez** Este estimador es válido siempre que $Se + Sp > 1$, lo cual asegura que la prueba diagnóstica tenga una capacidad de discriminación mejor que el azar. Si $1 - Sp \leq AP$, el denominador se vuelve cero o negativo, lo cual produce estimaciones donde $\hat{p}_{RG} < 1.0$ y si $AP \leq Se$ el $\hat{p}_{RG} > 1.0$.

La fórmula de Rogan y Gladen permite ajustar la estimación de prevalencia observada por los errores sistemáticos introducidos por un test imperfecto. En pruebas con alta sensibilidad y especificidad, la corrección tiende a ser mínima. Sin embargo, en contextos de baja prevalencia o desempeño diagnóstico limitado, esta corrección es fundamental para evitar sobreestimaciones o subestimaciones de la proporción de individuos enfermos en la población. Rogan et al. (6) derivo un intervalo de confianza, segun el estimador

derivado en 2.21 donde $\widehat{AP}$ es la proporción de resultados positivos observados en la muestra, y $Se$ y $Sp$ son conocidos y constantes, podemos derivar la varianza de $\hat{p}_{\text{RG}}$ utilizando propiedades de la varianza de funciones lineales.

Como la transformación es lineal en $AP$, se tiene:

$$\text{Var}(\hat{p}_{\text{RG}}) = \left(\frac{1}{Se + Sp - 1}\right)^2 \cdot \text{Var}(\widehat{AP}) \tag{2.22}$$

Ahora, dado que $\widehat{AP} = \frac{a+c}{n}$ y $x = a + c$ positivos con el test imperfecto $x \sim \text{Binomial}(n, AP)$, se tiene que:

$$\text{Var}(\widehat{AP}) = \frac{1}{n} AP\,[1 - AP] \tag{2.23}$$

como en 2.20 $AP = Se \cdot P + (1 - Sp) \cdot (1 - P)$

Se sustituye en la varianza:

$$\text{Var}(\hat{p}_{\text{RG}}) = \frac{1}{n(Se + Sp - 1)^2} \cdot AP(1 - AP) \tag{2.24}$$

Finalmente, al estimar $AP$ por $\widehat{AP}$, se obtiene un estimador práctico de la varianza de $\hat{p}_{\text{RG}}$:

$$\widehat{\text{Var}}(\hat{p}_{\text{RG}}) = \frac{1}{n(Se + Sp - 1)^2} \cdot \widehat{AP}(1 - \widehat{AP}) \tag{2.25}$$

Este resultado permite construir un intervalo de confianza aproximado para la prevalencia corregida, bajo el supuesto de normalidad asintótica:

$$IC_{(1-\alpha/2)} = \hat{p}_{\text{RG}} \pm z_{1-\alpha/2} \cdot \sqrt{\widehat{\text{Var}}(\hat{p}_{\text{RG}})} \tag{2.26}$$

Este intervalo es válido si $n$ es suficientemente grande y si la proporción observada $\widehat{AP}$ no está cercana a los extremos 0 o 1.

### Intervalo de Reiczigel

Reiczigel et al. (18) propuso la estimación por intervalo de la prevalencia por metodos exactos, para esto retoma el trabajo realizado por Blacker (19) y Sterne (20) citados en su trabajo original, donde ambos se derivan a partir de la inversión de sus respectivas pruebas estadísticas. Por consiguiente, en primer lugar define cómo pueden ajustarse dichas pruebas para tener en cuenta la sensibilidad y especificidad del test diagnóstico.

Ambas pruebas evalúan la hipótesis nula $H_0 : p = p_0$ frente a la alternativa $H_1 : p \neq p_0$, donde $p$ denota la verdadera prevalencia poblacional desconocida y $p_0$ representa un valor hipotético de dicha prevalencia.

Asumiendo $H_0$ y un modelo de muestreo binomial, la probabilidad de que una muestra aleatoria de tamaño $n$ contenga $k$ individuos con la enfermedad ($k = 0, 1, \ldots, n$) está dada por:

$$Pr(K = k \mid H_0) = \binom{n}{k} p_0^k (1 - p_0)^{n-k} \tag{2.27}$$

Dado que la muestra contiene $k$ sujetos realmente enfermos, y asumiendo que el test diagnóstico tiene sensibilidad $Se$ y especificidad $Sp$, se puede calcular la probabilidad de que el número de sujetos clasificados como positivos en la muestra sea exactamente $m$ ($m = 0, 1, \ldots, n$). Esta probabilidad está dada por:

$$P(M = m \mid K = k) = \sum_{i=0}^{m} \binom{k}{i} Se^i (1 - Se)^{k-i} \binom{n-k}{m-i} (1 - Sp)^{m-i} Sp^{(n-k)-(m-i)} \tag{2.28}$$

Combinando las ecuaciones 2.27 y 2.28, se obtiene que, bajo $H_0$, la probabilidad de observar exactamente $m$ positivos en la muestra puede expresarse como:

$$P_{H_0}(M = m) = \sum_{k=0}^{n} \binom{n}{k} p_{\text{hip}}^k (1 - p_{\text{hip}})^{n-k} \cdot \sum_{i=0}^{m} \binom{k}{i} \text{Se}^i (1 - \text{Se})^{k-i} \binom{n-k}{m-i} (1 - \text{Sp})^{m-i} \text{Sp}^{(n-k)-(m-i)} \tag{2.29}$$

El principio del método de Sterne consiste en ordenar el espacio muestral (es decir, los valores $m = 0, 1, \ldots, n$) de acuerdo con sus probabilidades bajo $H_0$. A partir de ello, se define el valor-p asociado a una observación $j$ de positivos como:

$$p_{j,H_0}^{\text{Sterne}} = \sum_{\substack{i: \\ P_{H_0}(i) \leq P_{H_0}(j)}} P_{H_0}(i) \tag{2.30}$$

donde $Pr_{H_0}(i)$ se refiere a la probabilidad definida por la ecuación 2.29.

En el caso de la prueba de Blaker, el espacio muestral se ordena utilizando una función denominada de aceptabilidad, definida como:

$$A_{H_0}(m) = \min \left\{ \sum_{i \geq m} P_{H_0}(i), \sum_{i \leq m} P_{H_0}(i) \right\} \tag{2.31}$$

A partir de esta función, el valor-p correspondiente es:

$$p_{j,H_0}^{\text{Blaker}} = \sum_{\substack{i: \\ A_{H_0}(i) \leq A_{H_0}(j)}} P_{H_0}(i) \tag{2.32}$$

La inversión de estas pruebas da lugar a intervalos de confianza bilaterales exactos para la prevalencia. Invertir el test significa que, al observar $j$ positivos en la muestra, el intervalo de nivel $1 - \alpha$ está compuesto por todos los valores hipotéticos $p_0$ tales que el valor-p correspondiente es mayor que $\alpha$.

El intervalo de Blaker presenta la ventaja de estar siempre contenido dentro del intervalo de Clopper–Pearson. Por el contrario, aunque el intervalo de Sterne suele ser ligeramente más estrecho en promedio, en algunas situaciones puede ser incluso más amplio que el intervalo de Clopper–Pearson.

Es importante señalar que los intervalos propuestos no distribuyen equitativamente el error en ambos extremos, por lo que no es posible derivar intervalos unilaterales directamente a partir de los bilaterales de la forma habitual. En consecuencia, si se requiere un intervalo exacto unilateral, se debe emplear el método de Clopper–Pearson.

Reiczigel et al. finalmente encontraron que según los resultados de su simulación, el intervalo de Sterne tiende a ser más estrecho que el de Blaker cuando la prevalencia verdadera está entre el 30% y el 70%, mientras que es más ancho para prevalencias inferiores al 20% o superiores al 80%. En casos extremos de baja prevalencia y sensibilidad con alta especificidad, el intervalo de Sterne puede incluso ser más largo que el de Clopper-Pearson. El intervalo de Blaker tiene la ventaja de que siempre está contenido

dentro del intervalo de Clopper-Pearson, por lo que su longitud nunca excede la de este último. Además, en algunos casos, el intervalo de Blaker resultó ser más estrecho que el intervalo de Wilson, manteniendo al mismo tiempo una cobertura más alta.

Dentro de este estudio también se documentó que ajustar los límites de confianza obtenidos para la prevalencia aparente mediante la fórmula de Rogan-Gladen (Ecuación (2.21) resulta en un intervalo de confianza exacto para la prevalencia verdadera, lo que facilita la implementación del método. Sin embargo, calcular un intervalo de confianza utilizando la prevalencia ajustada como si se hubiera observado directamente es un método inapropiado que conduce a intervalos incorrectos con una cobertura real muy baja.

Se destaca la importancia de utilizar métodos exactos, como los intervalos de Blaker y Sterne, para la construcción de intervalos de confianza para la prevalencia ajustada por sensibilidad y especificidad, especialmente para tamaños de muestra no muy grandes. Reiczigel et al. recomiendan el uso general del intervalo de Blaker debido a su buen comportamiento en diferentes escenarios y a que su longitud nunca excede la del intervalo de Clopper-Pearson.

### Intervalo de Agresti-Coull modificado

Lang et al. (21) proponen una alternativa práctica para la corrección de la mala clasificación en la estimación de la prevalencia verdadera; para esto retoman el intervalo para una proporción propuesto por Agresti y Coull (7).

Sean $Se$ y $Sp$ obtenidas a partir de un estudio independiente de una prueba diagnóstica imperfecta utilizada para estimar la prevalencia $p$ de una condición o enfermedad. La prevalencia aparente ($\widehat{AP}$) es la proporción de resultados positivos obtenidos mediante la prueba en la población objetivo. Dado el estimador $\widehat{p}_{RG}$ puntual de $p$ mostrado en la ecuación (2.21), si se aplica el ajuste al intervalo propuesto por Agresti y Coull sobre $\widehat{AP}$:

$$\tilde{n} = n + z_{\alpha}^{2}$$

$$\widehat{AP}' = \frac{n \cdot \widehat{AP} + \frac{z_{\alpha/2}^{2}}{2}}{\tilde{n}}$$

Así, tanto la estimación puntual corregida de $p$ como su intervalo de confianza se construyen de la siguiente manera:

$$\tilde{p} = \frac{\widehat{AP'} + Sp - 1}{Se + Sp - 1} \tag{2.33}$$

$$\tilde{p} \pm z_{\alpha/2} \frac{1}{(Se + Sp - 1)} \left( \frac{\widehat{AP'}(1 - \widehat{AP'})}{n'} \right)^{1/2} \tag{2.34}$$

El estudio de este intervalo propuesto por Lang et al. mostró que mantienen el nivel nominal de confianza, incluso para tamaños de muestra tan pequeños como $n = 30$. La cobertura mínima se mantuvo por encima del 88%, 93% y 98% para niveles nominales del 90%, 95% y 99%, respectivamente.

El IC de Lang para la prevalencia verdadera proporciona una buena cobertura para una amplia gama de valores de prevalencia, sensibilidad y especificidad, incluyendo prevalencias cercanas a 0 y 1, y sensibilidad y especificidad cercanas a 1. Debido a los ajustes aplicados, tiende a ser algo conservador cuando la prevalencia está cerca de 0 o 1.

### Intervalo bajo enfoque Bayesiano

Diferentes propuestas sobre un enfoque Bayesiano han sido realizadas Lew (22), Joseph (23) y Fraser (24), resaltando su versatilidad y mayor precisión de los estimados, además de solventar los problemas ya evidenciados en los métodos frecuentistas , más detalles en Speybroeck (25), especialmente en el estimador de Rogan-Gladen el cual puede dar estimaciones de prevalencia por debajo de cero ante el no cumplimiento de la condición de validez ($Se + Sp \leq 1$).

En un marco Bayesiano, los valores creíbles de los parámetros se describen mediante distribuciones de probabilidad. En el modelo Bayesiano de estimación de prevalencia, el conocimiento previo o la creencia sobre la prevalencia real, así como sobre la sensibilidad y especificidad de la prueba diagnóstica, se expresa en términos de distribuciones de probabilidad. En este contexto, el conocimiento previo se deriva de estudios de validación de las propiedades de la prueba diagnóstica, pero en otras situaciones también puede provenir de la opinión de expertos. A partir de los datos de la aplicación de la prueba para diagnosticar muestras de una población, el modelo actualiza las distribuciones de probabilidad que posteriormente describen el conocimiento posterior sobre la prevalencia real.

En un marco Bayesiano, la prevalencia verdadera ($p$) se modela como un parámetro aleatorio con una distribución a priori asociada. La sensibilidad ($Se$) y la especificidad ($Sp$) de la prueba diagnóstica, previamente definidas, se tratan como parámetros conocidos obtenidos a partir de estudios independientes. La probabilidad de observar $x$ casos positivos en una muestra de tamaño $n$, asumiendo que la probabilidad de éxito está dada por la prevalencia aparente, se modela mediante una distribución binomial $x \sim \text{Binomial}(n, AP)$.

La prevalencia aparente ($\widehat{AP}$) previamente descrita en la ecuación 2.20 presenta relación con la prevalencia verdadera $p$, así como con la sensibilidad y especificidad de la prueba diagnóstica. La función de verosimilitud para observar $x$ casos positivos es:

$$Pr(x|n, P, Se, Sp) = \binom{n}{x} \cdot AP^x \cdot (1 - AP)^{n-x}. \tag{2.35}$$

Basándonos en el teorema de Bayes, la distribución posterior de $p$ se expresa como:

$$Pr(P|x, n, Se, Sp) \propto Pr(x|n, P, Se, Sp) \cdot Pr(p) \tag{2.36}$$

$$p \sim \text{Beta}(1, 1) \tag{2.37}$$

Sustituyendo la verosimilitud binomial y la distribución a priori no informativa, obtenemos:

$$\begin{aligned}\mathbf{Pr}(p|x, n, Se, Sp) \propto \binom{n}{x} \cdot (p \cdot Se + (1 - p) \cdot (1 - Sp))^x \\ \cdot (1 - (p \cdot Se + (1 - p) \cdot (1 - Sp)))^{n-x} \cdot \text{Beta}(1, 1)\end{aligned}$$

La distribución posterior se simplifica a:

$$\begin{aligned}\mathbf{Pr}(p|x, n, Se, Sp) \propto (p \cdot Se + (1 - p) \cdot (1 - Sp))^x \cdot \\ (1 - (p \cdot Se + (1 - p) \cdot (1 - Sp)))^{n-x}\end{aligned} \tag{2.38}$$

Métodos numéricos como MCMC (Markov Chain Monte Carlo) pueden utilizarse para obtener la distribución posterior de $p$. Este enfoque permite la generación de muestras a partir de la distribución posterior de $p$ y la realización de inferencias sobre la prevalencia verdadera.

En algunos casos, para la implementación de las estimaciones Bayesianas, es posible que la información sobre $Se$ y $Sp$ no se encuentre disponible como parámetros puntuales, sino como resultado de estudios previos de validación de la prueba diagnóstica.

En tales casos, es posible incorporar explícitamente la incertidumbre asociada a $Se$ y $Sp$ modelándolas como variables aleatorias, informadas por los datos del estudio de validación. Por ejemplo, si en un estudio previo se identificaron $x_{Se}$ verdaderos positivos entre $n_{Se}$ individuos con la enfermedad, se puede modelar la sensibilidad como:

$$Se \sim \text{Beta}(x_{Se} + 1, n_{Se} - x_{Se} + 1) \tag{2.39}$$

De forma análoga, si $x_{Sp}$ sujetos sanos fueron correctamente clasificados como negativos entre un total de $n_{Sp}$ individuos sin la enfermedad, la especificidad se modela como:

$$\text{Sp} \sim \text{Beta}(x_{Sp} + 1, n_{Sp} - x_{Sp} + 1) \tag{2.40}$$

Estas distribuciones Beta actúan como distribuciones a posteriori conjugadas bajo una observación binomial, y permiten capturar la incertidumbre inherente en los parámetros diagnósticos. Al combinar estas distribuciones con la información observada en el estudio de prevalencia —por ejemplo, el número de positivos observados $x$ en una muestra de tamaño $n$— se puede construir un modelo jerárquico Bayesiano que incorpora la incertidumbre conjunta sobre la sensibilidad, especificidad y prevalencia.

Este enfoque permite una propagación coherente de la incertidumbre diagnóstica hacia la estimación de la prevalencia, produciendo intervalos de credibilidad más realistas, especialmente en estudios con tamaños de muestra pequeños o prevalencias bajas.

En las secciones siguientes se describirá el modelo jerárquico completo y las estrategias de inferencia computacional necesarias para su implementación.

Teniendo los datos de $x$ número de casos positivos observados en una muestra de tamaño $n$, con parámetros $p$, $Se$ y $Sp$, previamente descritos en ecuaciones (2.18) y (2.19) y la prevalencia aparente como en (2.20) Dado esto, la distribución del número

de positivos observados en la muestra es:

$$x \mid p, Se, Sp, n \sim \text{Binomial}(n, AP) \tag{2.41}$$

Para reflejar la incertidumbre sobre $p$, $Se$ y $Sp$, se asignan distribuciones Beta como priors conjugados, como se describió previamente en las ecuaciones 2.39, 2.40 y 2.37.

El modelo jerárquico completo puede resumirse de la siguiente manera:

$$\begin{aligned} x \mid p, Se, Sp, n &\sim \text{Binomial}(n, p \cdot Se + (1-p)(1-Sp)) \\ Se &\sim \text{Beta}(x_{Se} + 1, n_{Se} - x_{Se} + 1) \\ Sp &\sim \text{Beta}(x_{Sp} + 1, n_{Sp} - x_{Sp} + 1) \\ p &\sim \text{Beta}(1, 1) \end{aligned}$$

La distribución posterior conjunta se expresa, hasta una constante de proporcionalidad, como:

$$\begin{aligned} Pr(p, \text{Se}, \text{Sp} \mid x, n, x_{\text{Se}}, n_{\text{Se}}, x_{\text{Sp}}, n_{\text{Sp}}) \propto \binom{n}{x} \cdot [p \cdot \text{Se} + (1-p)(1-\text{Sp})]^x \cdot \\ [1 - p \cdot \text{Se} - (1-p)(1-\text{Sp})]^{n-x} \cdot \\ \text{Beta}(\text{Se} \mid x_{\text{Se}} + 1, n_{\text{Se}} - x_{\text{Se}} + 1) \cdot \\ \text{Beta}(\text{Sp} \mid x_{\text{Sp}} + 1, n_{\text{Sp}} - x_{\text{Sp}} + 1) \cdot \\ \text{Beta}(p \mid \alpha, \beta) \end{aligned} \tag{2.42}$$

Este modelo jerárquico puede ser implementado mediante técnicas de inferencia computacional como el MCMC. El intervalo de credibilidad del 95% para la prevalencia verdadera corregida se obtiene a partir de los cuantiles 2.5% y 97.5% de la distribución posterior de $p$:

$$IC_{95\%} = (Q_{0.025}, Q_{0.975}) \tag{2.43}$$

Este intervalo refleja la región más plausible para la prevalencia verdadera en vista de los datos observados y la incertidumbre sobre la sensibilidad y especificidad del test

diagnóstico.

Flor et al. (26) compararon diferentes métodos de estimación de prevalencia bajo error de clasificación, incluyendo métodos frecuentistas (Rogan-Gladen, Lang y Reiczel) y estimación Bayesiana, encontrando que la estimación puntual Bayesiana como la estimación de Rogan-Gladen mostraron distribuciones de error similares, con una ligera ventaja para el método Bayesiano.

Si bien el estimador puntual de Rogan-Gladen puede ofrecer estimaciones de prevalencia consistentes en general, presenta limitaciones significativas, especialmente el problema de truncamiento en cero o uno y la fuerte subcobertura de los intervalos de confianza construidos con métodos frecuentistas tradicionales. En contraste, el método Bayesiano de estimación de prevalencia, utilizando la media de la distribución posterior como estimador puntual y el intervalo entre percentiles como intervalo de credibilidad, se desempeña de manera superior.

### 2.2.3 Estimación puntual y por intervalo de la prevalencia por un test imperfecto y verificación parcial por gold estandar

La disponibilidad de una prueba gold estándar puede ser invasiva y/o costosa, lo que significa que en la práctica no se puede aplicar a todos los sujetos del estudio (1).

Es muy probable que a los sujetos que parecen tener un alto riesgo de enfermedad se les ofrezca la prueba estándar de oro, mientras que los sujetos con menor riesgo pueden tener menos probabilidades de someterse a ella. Este tipo de pruebas selectivas puede causar sesgos en la precisión estimada de la prevalencia de la enfermedad, a menos que se tengan debidamente en cuenta en la estimación. En algunos contextos, la prueba de gold estándar es muy invasiva, no es lo suficientemente costo-efectiva y, por razones éticas y de costos en salud, solo se puede aplicar a los sujetos que se consideran de alto riesgo de enfermedad según el resultado positivo de una prueba de detección. En tales casos, las probabilidades de error asociadas con la prueba de detección no son identificables.

En el caso específico de la estimación de condiciones (enfermedades) que son de baja prevalencia, esto es, enfermedades raras, enfermedades infecciosas de muy baja endemicidad local, se hace impráctico durante un estudio someter a la totalidad de la muestra seleccionada a dicha prueba gold estándar. De esta manera surgen alternativas, tanto en la aplicación de múltiples pruebas imperfectas con verificación al final, a sujetos que finalmente cumplen con criterios de positividad en dichos test de tamizaje, optimizando la implementación del estudio en términos de costos y facilidad de desarrollar

la estimación de este tipo de enfermedades con baja probabilidad de existencia entre la población estudiada.

Estrategias metodológicas como el uso de test imperfecto con aplicación condicionada (total de positivos o fracción de ellos y/o a una fracción de negativos) con prueba gold estándar, contienen implícitamente un problema de sesgo de verificación parcial, problema de estimación que ha sido abordado por diferentes autores, Mcnamee (27), Pickles (28), quienes a su vez plantearon sus trabajos con base en lo publicado previamente por Shrout (29).

Shrout et al. (29) propusieron un diseño óptimo para encuestas de prevalencia en dos fases dirigidas al estudio de trastornos raros (baja prevalencia), especialmente en el contexto de recursos limitados y la necesidad de incluir a todos los individuos que dieron positivo en la fase de cribado para una evaluación diagnóstica más exhaustiva. Su trabajo también se centró en comparar la eficiencia relativa de este diseño de dos fases con un diseño de una sola fase, donde todos los recursos se destinan a una única encuesta que emplea la evaluación diagnóstica. La propuesta de Shrout y Newman aborda situaciones en las que una evaluación inicial (test imperfecto), menos costosa, se aplica a una muestra grande, seguida de una evaluación diagnóstica (gold estándar) más completa en una submuestra. Se destaca que, para enfermedades raras (huérfanas), el diseño que minimiza la varianza del estimador de prevalencia a menudo requiere que el 100% de los individuos que dieron positivo en el test inicial se incluyan en la segunda fase de evaluación. Esto también se alinea con consideraciones éticas, especialmente cuando existen posibles intervenciones para aquellos que padecen el trastorno.

### Estimación de la prevalencia en dos fases

Sin pérdida de generalidad y bajo la notación propuesta por Shrout, Sea $D$ ($\bar{D}$) la presencia (ausencia) de un trastorno según lo indique una prueba gold estándar, y sea $p$ la prevalencia del trastorno. Sea $S$ ($\bar{S}$) un resultado positivo (negativo) de una prueba de tamizaje, y sea $\Pr(S) = \pi$ y $\Pr(\bar{S}) = 1-\pi$. Definimos $\lambda_1 = \Pr(D|S)$ y $\lambda_2 = \Pr(D|\bar{S})$, es decir, las prevalencias del trastorno en los grupos con resultado positivo y negativo al tamizaje, respectivamente. Se asume que $\lambda_1 > \lambda_2$.

En una encuesta en dos fases, se toma una muestra de tamaño $N_1$ de una población (asumida como grande en relación con $N_1$), se administra la prueba de tamizaje y se clasifica en $S$ y $\bar{S}$. Dentro del grupo $S$, una proporción $f_1$ se selecciona aleatoriamente para seguimiento en la fase 2, y dentro de $\bar{S}$ una proporción $f_2$ también se selecciona para seguimiento. Se asume que $N_1$ es suficientemente grande para que $N_1\pi f_1$ y $N_1(1-\pi)f_2$

no sean cero. La Tabla 2.2 muestra los valores esperados en las celdas de la encuesta.

Tabla 2.2: Valores esperados en las celdas de una encuesta en dos fases

| **Fase 1: Test imperfecto** | **Fase 2: Gold estándar** | | | **Total** |
|---|---|---|---|---|
| | $D$ | $D$ | No evaluado | |
| $S$ | $\lambda_1 f_1 \pi N_1$ | $(1-\lambda_1) f_1 \pi N_1$ | $(1-f_1)\pi N_1$ | $\pi N_1$ |
| $\bar{S}$ | $\lambda_2 f_2 (1-\pi) N_1$ | $(1-\lambda_2) f_2 (1-\pi) N_1$ | $(1-f_2)(1-\pi)N_1$ | $(1-\pi)N_1$ |

Una estimación de $\pi$ se obtiene a partir de la proporción muestral en $S$, y las estimaciones de $\lambda_1$ y $\lambda_2$ se derivan de las proporciones de individuos con el trastorno entre los evaluados en $S$ y $\bar{S}$. La estimación de $p$ y su varianza en muestras grandes se dan por las ecuaciones 2.44 y 2.45:

$$\hat{p} = \hat{\pi}\hat{\lambda}_1 + (1-\hat{\pi})\hat{\lambda}_2; \tag{2.44}$$

$$\text{var}(\hat{p}) = \frac{1}{N_1}\left[\frac{\lambda_1(1-\lambda_1)}{f_1\pi} + \frac{(1-\pi)\lambda_2(1-\lambda_2)}{f_2} + \pi(1-\pi)(\lambda_1-\lambda_2)^2\right] \tag{2.45}$$

Suponiendo que cada test imperfecto cuesta $c_s$ unidades y cada evaluación por el gold estándar cuesta $c_D$. Bajo la restricción de que el costo total es fijo, se pueden definir valores óptimos de $f_1$ y $f_2$ que minimizan la ecuación 2.45:

$$f_1^* = \left(\frac{\lambda_1(1-\lambda_1)}{(\lambda_1 - X_2)^2\pi}\frac{c_s}{c_D}\frac{1-\pi}{\pi}\right)^{1/2} \tag{2.46}$$

$$f_2^* = \left(\frac{\lambda_2(1-\lambda_2)}{(\lambda_1-\lambda_2)^2(1-\pi)}\frac{c_s}{c_D}\frac{\pi}{1-\pi}\right)^{1/2} \tag{2.47}$$

Se asume que $f_1$ y $f_2$ pertenecen al intervalo $(0, 1]$, aunque en la práctica no es infrecuente encontrar combinaciones de $\pi$, $\lambda_1$ y $\lambda_2$ que producen valores de $f_1$ y $f_2$ mayores que 1.0, incluso para razones $c_s/c_D$ pequeñas. Esto ocurre especialmente cuando el trastorno estudiado es raro y el test de la fase 1 está débilmente relacionado con el gold estándar, ya que tanto $\lambda_1$ como $\lambda_2$ pueden ser pequeños, y la diferencia cuadrática

entre ellos extremadamente baja. Como $\lambda_1$ tiende a ser mayor que $\lambda_2$, es posible que $f_1$ supere 1.0, incluso cuando $f_2$ no lo haga.

Cuando todas las personas con resultado en fase 1 positivas, por razones éticas son llevadas a la aplicación de gold estándar, el valor óptimo de $f_2$ es aquel que minimiza 2.45 cuando $f_1$ se fija en 1.0. Suponiendo costos $c_s$ y $c_D$ como antes, y que los recursos disponibles permiten seguir a todos los individuos en $S$, el valor óptimo de $f_2$ es:

$$f_2^{**} = \frac{\sqrt{\lambda_2(1-\lambda_2)(\pi + c_s/c_D)}}{\sqrt{\lambda_1(1-\lambda_1) + \pi(1-\pi)(\lambda_1-\lambda_2)^2}} \tag{2.48}$$

Si $f_2^{**} > 1.0$, el diseño en dos fases es menos eficiente que una encuesta de una sola fase.

Si bien el diseño y los estimadores $\hat{p}$ y var$(\hat{p})$ muestran una aplicabilidad en los casos de prevalencias de enfermedades huérfanas, por la optimización en costos y eficiencia estadística, en salud pública, en donde el uso de programas de tamizaje poblacional son de aplicación frecuente, surge la necesidad de estimaciones bajo esquemas donde $f_1 = 1.0$ y $f_2 = 0$, es decir, el 100% de sujetos positivos de $S$ son evaluados por gold estándar, con el fin de confirmar la presencia de la enfermedad, mientras que ninguno de los sujetos en $\bar{S}$ va a confirmación por gold estándar, debido a que el test inicial aplicado no identifica riesgo de padecimiento de la enfermedad. Este esquema es frecuente verlo en tamizajes de enfermedades infecciosas como VIH o en enfermedades crónicas como el cáncer. Para abordar este problema en la estimación de la prevalencia de la enfermedad, recientemente Thomas et al. (30) plantearon una solución bajo los dos enfoques, frecuentista y bayesiano, en el escenario en que las $Se$ y $Sp$ del test imperfecto son fijas y conocidas. A continuación se presentan los enfoques planteados.

### Estimación de la prevalencia con verificación por gold estándar solo entre positivos

En este contexto, Thomas et al. (30) derivaron estimadores de máxima verosimilitud para la prevalencia de la enfermedad utilizando enfoques tanto frecuentistas como Bayesianos, asumiendo que la sensibilidad ($Se$) y especificidad ($Sp$) de la prueba de tamizaje son conocidas con certeza.

#### Enfoque frecuentista .

Tabla 2.3: Tabla de contingencia para datos con verificación parcial por gold estándar solo en positivos

| **Gold Estándar X** | **Test diagnóstico Z** | | **Total** |
|---|---|---|---|
| | **Positivo (Z = 1)** | **Negativo (Z = 0)** | |
| **Positivo (X = 1)** | $n_{11}$ | $n_{10}$ | $n_{1.}$ |
| **Negativo (X = 0)** | $n_{01}$ | $n_{00}$ | $n_{0.}$ |
| **Total** | $n_{.1}$ | $n_{.0}$ | $n$ |

En la tabla 2.3 se pueden observar los datos $(Z_i = z_i, X_iZ_i = c_i)$ para todos los $n$ individuos en el estudio. Esto significa que para los individuos que dieron negativo en la prueba de detección $(Z_i = 0)$, su estado de enfermedad $(X_i)$ es desconocido. Sin embargo, conocemos el número total de test negativos $(n_{.0})$, el número de falsos positivos $(n_{01})$ y el número de verdaderos positivos $(n_{11})$ y $p$ es la prevalencia junto con la sensibilidad y especificidad.

La función de verosimilitud $\mathcal{L}_1(p)$ se define como la probabilidad de observar los datos dados un valor del parámetro de interés, en este caso, la prevalencia $p$. Matemáticamente, se expresa como el producto de las probabilidades de cada observación individual:

$$\mathcal{L}1(p) = \prod_{i=1}^{n} P(Z_i = z_i, X_iZ_i = c_i) \tag{2.49}$$

Esta probabilidad conjunta se puede descomponer considerando los dos posibles resultados de la prueba de detección ($Z_i = 0$ o $Z_i = 1$):

$$\mathcal{L}1(p) = \prod_{i:z_i=0} P(Z_i = 0) \times \prod_{i:z_i=1} P(X_i = 1, Z_i = 1)^{c_i} P(X_i = 0, Z_i = 1)^{1-c_i}$$

Ahora, para cada componente por separado utilizando las definiciones de sensibilidad $(Se = P(Z_i = 1|X_i = 1))$ y especificidad $(Sp = P(Z_i = 0|X_i = 0))$.

Para los individuos que dieron negativo en la prueba de detección $(Z_i = 0)$: La probabilidad de que un individuo dé negativo en la prueba de detección se puede obtener marginalizando sobre su estado real de la enfermedad:

$$P(Z_i = 0) = P(Z_i = 0|X_i = 1)P(X_i = 1) + P(Z_i = 0|X_i = 0)P(X_i = 0)$$

$$P(Z_i = 0) = (1 - Se)p + Sp(1 - p)$$

Dado que hay $n_{.0}$ individuos con $Z_i = 0$, la contribución de estos individuos a la función de verosimilitud es $[(1 - Se)p + Sp(1 - p)]^{n_{.0}}$.

Para los individuos que dieron positivo en la prueba de detección ($Z_i = 1$) se realiza la verificación con el gold estándar, por lo que observamos $c_i = X_i$. Tenemos dos posibles resultados:

Verdadero positivo ($X_i = 1, Z_i = 1$): La probabilidad de esto es:

$$P(X_i = 1, Z_i = 1) = P(Z_i = 1|X_i = 1)P(X_i = 1) = Se \cdot p$$

.

Hay $n_{11}$ tales individuos. La contribución a la verosimilitud es:

$$(Se \cdot p)^{n_{11}}$$

.

Falso positivo ($X_i = 0, Z_i = 1$) la probabilidad de es:

$$P(X_i = 0, Z_i = 1) = P(Z_i = 1|X_i = 0)P(X_i = 0) = (1 - Sp)(1 - p)$$

Hay $n_{01}$ tales individuos (donde $n_{01} = n_{.1} - n_{11}$). La contribución a la verosimilitud es:

$$((1 - Sp)(1 - p))^{n_{01}}$$

Combinando las contribuciones de los individuos con resultados negativos y positivos en la prueba de detección, obtenemos la función de verosimilitud:

$$\mathcal{L}(p) = [(1 - Se)p + Sp(1 - p)]^{n_{.0}}(Se \cdot p)^{n_{11}}((1 - Sp)(1 - p))^{n_{01}} \tag{2.50}$$

El estimador de máxima verosimilitud (MLE) de la prevalencia se obtiene al diferenciar el logaritmo de la verosimilitud con respecto a $p$:

$$\log \mathcal{L}(p) = n_{.0} \log[(1 - Se)p + Sp(1 - p)] + n_{11} \log(Sep) + n_{01} \log[(1 - Sp)(1 - p)]$$

donde $c$ es un constante que no depende de $p$, derivando $\ln L(p)$ with respect to $p$, tenemos:

$$\frac{d \ln L(p)}{dp} = n_{.0} \frac{1 - Se - Sp}{(1 - Se)p + Sp(1 - p)} + \frac{n_{11}}{p} - \frac{n_{01}}{1 - p} \tag{2.51}$$

$$p \neq 0, 1, \frac{Sp}{Sp - (1 - Se)}$$

igualando a cero la derivada (2.51) se obtiene una ecuación cuadrática en $p$, $Ap^2 + Bp + C = 0$ cuya solución explícita es:

$$\hat{p} = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \tag{2.52}$$

donde los términos $A$, $B$ y $C$ están definidos como:

$$A = -(1 - Se - Sp)n$$

$$B = (1 - Se - Sp)(n_{.0} + n_{11}) - Sp(n_{11} + n_{01})$$

$$C = n_{11}Sp$$

Aquí, $n$ representa el tamaño total de la muestra (la suma de los individuos que dieron positivo y negativo en la prueba de tamizaje).

La varianza asintótica de $\hat{p}$ se deriva de la teoría estándar de muestras grandes para estimadores de máxima verosimilitud, donde se obtiene a partir del inverso de la información de Fisher (es decir, la derivada segunda negativa del logaritmo de la verosimilitud (2.53)) evaluada en $\hat{p}$:

$$\frac{d^2 \ln L(p)}{dp^2} = -\frac{n_{11}}{p^2} - \frac{n_{01}}{(1 - p)^2} - \frac{n_{.0}(Se + Sp - 1)^2}{((1 - Se)p + Sp(1 - p))^2} \tag{2.53}$$

$$
\begin{aligned}
I(p) &= \mathbb{E}_{\hat{p}}\left[-\frac{d^2 \ln L(p)}{dp^2}\right] \\
&= \mathbb{E}_{\hat{p}}\left[\frac{n_{11}}{\hat{p}^2} + \frac{n_{01}}{(1-\hat{p})^2} + \frac{n_{.0}(Se+Sp-1)^2}{((1-se)\hat{p}+Sp(1-\hat{p}))^2}\right] \\
&= \left(\frac{1}{\hat{p}}\right)^2 \mathbb{E}_{\hat{p}}\left[n_{11}\right] + \left(\frac{1}{1-\hat{p}}\right)^2 \mathbb{E}_{p}\left[n_{01}\right] + \left(\frac{1-Se-Sp}{(1-Se)p+Sp(1-p)}\right)^2 \mathbb{E}_{\hat{p}}\left[n_{.0}\right] \\
&= n\left[\frac{Se}{\hat{p}} + \frac{1-Sp}{1-p} + \frac{(1-Se-Sp)^2}{(1-Se)\hat{p}+sp(1-\hat{p})}\right]
\end{aligned}
$$

luego $\sigma^2 = \frac{n}{I(p)}$, se obtiene:

$$
\hat{\sigma}^2 = \left[\frac{Se}{\hat{p}} + \frac{1-Sp}{1-\hat{p}} + \frac{(1-Se-Sp)^2}{(1-Se)\hat{p}+Sp(1-\hat{p})}\right]^{-1} \tag{2.54}
$$

Usando esta varianza, se puede construir un intervalo de confianza asintótico del $100(1-\alpha)\%$ de la siguiente manera:

$$
\hat{p} \pm z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}} \tag{2.55}
$$

**Enfoque Bayesiano** Partiendo de la verosimilitud definida 2.50 y asumiendo que $Se$ y $Sp$ son conocidos y fijos para la prueba de tamizaje, la prevalencia se estima como el parámetro de interés bajo una distribución a priori no informativa, definida como $P \sim \text{Beta}(1,1)$. Así, la distribución posterior de $p$ se expresa como:

$$
\mathbf{Pr}\left(Pr \mid \mathbf{D}\right) \propto \mathbf{P}\left(\mathbf{D} \mid p\right) \cdot \mathbf{Pr}(p) \tag{2.56}
$$

donde $\mathbf{D}$ corresponde a los verdaderos positivos ($n_{11}$), los falsos positivos entre los que resultaron positivos en la prueba ($n_{01}$), el total de negativos en la prueba ($n_{.0}$), y la sensibilidad y especificidad de la prueba ($Se, Sp$).

$$
\mathbf{Pr}\left(p \mid \mathbf{D}\right) = [(1-Se)p + Sp(1-p)]^{n_{.0}}(Se \cdot p)^{n_{11}}((1-Sp)(1-p))^{n_{01}} \cdot \text{Beta}(1,1) \tag{2.57}
$$

Finalmente, para obtener la distribución posterior de $p$ (2.38), pueden implementarse métodos MCMC (Markov Chain Monte Carlo), permitiendo calcular un intervalo creíble

para $p$ a partir de la distribución posterior estimada.

## 2.3 Aplicación práctica: Estimación de la prevalencia de Quilomicronemia familiar

### 2.3.1 Síndrome de Quilomicronemia Familiar FCS

La hipertrigliceridemia (HTG) es una condición médica común asociada con niveles elevados de lipoproteínas de muy baja densidad (VLDL) y quilomicrones, y en ciertos casos, y acorde a lo reportado por Cepeda (31) y Tamez (32) se asocia con un mayor riesgo cardiovascular y pancreatitis. Hegele (33) y Virani (34), basados en datos genéticos recientes, redefinieron el trastorno en dos tipos: HTG severa, con niveles de triglicéridos (TG) mayores a 10 mmol/L ($\geq$ 880 mg/dL) o $\geq$ 500 mg/dL, que es más probable que tenga una causa monogénica, y HTG leve a moderada, con niveles de TG entre 2–10 mmol/L (150–880 mg/dL) o 150–500 mg/dL. Herrera Del Águila (35) afirma que la HTG severa puede ser producida por causas primarias asociadas con trastornos genéticos en el metabolismo de los lípidos y por causas secundarias, como el alcoholismo, el tabaquismo, las enfermedades biliares y la diabetes no controlada.

El FCS es un trastorno del metabolismo lipídico, especialmente de los quilomicrones, en el cual los triglicéridos (TG) se encuentran severamente elevados. Segun lo reportado por Davidson (36), Bchetnia (37) y Gaudet (38) se estima que podría haber aproximadamente entre 3000 y 5000 pacientes con FCS en todo el mundo . A nivel general, se han reportado tasas de prevalencia de 1 a 2 casos por cada 1.000.000 de habitantes, y en algunas poblaciones con efecto fundador, se han encontrado tasas de prevalencia de 1 caso por cada 10.000 habitantes.

Hasta la fecha, solo 166 pacientes con síndrome de quilomicronemia familiar han sido diagnosticados y confirmados mediante estudios genéticos moleculares en Colombia, debido a las limitaciones en el acceso de los pacientes a los servicios de salud y a laboratorios especializados en el país.

### 2.3.2 Diseño del estudio y estrategia de muestreo

Se desarrolló un estudio de corte transversal descriptivo de base hospitalaria, orientado a la detección de casos de Quilomicronemia familiar en adultos. El diseño siguió una estrategia de cribado en dos fases, recomendada para enfermedades raras, conforme a los

lineamientos metodológicos de Shrout (29) y McNamee (27) para estudios de prevalencia.

En la primera fase de cribado, se identificaron pacientes con la aplicación del puntaje clínico de síndrome de quilomicronemia familiar (FCS score), el cual posee una sensibilidad de 0.88 y una especificidad de 0.85 previamente estimadas por Moulin (39). El test es considerado positivo para aquellos sujetos con puntajes superiores o iguales a 8.

La segunda fase consistió en confirmación por gold estándar (prueba molecular para la identificación de mutación de la enfermedad) realizada solo al 100% de los positivos en la primera fase.

### 2.3.3 Criterios de inclusión y exclusión

- **Criterios de inclusión**: adultos mayores de 18 años con al menos una medición de TG registrados en el sistemas de información hospitalaria de la Clinica Comfamiliar, Pereira, Colombia entre los años 2010 a 2020.
- **Criterios de exclusión**: Individuos con dislipidemias mixtas (definidas como LDL $\geq$ 100 mg/dL y/o APOB elevado).

La recolección de datos se efectuó en dos etapas:

1. Extracción de registros clínicos históricos a través de bases de datos de laboratorio institucionales, filtrando los resultados conforme a los criterios establecidos.
2. Revisión manual de historias clínicas para excluir causas secundarias y recoger variables clínicas adicionales (presencia de hipertensión, diabetes, pancreatitis previa, obesidad, entre otros).

La información fue sistematizada mediante la plataforma RedCap para asegurar la estandarización y el control de calidad en la captura de datos.

### 2.3.4 Estudios genéticos

Los pacientes seleccionados con FCS score $\geq$ 8 fueron sometidos a secuenciación dirigida de genes asociados a Quilomicronemia severa, incluyendo *APOA5*, *APOC2*, *GPIHBP1*, *LMF1* y *LPL*. La secuenciación fue realizada sobre exoma completo con cobertura superior al 98% y profundidad mínima de lectura de 20X. Las variantes fueron analizadas en las bases de datos ClinVar, HGMD, LOVD, dbSNP y gnomAD, y clasificadas conforme a su significancia clínica (patogénicas, probablemente patogénicas, variantes de significado incierto).

### 2.3.5 Análisis estadístico metodológico

Con el fin de realizar el estudio del impacto del sesgo asociado a la estimación de la prevalencia de síndrome de quilomicronemia familiar asumiendo tres diferentes escenarios:

1. Estimación de la prevalencia con los casos detectados sobre la población de estudio mediante gold estándar, ignorando el cribado (muestreo en fases).

2. Estimación de la prevalencia unicamente bajo el test imperfecto con Sensibilidad y Especificidad conocidas (Score FCS)

3. Estimación de la prevalencia asumiendo la estructura de 2-fases con test imperfecto y Gold estándar unicamente entre positivos.

En los tres escenarios se aplicaron técnicas de estimación puntual y por intervalo descritas en la sección 2.2 bajo enfoque frecuentista y bayesiano. Métodos de estimación aplicados fueron los que en sus estudios metodológicos reportaran una cobertura cercana al nivel nominal, además de ser altamente implementables en la práctica, los cuales fueron: IC de Agresti-Coull (7), Intervalo bayesiano de Jeffrey (5), Intervalo Agresti-Coull modificado con el estimador de Rogan-Gladen propuesto por Lang et al. (21), IC bayesiano con corrección por características del Test (26) y los intervalos frecuentista y bayesiano propuestos por Thomas et al. (30). Para un enfoque más claro en el análisis metodológico y dado que el estudio siguió una estrategia de diagnóstico y búsqueda de casos en fases, el método propuesto por Thomas et al. fue considerado el punto de referencia para definir la mejor estimación de la prevalencia en este estudio.

Las estimaciones fueron realizadas utilizando el software R versión 4.4.3 (40) con las librerías `binom` (41) y funciones programadas autoría propia a partir de R `base`. Para el caso de la inferencia Bayesiana se utilizo `Stan` (42) y la interface para R denominada `RStan` (43).

## 2.4 Resultados

Se identificaron 274466 mediciones de triglicéridos (TG) en 79570 pacientes. En cuanto a los resultados en la primera fase, 18 pacientes obtuvieron una puntuación $\geq 8$, siendo clasificados como positivos. La totalidad de estos se sometieron a pruebas moleculares, identificándose 6 pacientes con variantes únicas en el gen APOA5 (c.694 T > C; p. Ser232Pro) o en el gen GPIHBP1 (c.523G > C DNA, p. Gly175Arg).

Tabla 2.4: Estimaciones de prevalencia de quilomicronemia familiar según distintos métodos.

| **Método** | **Tipo** | **Estimación** ($\times 10^{-4}$) | $I_{95\%}$ | $S_{95\%}$ |
|---|---|---|---|---|
| Agresti-Coull | Frecuentista | 0.81 | 0.32 | 1.81 |
| Jeffrey | Bayesiano | 0.88 | 0.28 | 1.56 |
| Lang | Frecuentista | 0.00 | 0.00 | 1.62 |
| Flor | Bayesiano | 0.16 | 0.004 | 0.58 |
| $\text{Thomas}_{Fr}$ | Frecuentista | 0.94 | 0.20 | 1.69 |
| $\text{Thomas}_{Bay}$ | Bayesiano | 1.04 | 0.45 | 2.01 |

Frecuentista: método de inferencia clásico. Bayesiano: método basado en inferencia Bayesiana. La estimación y los límites corresponden a la prevalencia por cada 10.000 individuos. Métodos basados en gold estándar (Agresti-Coull, Jeffrey), corrección por test imperfecto (Lang, Flor) y ajuste de muestreo en dos fases ($\text{Thomas}_{Fr}$, $\text{Thomas}_{Bay}$). [$I_{95\%}$,$S_{95\%}$]: Intervalos de Confianza/Credibilidad al 95%

En la Tabla 2.4 se presentan las estimaciones puntuales de prevalencia de quilomicronemia familiar obtenidas mediante distintos métodos de inferencia, junto con sus respectivos intervalos de incertidumbre. Las estimaciones basadas en la aplicación de un *gold estándar* a todos los individuos (métodos Agresti-Coull y Jeffrey) mostraron valores similares, de $0.81 \times 10^{-4}$ y $0.88 \times 10^{-4}$, respectivamente.

En contraste, los métodos que incorporaron la corrección por error de clasificación asociado a un test imperfecto (Lang y Flor) mostraron resultados divergentes. El método de Lang estimó una prevalencia puntual de $0.00 \times 10^{-4}$ con un límite superior de hasta $1.62 \times 10^{-4}$, mientras que el método de Flor, basado en inferencia Bayesiana, presentó una estimación distinta de $0.16 \times 10^{-4}$ y un intervalo de incertidumbre considerablemente más estrecho. Por su parte, los métodos que tuvieron en cuenta la estructura de muestreo en dos fases (Thomas-frecuentista y Thomas-bayesiano) evidenciaron estimaciones de $0.94 \times 10^{-4}$ y $1.04 \times 10^{-4}$, respectivamente.

La Figura 2.3 muestra las estimaciones de prevalencia junto con sus respectivos intervalos de confianza o credibilidad, según el método de inferencia aplicado. Se observa que los métodos frecuentistas basados en la aplicación de un *gold estándar* (Agresti-Coull y Jeffrey) presentan amplitudes de intervalo moderadas, siendo ligeramente más estrecho en el caso de Jeffrey.

Los métodos que incorporan la corrección por test imperfecto evidencian diferencias notables: mientras que Lang presenta un intervalo de máxima amplitud, abarcando desde 0 hasta $1.62 \times 10^{-4}$, el método de Flor presenta un intervalo considerablemente

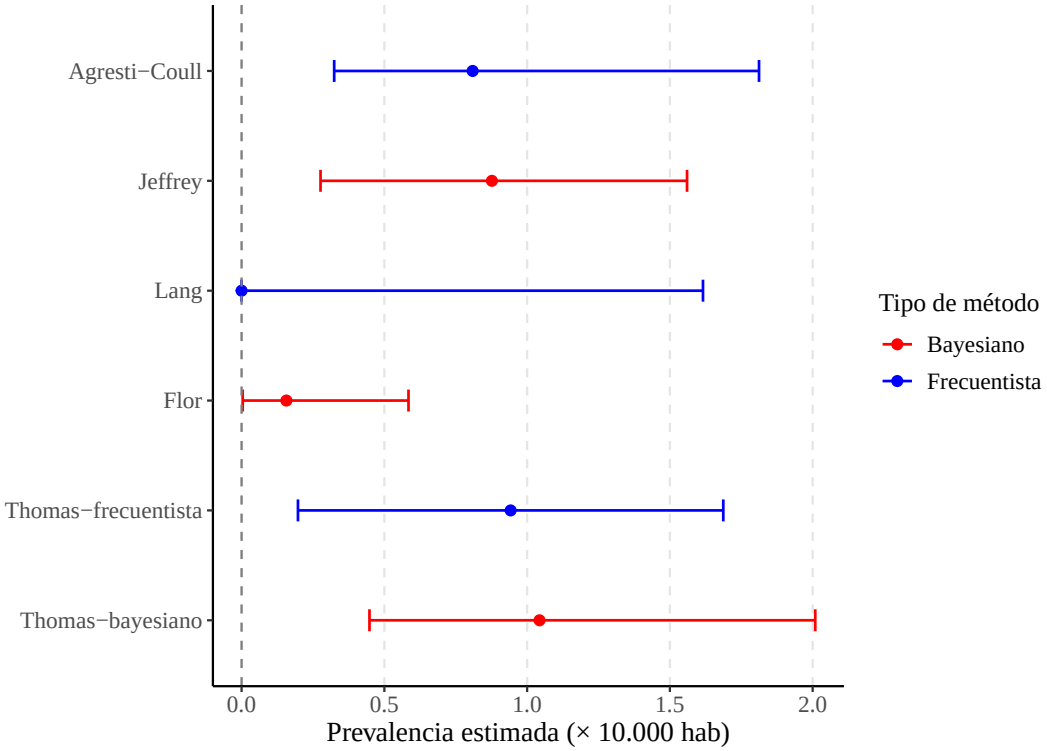


Figura 2.3: Estimaciones de la Prevalencia de Quilomicronemia Familiar por diferentes métodos.

*Intervalos de confianza (frecuentistas, azul) o credibilidad (Bayesianos, rojo)*

más estrecho, lo que sugiere una mejor precisión bajo un enfoque Bayesiano.

## 2.5 Discusión

En escenarios donde se asume un esquema de muestreo distinto al utilizado en la práctica, las estimaciones de prevalencia pueden parecer razonablemente consistentes, especialmente cuando se emplean métodos estadísticos con buen desempeño para proporciones bajas (5). Tal como se mostró en la sección de métodos del presente capítulo 2, tanto el intervalo de Agresti-Coull como el intervalo basado en la prior de Jeffreys presentan un comportamiento óptimo en este tipo de situaciones.

Sin embargo, al no integrar explícitamente la fase de clasificación diagnóstica imperfecta, es decir, el uso de un test imperfecto en lugar de un gold estándar, pueden surgir diversos problemas que afectan la validez de las estimaciones realizadas.

Primero, es posible que se produzcan sesgos sistemáticos en la estimación de la

prevalencia. Si no se ajusta por la sensibilidad y especificidad del test, se incurre en el riesgo de sobreestimar o subestimar la prevalencia real. Por ejemplo, si los individuos seleccionados para una segunda fase de verificación diagnóstica presentan una mayor probabilidad de estar enfermos que la población general, la prevalencia estimada se verá sobrestimada. Por el contrario, si la sensibilidad de la prueba es baja y muchos sujetos verdaderamente enfermos no son identificados inicialmente, se producirá una subestimación de la prevalencia.

En segundo lugar, existe un problema de subestimación de la incertidumbre. Al ignorar tanto el error de clasificación inherente al test imperfecto como la variabilidad introducida por el proceso de selección a segunda fase, se generan intervalos de confianza excesivamente estrechos. Esto puede llevar a una falsa sensación de precisión en las estimaciones puntuales.

Frente a estos desafíos, el enfoque Bayesiano muestra una superioridad. En particular, al corregir por sensibilidad y especificidad mediante inferencia Bayesiana, se obtiene una ventaja importante: las estimaciones de prevalencia siempre se restringen naturalmente al intervalo $[0, 1]$. Esto se debe a que la distribución a priori impuesta sobre la prevalencia garantiza su coherencia matemática, evitando valores negativos o mayores que uno, situación que puede ocurrir en métodos frecuentistas bajo ciertas condiciones.

Un ejemplo notable es el intervalo propuesto por Lang (21), el cual integra el estimador de Rogan-Gladen al intervalo de Agresti-Coull. Si bien este método representa una mejora, debe cumplir con las condiciones de validez descritas en la Sección 2.2.2, en el caso de la aplicación empírica realizada en este estudio, una de esas condiciones no se cumplió ($1 - \hat{Sp} \geq \widehat{AP}$), siendo necesario truncar la estimación puntual y el límite inferior del intervalo a 0, lo cual es metodológicamente subóptimo. Esta situación resalta aún más la ventaja del enfoque Bayesiano en este contexto.

Adicionalmente, como se ha documentado, la corrección propuesta por Rogan y Gladen tiende a subestimar sistemáticamente la prevalencia verdadera, especialmente en contextos de baja prevalencia, situación que también fue evidenciada en los resultados de este análisis, donde sus estimaciones fueron consistentemente más bajas que las obtenidas mediante otros métodos.

En cuanto al diseño de muestreo, el estimador propuesto por Thomas et al. resulta adecuado. Dicho estimador preserva la estructura del diseño en dos fases con verificación parcial, permitiendo estimaciones de prevalencia más coherentes y menos sesgadas que aquellas basadas únicamente en test imperfectos o que suponen la aplicación universal de un gold estándar. Además, su planteamiento es práctico en investigación en salud pública

y clínica, alineándose con esquemas de uso eficiente estadísticamente y financieramente y de consideraciones éticas relevantes, tal como se discute en McNamee et al. (44).

**Limitaciones del enfoque** Pese a las ventajas del enfoque Bayesiano, deben reconocerse algunas limitaciones:

- **Requisitos computacionales:** La implementación práctica de métodos Bayesianos, especialmente cuando se requiere inferencia mediante técnicas de MCMC, demanda habilidades en programación estadística (por ejemplo, en STAN o JAGS) y, en ocasiones, recursos computacionales elevados. Esto puede limitar su adopción masiva entre investigadores en salud pública.
- **Supuestos sobre el diseño de muestreo:** Tanto los métodos frecuentistas como Bayesianos asumen, en su formulación básica, que los datos provienen de una muestra aleatoria simple de la población. En la práctica, esto resulta difícil de garantizar, especialmente en programas masivos de tamizaje donde la selección probabilística es éticamente o logísticamente inviable.

Por estas razones, la necesidad de aplicar correcciones metodológicas adicionales se vuelve crítica para ajustar las estimaciones de prevalencia y aprovechar de manera válida la información recolectada en contextos reales de aplicación. Este tema será desarrollado con mayor profundidad en el capítulo 3 de este trabajo doctoral.

En el presente estudio no fue posible estimar el sesgo de los estimadores de prevalencia, debido a la ausencia de información sobre la verdadera prevalencia poblacional en las bases de datos empleadas. Esta limitación impide la comparación directa entre los valores estimados y el parámetro real, lo cual es necesario para la evaluación formal del sesgo.

En conjunto, los resultados muestran variabilidad entre las estimaciones dependiendo del modelo estadístico y los supuestos incorporados, lo que resalta la importancia de considerar adecuadamente el diseño del estudio y la imperfección de los tests diagnósticos en la estimación de prevalencias bajas.

# Capítulo 3

# Estimación de la Prevalencia Usando Test Imperfecto, Estándar de Referencia y Muestra No Probabilistica: Corrección Post-Estratificación

## 3.1 Introducción

La estimación de la prevalencia en estudios epidemiológicos es un componente fundamental de la planificación en salud pública, particularmente para condiciones patológicas con baja prevalencia en la población. Sin embargo, este proceso enfrenta numerosos desafíos metodológicos, tales como la necesidad de tamaños de muestra grandes y la implementación de diseños complejos para la selección probabilística de la muestra, para aspectos detallados en Ravindra y Fraser (24; 45). Además, el uso de pruebas diagnósticas con imprecisiones en ausencia de un estándar de referencia (gold standard) o la aplicación limitada de dichas pruebas a solo una fracción de la muestra añade complejidad al análisis y puede introducir sesgo. Asimismo, Agresti (7) afirma que es crucial emplear métodos avanzados de estimación por intervalos que superen las limitaciones inherentes a tasas de prevalencia extremadamente bajas, en particular aquellas por debajo de 0.1.

Diversos autores han abordado los desafíos metodológicos asociados con la estimación

de la prevalencia en el contexto de condiciones de baja prevalencia y pruebas diagnósticas imperfectas. Por ejemplo, Rogan y Gladen (6) propusieron un enfoque clásico para la inferencia y corrección de sesgo en la prevalencia estimada cuando se utilizan pruebas con sensibilidad y especificidad inferiores a 1.0. Este enfoque ha sido revisitado y ampliado más recientemente por Izbicki (46). De manera similar, Reiczigel (18) propuso un método exacto para la construcción de intervalos de confianza cuando se emplean pruebas con sensibilidad y especificidad conocidas, una contribución que posteriormente fue desarrollada por Lang et al. (21).

Otro desafío metodológico en la estimación de baja prevalencia es el uso combinado de una prueba diagnóstica falible dentro de un diseño transversal de dos fases, en el cual la verificación mediante un estándar de referencia se realiza únicamente en los casos positivos. Este problema fue abordado por Thomas et al. (30), quienes desarrollaron estimadores frecuentistas y bayesianos para corregir el sesgo introducido por este esquema de verificación.

Sin embargo, si bien las muestras probabilísticas permiten una inferencia robusta a nivel poblacional, su implementación suele implicar costos elevados y limitaciones significativas en entornos con restricciones de recursos. En tales contextos, las muestras de conveniencia y no probabilísticas representan una alternativa práctica. Diversos enfoques metodológicos han sido propuestos para la estimación con muestras no probabilísticas , para más detalles Lohr y Elliot (47; 48). Un ejemplo notable es el uso de estimadores de post-estratificación, como el propuesto por Smith (49), que aprovechan las distribuciones de covariables a nivel poblacional como auxiliares para corregir las estimaciones de prevalencia. Diseñada inicialmente para mitigar el sesgo de no respuesta en el muestreo por encuestas, esta técnica es una herramienta efectiva para ajustar el sesgo en muestras no probabilísticas mediante la estimación de probabilidades de selección.

Este capítulo tiene como objetivo presentar y evaluar algunos de los métodos previamente descritos para estimar prevalencias bajas ($< 0.1$), incorporando los ajustes propuestos por diversos autores. Estos métodos abordan contextos en los que se emplean pruebas diagnósticas falibles y esquemas de verificación parcial aplicados exclusivamente a individuos con resultados positivos. Además, como aportación se integra a estimadores definidos previamente un enfoque de post-estratificación como estrategia para corregir las estimaciones de prevalencia derivadas de muestras no probabilísticas. Los métodos y ajustes son aplicados y analizados en diferentes escenarios utilizando un estudio de prevalencia previamente publicado.

## 3.2 Métodos

Se establecieron tres escenarios utilizando diferentes métodos de estimación por intervalos dentro de los enfoques frecuentista y bayesiano. Para cada enfoque, se compararon estimaciones con y sin ajuste por post-estratificación. Este análisis de sensibilidad permite evaluar el impacto de la post-estratificación en la reducción del sesgo, mostrando que, aunque la eliminación completa del sesgo no está garantizada, el ajuste probablemente mejora la precisión de las estimaciones de prevalencia en muestras no probabilísticas.

### 3.2.1 Estimación de una Proporción en Muestras No Probabilísticas con Ajuste por Post-estratificación

#### Muestreo Probabilistico Vs No probabilistico

Estimar la prevalencia de una condición o enfermedad en la población es uno de los problemas de inferencia más frecuentes en salud pública, debido a la necesidad de calcular con precisión qué proporción de personas sufren dicha enfermedad. Para que estas estimaciones sean válidas, es fundamental que los datos para realizar inferencias provengan de una muestra probabilística y se utilicen técnicas de muestreo adecuadas. Sin embargo, obtener este tipo de muestras puede ser difícil, especialmente cuando no existe un marco muestral completo o la selección equitativa no es posible. En el ámbito clínico y de salud pública, identificar casos prevalentes se realiza a menudo mediante registros administrativos o evaluando poblaciones específicas con ciertos riesgos para detectar la presencia de una enfermedad; sin embargo, esto se lleva a cabo bajo esquemas no probabilísticos que pueden introducir sesgos al seleccionar elementos basándose en criterios distintos a la aleatoriedad.

En el contexto de estimaciones de prevalencia a partir de muestras no probabilísticas es común encontrar registros administrativos, como lo son registros médicos electrónicos, registros individuales de datos de vigilancia de enfermedades de interés en salud pública, entre otros.

En el caso de los registros médicos electrónicos, habría personas con características como, por ejemplo, el tomar servicio en determinado hospital o que, por características como la edad o sexo, siempre tengan una prueba diagnóstica que se considera una prueba de tamiz y que, por ende, otros nunca la tengan.

En las muestras no probabilísticas, ciertas unidades de la población pueden tener una probabilidad cero de ser parte de la muestra. Lo mismo puede ocurrir en muestras

probabilísticas que sufren de no respuesta, donde algunas personas seleccionadas para la muestra nunca acceden a participar a pesar de los esfuerzos realizados. Esto implica que algunos elementos de la población no tienen ninguna posibilidad de estar presentes en la muestra. En encuestas o conjuntos de datos administrativos que involucran personas, esto puede incluir a individuos sin acceso al servicio de salud, o individuos que durante su contacto con servicios de salud nunca, por criterio médico, hayan recibido una prueba tamiz o rápida para identificación de alguna condición patológica. Igualmente, en salud pública, el desarrollo de encuestas basadas en hogares puede excluir zonas inaccesibles por carretera.

El factor de participación en dichos registros a menudo depende de criterios bien definidos que afectan la distribución de una variable de participación de la población en estudio. Llamemos estas variables de participación $R_i$, en la cual $R_i = 1$ cuando el individuo $i$ está en la muestra y $R_i = 0$ en caso contrario. Puede suceder que, en cualquier situación concebible, haya un subconjunto de individuos que siempre van a estar incluidos $P(R_i = 1) = 1$ y otro subconjunto que siempre estará excluido $P(R_i = 1) = 0$.

Para cualquier procedimiento de muestreo donde algunas unidades tienen $P(R_i = 0) = 1$, y para cualquier estimador, siempre es posible concebir una variable de respuesta y para la cual el estimador esté sesgado. La única posibilidad, para las variables de respuesta que interesan, es que las unidades de población con cero posibilidad de participar sean lo suficientemente similares a las unidades en la muestra de modo que el sesgo sea despreciable.

Siguiendo la misma estructura analítica descrita por Lorh (48) para evaluar el grado del efecto del sesgo sobre un estimador bajo una muestra no probabilística, incluso sin conocer la distribución exacta de probabilidad de las $R_i$, los investigadores pueden analizar cómo sería el comportamiento del estimador, como en la proporción en una muestra, bajo ciertas suposiciones acerca de estas probabilidades.

Supongamos que en una población de tamaño $N$ la prevalencia de una enfermedad es $p$ que corresponde a la proporción de casos de la enfermedad en la población ($p = \frac{C}{N}$), donde $C$ corresponde a un total de casos en la población. Sea $C$ el subconjunto de unidades $i$ de $N$ que hace parte de una posible muestra no probabilística $n$ incluida en algún registro administrativo y en la cual fue objeto de aplicación de una prueba diagnóstica gold estándar para definir la presencia de una enfermedad donde $y_i = 1$ con $i$ un individuo que pertenece a la muestra, cuando es positiva y $y_i = 0$ cuando es negativa, con el fin de estimar $p$ se puede utilizar el estimador de razón, asumiendo que

la variable de respuesta de interés es $y_i$ donde:

$$\hat{c} = \sum_{i=1}^{n} = y_i$$

donde $\hat{c}$ es el total de casos en la muestra $n$ de esta manera el estimador de la prevalencia sería:

$$\hat{p} = \frac{\sum_{i=1}^{n} = y_i}{n} = \bar{y} \tag{3.1}$$

Entonces, la variable aleatoria $R_i$ se observa que es 1 para cada unidad en $C$, y se observa que es 0 para todas las unidades de la población que no están en $C$. La proporción muestral se calcula tomando en cuenta sólo las unidades en las que $R_i = 1$, es decir, solo aquellas unidades que forman parte del conjunto $C$. La proporción muestral se basa sólo en las observaciones de las unidades que han sido seleccionadas para incluirse en la muestra $C$, excluyendo todas las demás unidades de la población.

Se puede entonces expresar el estimador muestral en función de la variable de participación $R_i$:

$$\hat{p} = \frac{\sum_{i=1}^{N} R_i y_i}{\sum_{i=1}^{N} R_i}.$$

Denotemos la desviación estándar de la variable de respuesta $y$ y de la variable de participación $R$ y el coeficiente de correlación entre $R$ y $y$.

$$S_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - \bar{y}_N\right)^2} \tag{3.2}$$

$$S_R = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(R_i - \bar{R}_N\right)^2} \tag{3.3}$$

$$\text{Corr}(R, y) = \frac{\sum_{i=1}^{N} \left(R_i - R_N\right)\left(y_i - \bar{y}_N\right)}{(N-1) S_y S_R} \tag{3.4}$$

El error del estimador de razón en la muestra $\bar{y}$ ($\hat{p}$ , cuando $y_i = 1$ o cero en otro caso) al estimar el de la población $p$, cuando $Sy > 0$ y $n < N$, el error del estimador de razón para estimar el parámetro poblacional puede expresarse como:

$$\bar{y} - \bar{Y} = \text{Corr}(R, y) \times \sqrt{\frac{N-1}{n}\left(1 - \frac{n}{N}\right)} \times S_y$$

cada uno de los factores de la expresión anterior tiene un papel importante para el sesgo producido por la muestra no probabilística: El primero de ellos $\text{Corr}(R, y)$, Si existe una correlación positiva entre la realización de una prueba diagnóstica y el indicador de participación en la muestra, entonces las personas con el acceso al registro administrativo (registro clínico electrónico, asistencia a un centro médico) estarán sobrerrepresentadas, y nuestra estimación de la proporción de personas con la enfermedad será demasiado alta. Por otro lado, si la correlación es negativa, las personas que no hacen parte del registro administrativo estarán subrepresentadas y la proporción estimada será demasiado baja.

El segundo factor en la expresión está en función de la fracción muestral que se hace más pequeña cuando la fracción de muestreo n/N es más grande. Cuando la fracción de muestreo es igual a 1, la media de la muestra es idéntica a la media poblacional y no hay error.

El tercer y último factor corresponde a la variabilidad de la variable de interés en la población. Si todos los valores de $y_i$ son iguales (por ejemplo, en el caso anterior, esto ocurriría si todas las unidades en la población tienen acceso al servicio de salud donde se conforma el registro administrativo o si nadie tiene acceso), entonces cualquier muestra dará el mismo valor de la media muestral y no habrá sesgo.

En resumen, en las muestras no probabilísticas, es fundamental tener en cuenta el sesgo que puede surgir debido a la falta de conocimiento de las probabilidades de inclusión. Este sesgo puede ser corregido mediante ajustes de ponderación adecuados que tienen en cuenta las diferencias entre la muestra y la población. Para abordar este sesgo de selección en muestras no probabilísticas, se pueden utilizar métodos de ajuste de ponderación. Uno de ellos es la post-estratificación, en la cual se ajustan los pesos de las unidades de la muestra para que reflejen las proporciones conocidas en la población; este método será explicado en la siguiente sección.

### Post-estratificación

La post-estratificación es una técnica de ajuste por ponderación utilizada en muestras no probabilísticas. Fue inicialmente propuesta para corregir diversos sesgos en muestras probabilísticas, tales como la subcobertura, la sobrecobertura y el sesgo de no respuesta. Esta técnica implica ajustar los pesos de las unidades muestrales para reflejar las proporciones conocidas de una población objetivo dividida en estratos. Sin embargo, estos

estratos se construyen después de que la muestra ha sido seleccionada. En el caso ideal, utilizando una prueba de referencia (gold estándar), la estimación de prevalencia ajustada por post-estratificación, como lo propusieron Holt et al. (50), en notación estándar, asume que la población está compuesta por $N$ unidades, que pueden dividirse de manera única en $H$ estratos de tamaños $N_1, N_2, \ldots, N_H$, de modo que $\sum_{h=1}^{H} N_h = N$.

Sea $y_i$ una variable que toma valores $y_{hi}$, donde $h = 1, \ldots, H$ y $i = 1, \ldots, N_h$. La muestra seleccionada tiene un tamaño fijo $n$, que, tras la selección, se distribuye entre los estratos de acuerdo con el vector $n = (n_1, n_2, \ldots, n_H)$, donde $\sum_{h=1}^{H} n_h = n$. Los componentes de $n$ son desconocidos hasta que la muestra es extraída. El estimador de prevalencia para la población en cada estrato $h$ se define como:

$$\hat{p}_h = \frac{\sum_{i=1}^{n_h} y_{ih}}{n_h} \tag{3.5}$$

donde $y_i$ es una variable indicadora tal que $y_i = 1$ si el individuo en la muestra es positivo y $y_i = 0$ en caso contrario. Además, $x_h = \sum_{i=1}^{n_h} y_{ih}$ representa el número total de casos identificados en el estrato $h$.

Así, el estimador de prevalencia poblacional ajustado por post-estratificación, $\hat{p}_w$, y su varianza, $\hat{V}(\hat{p}_w)$, se expresan de la siguiente manera:

$$\hat{p}_w = \sum_{h=1}^{H} \frac{N_h}{N} \hat{p}_h \tag{3.6}$$

La varianza del estimador de prevalencia ponderada $\hat{p}_w$ se deriva utilizando la fórmula de la varianza de una suma ponderada de variables aleatorias independientes, incorporando el factor de corrección por población finita para cada estrato. Específicamente, la expresión considera la varianza de $\hat{p}_h$ en cada estrato y la proporción de la población total en dicho estrato:

$$\hat{V}(\hat{p}_w) = \sum_{h=1}^{H} \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1} \tag{3.7}$$

Aquí, $\hat{V}(\hat{p}_h) = \frac{\hat{p}_h(1-\hat{p}_h)}{n_h-1}$ representa la varianza estimada para cada estrato, con el factor de corrección por población finita $(1 - \frac{n_h}{N_h})$ aplicado.

### 3.2.2 Escenario 1: Estimación de Prevalencia con Verificación Completa Usando un Estándar de Referencia

**Enfoque Frecuentista**

Sea $p$ la prevalencia poblacional desconocida que se desea estimar a partir de una muestra probabilística de tamaño $n$. Siguiendo el intervalo propuesto por Agresti y Coull (7), empleamos un ajuste —conocido como el método de "agregar dos éxitos y dos fracasos"— que modifica tanto el tamaño de la muestra como el número de éxitos observados. Específicamente, definimos el tamaño de muestra ajustado de la siguiente manera:

$$\tilde{n} \equiv n + z_{\alpha}^2$$

donde $z_{\alpha/2}$ denota el valor crítico de la distribución normal estándar correspondiente a un nivel de significancia $\alpha$. De manera similar, el estimador de prevalencia ajustado se define como:

$$\tilde{p} = \frac{x + \frac{z_{\alpha/2}^2}{2}}{\tilde{n}} \tag{3.8}$$

donde $x$ representa el número de éxitos observados. Con base en estas definiciones, un intervalo para la prevalencia poblacional $p$ se expresa de la siguiente manera:

$$p \approx \tilde{p} \pm z_{\alpha}\sqrt{\frac{\tilde{p}}{\tilde{n}}(1-\tilde{p})} \tag{3.9}$$

- Ajuste por Post-estratificación

  Sea $\tilde{p}$ la prevalencia ajustada, donde $w_h$ representa el peso del estrato $h$ en la población y $\hat{p}_h$ es la prevalencia estimada dentro del estrato $h$. El ajuste para el tamaño de muestra por estrato, $\tilde{n}_h$, se define como:

  $$\tilde{n}_h = n_h + z_{\alpha/2}^2$$

  El estimador $\tilde{p}$ también debe ajustarse utilizando los pesos y las prevalencias de cada estrato, y se expresa como sigue:

  $$\tilde{p}_h = \frac{x_h + \frac{z_{\alpha}^2}{2}}{\tilde{n}_h}$$

Para la prevalencia ponderada, la fórmula es la siguiente:

$$\tilde{p} = \sum_{h=1}^{H} w_h \tilde{p}_h \tag{3.10}$$

La varianza ponderada se define como:

$$\text{Var}(\tilde{p}) = \sum_{h=1}^{H} w_h^2 \frac{\tilde{p}_h(1-\tilde{p}_h)}{\tilde{n}_h} \tag{3.11}$$

El intervalo de confianza se expresa como:

$$p \approx \tilde{p} \pm z_\alpha \sqrt{\text{Var}(\tilde{p})} \tag{3.12}$$

**Enfoque Bayesiano**

Desde una perspectiva Bayesiana, un intervalo para el parámetro $p$ puede derivarse utilizando un modelo beta-binomial, el cual es reconocido por sus favorables propiedades frecuentistas (51). Este enfoque también permite incorporar ajustes por post-estratificación para corregir sesgos en el muestreo.

Para la prevalencia en cada estrato, utilizamos una distribución beta como prior:

$$p_h \sim \text{Beta}(\alpha_h, \beta_h)$$

En este caso específico, usamos una priori no informativa $p_h \sim \text{Beta}(1,1)$. La verosimilitud de observar $x_h$ casos positivos en una muestra de tamaño $n_h$ en el estrato $h$ sigue una distribución binomial; es decir, $x_h \sim \text{Binomial}(n_h, p_h)$. Por lo tanto,

$$\mathbf{p}(x_h|n_h, p_h) = \binom{n_h}{x_h} p_h^{x_h}(1-p_h)^{n_h - x_h}$$

Aplicando el teorema de Bayes, la distribución posterior de $p_h$ se obtiene combinando la distribución a priori y la verosimilitud:

$$\mathbf{p}(p_h|x_h, n_h) \propto \mathbf{p}(x_h|n_h, p_h) \cdot \mathbf{p}(p_h)$$

Sustituyendo la verosimilitud binomial y la distribución a priori $\text{Beta}(\alpha_h, \beta_h)$ en esta

expresión, obtenemos:

$$\mathbf{p}(p_h|x_h, n_h) \propto p_h^{x_h}(1-p_h)^{n_h-x_h} \cdot p_h^{\alpha_h-1}(1-p_h)^{\beta_h-1}. \tag{3.13}$$

Cuando $\alpha_h = 1$ y $\beta_h = 1$, la distribución posterior se simplifica a:

$$\mathbf{p}(p_h|x_h, n_h) \propto p_h^{x_h}(1-p_h)^{n_h-x_h} \tag{3.14}$$

La prevalencia ajustada, $(\tilde{p})$, es una combinación ponderada de las prevalencias de los subgrupos, donde los pesos corresponden a la proporción de cada estrato en la población total. Durante el proceso de inferencia bayesiana, se extraen muestras de la distribución posterior de $p_h$ para cada estrato $h$, lo cual denotamos como $p_h^s$, donde $s$ indica una muestra específica de la distribución posterior.

$$\tilde{p}^s = \sum_{h=1}^{H} w_h p_h^s \tag{3.15}$$

Con base en lo anterior, un intervalo creíble del 95% para la prevalencia ajustada $\tilde{p}$ se define como:

$$\text{ICr}_{95\%}(\tilde{p}) = [\tilde{p}_{2.5}, \tilde{p}_{97.5}] \tag{3.16}$$

donde $\tilde{p}_{2.5}$ y $\tilde{p}_{97.5}$ representan los percentiles 2.5 y 97.5, respectivamente, de la distribución posterior de $\tilde{p}$.

### 3.2.3 Escenario 2: Estimación con una Prueba Diagnóstica Única con Sensibilidad y Especificidad Conocidas

En la mayoría de las situaciones prácticas, el uso de una prueba de referencia (gold standard) no es factible, ya sea por la complejidad de su aplicación o por los altos costos asociados. Esta limitación se agrava aún más cuando, debido a los esquemas de muestreo, ya sean simples o complejos, la prueba de referencia no está disponible para todos los individuos seleccionados. Además, su uso puede estar contraindicado en algunos casos debido a indicaciones específicas.

#### Enfoque Frecuentista

Reiczigel (18) propone la estimación de la prevalencia utilizando un método exacto que corrige el sesgo introducido por una prueba diagnóstica imperfecta, asumiendo que su

sensibilidad y especificidad son conocidas. Posteriormente, Lang et al. (21) ampliaron esta propuesta combinando la fórmula de Rogan y Gladen con un ajuste basado en el método propuesto por Agresti y Coull.

Tanto el estimador puntual corregido con el intervalo de confianza, bajo el enfoque propuesto por Lang, dadas en la ecuación (2.33) y (2.34) del capitulo 2 son:

$$\widehat{AP'} = \frac{n \cdot \widehat{AP} + \frac{z^2_{\alpha/2}}{2}}{\tilde{n}}$$

$$\tilde{p} = \frac{\widehat{AP'} + Sp - 1}{Se + Sp - 1}$$

$$\tilde{p} \pm z_{\alpha/2} \frac{1}{(Se + Sp - 1)} \left( \frac{\widehat{AP'}(1 - \widehat{AP'})}{n'} \right)^{1/2}$$

- Ajuste por Postestratificación

  Aplicando el ajuste propuesto por Agresti y Coull para cada estrato $h$:

$$\widehat{AP'_h} = \frac{x_h + \frac{z^2_{\alpha/2}}{2}}{n_h + z^2_{\alpha/2}} \tag{3.17}$$

$$\tilde{p}_h = \frac{\widehat{AP'_h} + Sp - 1}{Se + Sp - 1} \tag{3.18}$$

  El estimador de la prevalencia ajustada por post-estratificación y la sensibilidad y especificidad conocidas de la prueba, $\tilde{p}$, se define como:

$$\tilde{p} = \sum_{h=1}^{H} w_h \tilde{p}_h \tag{3.19}$$

  La varianza para cada estrato se expresa como:

$$\text{Var}(\tilde{p}_h) = \frac{1}{(Se + Sp - 1)^2} \cdot \frac{\widehat{AP'_h}(1 - \widehat{AP'_h})}{\tilde{n}_h}$$

  La varianza total combinada se calcula de la siguiente manera:

$$\text{Var}(\tilde{p}) = \sum_{h=1}^{H} w_h^2 \cdot \text{Var}(\tilde{p}_h) \tag{3.20}$$

Finalmente, el intervalo de confianza ajustado por post-estratificación se expresa como:

$$\tilde{p} \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(\tilde{p})} \tag{3.21}$$

**Enfoque Bayesiano**

En un marco bayesiano, la descripción en el metodo de inferencia para la prevalencia con corrección por sensibilidad y especificidad fueron revisadas en el sección 2.2.2 en la que se definio la ecuación 2.38 que corresponde a la distrbución posterior de $p$

$$\begin{aligned}\mathbf{Pr}(p|x,n,Se,Sp) \propto (p \cdot Se + (1-p) \cdot (1-Sp))^{x} \cdot \\ (1-(p \cdot Se + (1-p) \cdot (1-Sp)))^{n-x}\end{aligned}$$

y sobre la cual se propone acontinuación un ajuste o corrección por post-estratificación como método aplicable en los escenarios con disponibilidad unicamente de muestras no probabilisticas.

- Ajuste por Post-estratificación

  La estimación bayesiana anterior puede complementarse con un ajuste por post-estratificación. Para $H$ estratos, cada uno con una prevalencia $p_h$, y cada estrato con una proporción $w_h$ de la población total, la prevalencia total ajustada por post-estratificación se puede calcular como sigue:

$$\begin{aligned}\mathbf{p}(p_h|x_h,n_h,Se,Sp) \propto (p_h \cdot Se + (1-p_h) \cdot (1-Sp))^{x_h} \cdot \\ (1-(p_h \cdot Se + (1-p_h) \cdot (1-Sp)))^{n_h-x_h}\end{aligned} \tag{3.22}$$

$$\tilde{p} = \sum_{h=1}^{H} w_h \cdot p_h \tag{3.23}$$

  Sin embargo, al estimar la prevalencia $p_h$ para cada estrato $h$ bajo un enfoque Bayesiano, no se obtiene una única estimación puntual, sino una distribución posterior para $p_h$. Por lo tanto, esta post-estratificación debe basarse en distribuciones posteriores.

  Supongamos que hemos obtenido $S$ muestras de la distribución posterior de $p_h$ para $H$ subgrupos. Denotamos estas muestras para cada subgrupo $h$ de la siguiente manera:

$$p_h^{(1)}, p_h^{(2)}, \ldots, p_h^{(S)}$$

Para cada iteración $s$ en la distribución posterior, ponderamos la prevalencia $p_h^{(s)}$ por la proporción $w_h$ del estrato $h$ en la población total. De esta manera, la prevalencia ajustada en la iteración $s$ es:

$$\tilde{p}^{(s)} = \sum_{h=1}^{H} w_h \cdot p_h^{(s)}$$

Este proceso se repite para cada muestra $s = 1, \ldots, S$, generando una distribución posterior combinada para la prevalencia total ajustada por post-estratificación:

$$\tilde{p}^{(1)}, \tilde{p}^{(2)}, \ldots, \tilde{p}^{(S)}$$

A partir de las muestras $\tilde{p}^{(s)}$, es posible calcular cualquier intervalo creíble. por ejemplo, los percentiles 2.5 y 97.5 de la distribución posterior de $\tilde{p}$ pueden utilizarse para definir un intervalo creíble del 95%. De manera similar, la estimación puntual se obtiene como la mediana de la distribución posterior de $\tilde{p}$.

### 3.2.4 Escenario 3: Estimación cuando el Estado de la Enfermedad Se Verifica Solo en los Casos Positivos en la Prueba

#### Enfoque Frecuentista

En los procesos de diagnóstico o encuestas de salud pública, es común que la verificación del estado real de la enfermedad se realice solo en los individuos que obtienen un resultado positivo en una prueba de tamizaje inicial. Por razones de costo o ética, esta verificación suele restringirse a aquellos con una alta probabilidad de enfermedad basada en los resultados del tamizaje. Como se describió en la sección 2.2.3 Thomas et al. (30) derivaron estimadores frecuentistas para este caso, con base en las derivaciones realizadas para el estimador puntual y su varianza (Ecuaciones 2.52 2.54)

En el desarrollo teórico presentado hasta el momento, se ha asumido que la muestra de tamaño $n$ utilizada para la estimación de la prevalencia de la enfermedad es una muestra aleatoria simple de la población objetivo. Esta suposición es fundamental para garantizar que las inferencias realizadas a partir de los datos observados sean válidas y representativas del conjunto poblacional.

Sin embargo, en aplicaciones prácticas, como en programas de tamizaje o cribado poblacional, esta condición rara vez se cumple. En dichos programas, la muestra de individuos evaluados generalmente no es aleatoria, sino que está determinada por factores como la autoselección de los participantes, criterios logísticos, accesibilidad a los servicios de salud o sesgos de derivación. Estas fuentes de sesgo pueden inducir una composición muestral que difiere sistemáticamente de la estructura real de la población.

Como resultado, las estimaciones directas de la prevalencia basadas en la muestra recolectada pueden ser sesgadas y no reflejar adecuadamente la verdadera prevalencia poblacional. Este problema compromete la validez externa de los resultados y puede conducir a interpretaciones erróneas si no se aplican técnicas de corrección adecuadas.

En este contexto, la **post-estratificación** surge como una estrategia metodológica pertinente para ajustar las estimaciones de prevalencia y mitigar el sesgo introducido por el muestreo no aleatorio. A continuación se propone una corrección a estos estimadores.

- Ajuste por Post-estratificación

  La prevalencia para cada estrato se calcula utilizando el método propuesto por (30). Se define un valor $\hat{p}_h$ para cada uno de los $H$ estratos en la muestra:

$$\hat{p_h} = \frac{-B_h - \sqrt{B_h^2 - 4A_hC_h}}{2A_h} \tag{3.24}$$

  donde $A_h$, $B_h$ y $C_h$ se han definido previamente en (2.52). La varianza para cada estrato se expresa como:

$$\hat{\sigma}_h^2 = \left[\frac{Se}{\hat{p_h}} + \frac{1-Sp}{1-\hat{p_h}} + \frac{(1-Se-Sp)^2}{(1-Se)\hat{p_h} + Sp(1-\hat{p_h})}\right]^{-1} \tag{3.25}$$

  La prevalencia total ajustada por post-estratificación se calcula como:

$$\tilde{p} = \sum_{h=1}^{H} w_h \cdot \hat{p}_h \tag{3.26}$$

  La varianza ponderada se expresa como:

$$Var(\tilde{p}) = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 \hat{\sigma}_h^2 \tag{3.27}$$

Un intervalo de confianza para $\tilde{p}$ se define como:

$$CI(\tilde{p}) = \tilde{p} \pm z_{1-\alpha/2}\sqrt{Var(\tilde{p})} \tag{3.28}$$

**Enfoque Bayesiano**

En el mismo trabajo publicado por Thomas et al. (30), se presenta la versión bayesiana (descrita previamente en la sección 2.2.3) para la estimación de la prevalencia en casos donde la verificación mediante un estándar de referencia solo se realiza en individuos que resultan positivos en una prueba de tamizaje. Tomando como punto de partida la distribución posterior en la ecuación 2.57 se propone a continuación una corrección por post-estratificación sobre la estimación de la distribución posterior de $p$.

- Ajuste por Post-estratificación

  La estimación bayesiana anterior también puede incluir un ajuste por post-estratificación. En este caso, como se definió previamente, para los $H$ estratos formados, $p_h$ se estima estratificando la distribución posterior de $p$:

$$\begin{aligned}\mathbf{p}\left(p_h \mid \mathbf{D}\right) = &\left[(1-Se)p_h + Sp(1-p_h)\right]^{n_{\cdot 0_h}} \\ &\times \left(Se \cdot p_h\right)^{n_{11_h}} \\ &\times \left[(1-Sp)(1-p_h)\right]^{n_{01_h}} \cdot \text{Beta}(1,1)\end{aligned}$$

$$\tilde{p} = \sum_{h=1}^{H} w_h \cdot p_h \tag{3.29}$$

  Como se definió anteriormente, al estimar la prevalencia $p_h$ para cada estrato $h$ bajo un enfoque Bayesiano, no se obtiene una única estimación puntual, sino que se genera una distribución posterior para $p_h$. En este contexto, la post-estratificación debe basarse en estas distribuciones posteriores, permitiendo la incorporación de la incertidumbre en las estimaciones ajustadas por estrato.

  Hemos obtenido $S$ muestras de la distribución posterior de $p_h$ para $H$ estratos. Estas muestras para cada estrato $h$ se denotan como:

$$p_h^{(1)}, p_h^{(2)}, \ldots, p_h^{(S)}$$

  Para cada iteración $s$ de la distribución posterior, ponderamos la prevalencia $p_h^{(s)}$ por la proporción $w_h$ del estrato $h$ en la población total. De esta manera, la

prevalencia ajustada en la iteración $s$ se expresa como:

$$\tilde{p}^{(s)} = \sum_{h=1}^{H} w_h \cdot p_h^{(s)} \tag{3.30}$$

Este proceso se repite para cada muestra $s = 1, \ldots, S$, generando una distribución posterior combinada para la prevalencia total ajustada por post-estratificación:

$$\tilde{p}^{(1)}, \tilde{p}^{(2)}, \ldots, \tilde{p}^{(S)}$$

A partir de las muestras $\tilde{p}^{(s)}$, se puede calcular cualquier intervalo creíble.

### 3.2.5 Aplicación de la Estimación de Prevalencia para la Mutación en la Quilomicronemia Familiar con Ajuste por Post-estratificación

Se utilizaron datos del estudio descriptivo transversal de dos fases, en el cual se desarrolló para el capítulo 2, para evaluar la aplicabilidad del ajuste por post-estratificación propuesto.

Los estratos utilizados para el ajuste por post-estratificación se basaron en el sexo y la edad en años individuales (rango de 18 a 85 años o más), obtenidos a partir de las proyecciones oficiales de población publicadas por el Departamento Administrativo Nacional de Estadística (DANE) de Colombia.

Las estimaciones frecuentistas se realizaron utilizando el *software R, versión 4.2.3*, mientras que las estimaciones Bayesianas se llevaron a cabo en Stan a través de la biblioteca RStan. En la simulación MCMC, la convergencia hacia una distribución estacionaria se verificó ejecutando cuatro cadenas con 1000 iteraciones cada una, asegurando que $\hat{R} < 1.1$ como criterio de convergencia. Este enfoque permite obtener la distribución posterior completa, facilitando así la selección de intervalos creíbles del 95% (52). El código de $R$ para los resultados presentados se puede encontrar en el apéndice B

## 3.3 Resultados

La Tabla 3.1 presenta los resultados obtenidos para la estimación con y sin ajuste por post-estratificación, correspondientes a cada escenario. La Tabla 3.2 muestra las longi-

tudes de cada intervalo con y sin ajuste por post-estratificación. Según cada escenario, se encontraron los siguientes resultados:

Tabla 3.1: Estimaciones puntuales y por intervalo de la prevalencia poblacional ($p$) obtenidas mediante los métodos frecuentista y Bayesiano, tanto sin ajuste como con ajuste por post-estratificación.

| **Escenario** | **Método** | **Sin ajuste (%)** | | | **Con ajuste (%)** | | |
|---|---|---|---|---|---|---|---|
| | | ***p*** | **Inferior** | **Superior** | ***p*** | **Inferior** | **Superior** |
| Escenario 1 | Agresti–Coull | 0.0081 | 0.0032 | 0.0181 | 0.4621 | 0.4015 | 0.5226 |
| | Bayesiano (beta-binomial) | 0.0088 | 0.0028 | 0.0156 | 0.2417 | 0.2014 | 0.2886 |
| Escenario 2 | Lang | 0.0000 | 0.0000 | 0.0162 | 0.0000 | 0.0000 | 0.0836 |
| | Bayesiano–Flor | 0.0016 | 0.0000 | 0.0058 | 0.2740 | 0.2280 | 0.3286 |
| Escenario 3 | Frecuentista–Thomas | 0.0094 | 0.0020 | 0.0169 | 0.0075 | 0.0014 | 0.0135 |
| | Bayesiano–Thomas | 0.0104 | 0.0045 | 0.0201 | 0.2797 | 0.2328 | 0.3352 |

Tabla 3.2: Comparación de las longitudes de los intervalos según el método de estimación.

| **Método** | **Sin Ajuste** | **Con Ajuste** |
|---|---|---|
| Agresti–Coull | 0.0149 | 0.1212 |
| Bayesiano (beta-binomial) | 0.0128 | 0.0872 |
| Lang | 0.0162 | 0.0836 |
| Flor–Bayesiano | 0.0058 | 0.1005 |
| Frecuentista–Thomas | 0.0149 | 0.0121 |
| Bayesiano–Thomas | 0.0156 | 0.1025 |

### 3.3.1 Escenario 1: Estimación con Verificación Completa Usando un Estándar de Referencia

En el primer escenario, el método de Agresti–Coull mostró una estimación puntual de la prevalencia de 0.0081% sin ajuste por post-estratificación, con un intervalo de confianza de $[0.0032\%, 0.0181\%]$. Después de aplicar el ajuste por post-estratificación, la estimación aumentó a $p = 0.4621\%$. Este resultado evidencia un aumento significativo en la estimación de la prevalencia, probablemente debido a la corrección de sesgos en la muestra.

La longitud del intervalo de confianza sin post-estratificación fue de 0.0149, mientras que aumentó a 0.1211 después del ajuste, reflejando un incremento notable. Este resultado concuerda con el objetivo de la post-estratificación: considerar la variabilidad previamente ignorada debido al uso de una muestra no probabilística.

De manera similar, el enfoque bayesiano inicialmente presentó una prevalencia de $p = 0.0088\%$, con un intervalo creíble del 95% ($CIr_{95\%}$) de $[0.0028\%, 0.0156\%]$. Con el ajuste, la estimación aumentó a $p = 0.2417\%$, con un intervalo ajustado de $[0.2014\%, 0.2886\%]$. La longitud del intervalo aumentó de 0.0128 sin post-estratificación a 0.0872 con ajuste, reflejando la misma tendencia observada con el método frecuentista.

### 3.3.2 Escenario 2: Estimación con una Prueba Diagnóstica de Sensibilidad y Especificidad Conocidas

En este escenario, el método de Lang presentó una prevalencia inicial de $p = 0.000\%$ sin ajuste por post-estratificación, lo que indica una subestimación severa debido a la sensibilidad limitada del método. Después de aplicar la post-estratificación, el estimador aumentó a 0.0836%, resaltando la importancia del ajuste para corregir el sesgo en diseños con pruebas diagnósticas imperfectas. La longitud del intervalo sin post-estratificación fue de 0.0162, mientras que aumentó a 0.0836 con post-estratificación.

El método Bayesiano propuesto por Flor mostró una prevalencia inicial de $p = 0.0016\%$, con un intervalo creíble del 95% ($CIr_{95\%}$) de $[0.0000\%, 0.0058\%]$. Tras el ajuste, el estimador aumentó significativamente, como se muestra en la Tabla 3.1; este resultado enfatiza la flexibilidad del enfoque Bayesiano para abordar las limitaciones de pruebas diagnósticas falibles y ajustar las estimaciones de prevalencia. En este caso, la longitud del intervalo aumentó de 0.0058 a 0.1006. Este incremento significativo se debe a que el método Bayesiano incorpora explícitamente la incertidumbre adicional asociada con la muestra no probabilística.

### 3.3.3 Escenario 3: Estimación Cuando el Estado de la Enfermedad Se Verifica Solo en Casos Positivos

En este escenario final, el método frecuentista de Thomas estimó una prevalencia inicial de $p = 0.0094\%$, con un intervalo de confianza del 95% ($CI_{95\%}$) de $[0.0020\%, 0.0169\%]$. Después de aplicar la corrección por el inverso de la inclusión en la muestra, el estimador disminuyó, resultando en un intervalo más estrecho. La longitud del intervalo fue de

0.0149 sin post-estratificación y se redujo ligeramente a 0.0121 con el ajuste. Este resultado es único en este escenario, ya que la post-estratificación parece reducir la variabilidad al abordar las limitaciones específicas de la verificación parcial. Este cambio refleja la capacidad del ajuste para mejorar la precisión en diseños de verificación parcial, una situación más representativa de la práctica diaria.

El enfoque Bayesiano de Thomas et al. estimó inicialmente una prevalencia de $p$ = 0.0104%, con un intervalo creíble del 95% ($CIr_{95\%}$) de $[0.0045\%, 0.0201\%]$. El estimador ajustado mediante el método de post-estratificación aumentó, demostrando la capacidad de las estimaciones Bayesianas para integrar información adicional y ajustar las estimaciones de prevalencia en escenarios de verificación parcial. La longitud del intervalo creíble aumentó de 0.0156 a 0.1024 con el ajuste. Este incremento considerable es consistente con la capacidad del enfoque Bayesiano para modelar la incertidumbre adicional introducida por la post-estratificación.

## 3.4 Discusión

La estimación precisa de la prevalencia en estudios epidemiológicos es fundamental para la planificación en salud pública, especialmente en enfermedades raras o condiciones de baja prevalencia. Los métodos evaluados en este estudio abordan desafíos clave, como pruebas diagnósticas imperfectas, verificación parcial del estado de la enfermedad y el uso de muestras no probabilísticas. Estas limitaciones pueden generar estimaciones sesgadas y decisiones inadecuadas en salud pública si no se corrigen. Por ello, este estudio destaca la necesidad de implementar ajustes metodológicos robustos para mejorar la precisión en las estimaciones de prevalencia. La comparación entre los enfoques frecuentistas y Bayesianos reveló diferencias significativas en su capacidad para abordar las limitaciones inherentes a los diseños de estudio y las pruebas diagnósticas. Estas observaciones concuerdan con McNamee et al. (27), quienes señalaron que, a pesar de su complejidad, los diseños en dos fases pueden optimizar la precisión y reducir costos en estudios de prevalencia.

### 3.4.1 Primer Escenario

En el primer escenario, los resultados muestran que ambos enfoques metodológicos mitigan de manera indirecta los sesgos en las estimaciones sin ajuste, como lo evidencian las diferencias observadas en las estimaciones puntuales y la longitud de los interva-

los. Sin embargo, el ajuste por post-estratificación tuvo un impacto significativo al incorporar información adicional sobre la distribución de variables poblacionales. Este resultado es consistente con Holt et al. (50), quienes reportaron la efectividad de la post-estratificación para abordar la falta de representatividad en muestras no probabilísticas. Además, el enfoque Bayesiano ajustado por post-estratificación generó intervalos de confianza más amplios que el método de Agresti–Coull, como se observó en el estudio de Flor et al. (26). Este comportamiento refleja la capacidad del enfoque Bayesiano para modelar explícitamente la incertidumbre, lo que lo convierte en una alternativa ventajosa en contextos de alta variabilidad.

### 3.4.2 Segundo Escenario

El segundo escenario, donde se utilizó una prueba diagnóstica con sensibilidad y especificidad conocidas, resalta la importancia de considerar la calidad de las pruebas iniciales en el diseño del estudio. Como lo señalaron Shrout y Newman (53), la sensibilidad limitada de los métodos de tamizaje puede subestimar significativamente la prevalencia, en concordancia con los resultados sin ajuste obtenidos en esta investigación. No obstante, los ajustes por post-estratificación y las estimaciones Bayesianas permitieron corregir estas limitaciones, en línea con estudios previos que destacan la flexibilidad del enfoque Bayesiano para incorporar información previa y mejorar la precisión. En cuanto a la longitud de los intervalos, se observó que los intervalos sin ajuste eran más cortos, lo que podría generar una falsa confianza en estimaciones subrepresentadas. En contraste, el ajuste por post-estratificación, aunque aumentó la longitud de los intervalos, corrigió el sesgo inherente en los datos y mejoró la representatividad poblacional.

Una situación particular, bien documentada en la literatura, se hizo evidente en el intervalo propuesto por Lang y también fue observada en este estudio. Al utilizar el estimador de prevalencia ajustado propuesto por Rogan y Gladen, la estimación ajustada puede quedar fuera del rango permitido $[0, 1]$, especialmente con valores negativos cuando la prevalencia real es muy baja. Esta situación, previamente estudiada por autores como Viana (54) y Hasselt (55), se reflejó en el presente estudio, donde la prevalencia reportada (Tabla 3.1) fue cero, subestimando consistentemente la prevalencia real en escenarios de baja prevalencia.

En contraste, el enfoque bayesiano resultó beneficioso, como lo evidencian los resultados de este estudio. Al utilizar una distribución a priori $p \sim \text{Beta}(\alpha, \beta)$, este enfoque garantiza que la prevalencia estimada permanezca dentro del rango permitido $[0, 1]$, incluso en situaciones de prevalencia extremadamente baja. Esta ventaja también

se reflejó en el enfoque Bayesiano ajustado por post-estratificación en el Escenario 2, donde se obtuvo una estimación más robusta y representativa, abordando eficazmente las limitaciones observadas en los métodos frecuentistas.

### 3.4.3 Tercer Escenario

En el Escenario 3, donde el estado de la enfermedad se verifica solo en los casos positivos, los resultados confirman las ventajas de combinar métodos de máxima verosimilitud y ajustes por post-estratificación. El enfoque frecuentista, aunque tradicional, mostró una reducción en la longitud del intervalo tras el ajuste, lo que resalta su capacidad para abordar sesgos específicos en estos diseños. Por otro lado, el enfoque Bayesiano proporcionó intervalos más amplios pero con una mejor representación de la incertidumbre, en concordancia con las observaciones de Flor et al. (26) en escenarios de verificación parcial.

En cuanto a las longitudes de los intervalos, los intervalos sin ajuste fueron más cortos, lo que podría llevar a una confianza excesiva en estimaciones subrepresentadas. En contraste, el ajuste por post-estratificación, aunque incrementó la longitud de los intervalos, corrigió el sesgo inherente en los datos y mejoró la representatividad poblacional.

En este escenario, se observó un patrón diferente en las longitudes de los intervalos de confianza. El enfoque frecuentista mostró una reducción en la longitud del intervalo tras el ajuste. En contraste, los enfoques Bayesianos presentaron intervalos más amplios después del ajuste, reflejando su capacidad para incorporar la incertidumbre inherente a los diseños de verificación parcial mediante una prueba de referencia. Estos resultados subrayan la importancia de seleccionar el método más adecuado según el contexto del estudio, como lo sugirió McNamee (27) en investigaciones previas.

Finalmente, Bayer et al. (56) propusieron un enfoque innovador para la estimación de prevalencia derivada de encuestas con muestreo complejo, basado en un método de combinación melding method que utiliza distribuciones gamma y beta. Si bien su método es particularmente robusto para abordar el sesgo introducido por la sensibilidad y especificidad imperfectas de las pruebas diagnósticas, su aplicación se limita a muestras probabilísticas, garantizando la representatividad poblacional a partir del diseño muestral, pero sin abordar directamente el sesgo en muestras no probabilísticas ni ajustes por post-estratificación.

En contraste, el enfoque adoptado en este estudio se centra en contextos donde las muestras son no probabilísticas, presentando un desafío adicional. La integración de métodos Bayesianos con ajustes por post-estratificación permite corregir sesgos rela-

cionados con la representatividad poblacional, una dimensión no abordada explícitamente en el método de Bayer. Esta diferencia metodológica hace que este trabajo sea complementario al de Bayer et al., proporcionando una solución flexible para situaciones en las que limitaciones prácticas—como la imposibilidad de utilizar diseños de muestreo probabilísticos o la verificación incompleta de pruebas diagnósticas—requieren métodos que equilibren representatividad y adaptabilidad.

En estudios epidemiológicos, los métodos frecuentistas ofrecen eficiencia computacional e interpretación sencilla en grandes muestras, pero pueden subestimar la incertidumbre en escenarios de baja prevalencia o en diseños con verificación parcial. En contraste, los métodos Bayesianos integran información previa y proporcionan una cuantificación más completa de la incertidumbre mediante intervalos creíbles. La integración de ambos enfoques, junto con ajustes por post-estratificación, permite corregir eficazmente los sesgos relacionados con el muestreo no probabilístico y las pruebas diagnósticas imperfectas. Esta estrategia combinada ofrece un marco sólido para la estimación de prevalencia, mejorando la representatividad y respaldando una toma de decisiones en salud pública más informada.

A pesar de los hallazgos favorables, este estudio presenta algunas limitaciones. La dependencia de métricas conocidas, como la sensibilidad y especificidad de las pruebas diagnósticas, puede restringir la aplicabilidad de los métodos en contextos donde estos parámetros no están bien establecidos. Además, aunque el enfoque Bayesiano es robusto, su precisión depende críticamente de la calidad de la información previa (es decir, sin datos confiables, el uso de distribuciones a priori no informativas puede generar estimaciones menos precisas).

No fue posible generar o adquirir un conjunto de datos sintéticos que represente con precisión las complejidades asociadas con los procesos de muestreo no probabilístico, que constituyen un elemento crucial en el diseño de este estudio. Esta limitación resalta la necesidad de investigaciones futuras para evaluar rigurosamente las propiedades de cobertura de los intervalos ajustados mediante post-estratificación y cuantificar el sesgo asociado a los estimadores puntuales bajo diferentes condiciones. Estas evaluaciones son esenciales para determinar la solidez y validez de las metodologías propuestas, especialmente en comparación con enfoques alternativos.

Los resultados obtenidos tienen implicaciones importantes para la práctica en salud pública y la investigación epidemiológica. Ajustar las estimaciones de prevalencia mediante post-estratificación mejora significativamente la representatividad en contextos con muestras no probabilísticas y pruebas diagnósticas imperfectas. Los métodos Bayesianos

han demostrado ser herramientas valiosas en escenarios de alta incertidumbre, particularmente cuando se incorpora información previa.

# Capítulo 4

# Estimacion de la prevalencia con test imperfecto mediante Regresión Logística

## 4.1 Introducción

La estimación precisa de la prevalencia de enfermedades infecciosas en programas de salud pública es crucial para la planificación, implementación y evaluación de estrategias de intervención. Habitualmente, dicha prevalencia se estima mediante la aplicación masiva de pruebas diagnósticas, cuyo desempeño raramente es perfecto, resultando en errores de clasificación como falsos positivos y falsos negativos, que afectan la precisión de las estimaciones obtenidas, según lo afirmado por Magder et al. (57).

Considerar la posible clasificación errónea en la variable de resultado binaria es fundamental, dado que esta puede generar sesgos no despreciables en las estimaciones de los parámetros si se ignora. Si se sospecha una clasificación errónea y se dispone de información sobre las tasas de clasificación errónea (sensibilidad y especificidad), o se pueden asumir sus valores, se deben utilizar métodos que incorporen explícitamente esta información para corregir la prevalencia estimada y evitar sesgos hacia el valor nulo en las estimaciones obtenidas.

La regresión logística es una herramienta comúnmente utilizada para estimar prevalencias con la incorporación de ajustes por covariables y correcciones de errores de medición en modelos de regresión logística; mejora la estimación de la prevalencia de una condición, permitiendo además la incorporación directa de covariables demográficas y

epidemiológicas relevantes. No obstante, para abordar los desafíos inherentes a la clasificación errónea y aprovechar al mismo tiempo la flexibilidad en la incorporación de covariables de interés, se requieren métodos que integren ambos elementos. Esta combinación permite obtener estimaciones de prevalencia ajustadas con mayor validez, especialmente importantes cuando múltiples enfermedades infecciosas están correlacionadas epidemiológicamente.

Diferentes estrategias han sido desarrolladas para manejar estos errores diagnósticos en la estimación de prevalencias: desde ajustes post-hoc mediante estandarización marginal de probabilidades predichas como la propuesta por Muller et al. (58), que permite un ajuste extrínseco con resultados a partir de las estimaciones realizadas con el modelo. En esta misma línea de trabajo Liu et al (59) plantean un modelo de regresión logística con corrección por clasificación errónea en la variable de resultado binaria. El método se basa en la incorporación de parámetros para las tasas de falsos positivos (FP) y falsos negativos (FN) en el modelo de regresión logística estándar, permitiendo a su vez estimar coeficientes ajustados por dichas tasas de error. Este modelo modificado por Liu et al. plantea varias variantes en función de la necesidad de estimación de no solo uno de los dos errores, sino también la posibilidad de ambos errores, asumiendo que puedan ser iguales o distintos.

Valle y cols. (60) proponen una perspectiva de inferencia Bayesiana sobre la regresión logística para corregir sesgos sobre las estimaciones con datos de prevalencia cuando se usan test imperfectos; sin embargo, sus resultados son orientados con mayor énfasis en el sesgo inducido por la mala clasificación sobre los estimados de asociación en factores de riesgo para la enfermedad y no como método de ajuste de estimaciones de prevalencia en presencia de covariables.

En este capítulo se aborda directamente la problemática relacionada con la estimación de prevalencia de enfermedades con el uso de un test imperfecto, pero también con la necesidad de controlar por covariables, de índole epidemiológica y/o demográfica. Para esto se propuso como objetivo proporcionar una comparación de cuatro metodologías estadísticas en la estimación de la prevalencia bajo un enfoque práctico para estimaciones puntuales ajustadas y corregidas, resaltando la importancia del abordaje integral que incluye tanto la incertidumbre diagnóstica como el ajuste por covariables. Para dar respuesta a este objetivo se planteó una revisión sistemática de literatura que permitió identificar metodologías estadísticas relevantes y previamente validadas en cuanto a su desempeño estadístico en corrección del sesgo, y posteriormente fueron comparadas bajo su aplicación empírica sobre una muestra representativa de 11452 sujetos

evaluados simultáneamente para seroprevalencia de VIH y sífilis.

## 4.2 Revisión sistemática de literatura sobre métodos de Estimación de Prevalencia en presencia de test imperfectos mediante modelos de regresión

La incorporación de ajustes por covariables y la corrección de errores de medición en los modelos de regresión contribuyen significativamente a mejorar la estimación de la prevalencia de una condición. Diversos estudios han destacado la necesidad de emplear métodos que permitan realizar ajustes simultáneos, tanto por variables de confusión que introducen sesgos sistemáticos en la estimación, como por errores de clasificación derivados del uso de pruebas diagnósticas imperfectas, cuya sensibilidad y especificidad son inferiores a 1.0.

Para evaluar la literatura existente se llevó a cabo una búsqueda bibliográfica estructurada en bases de datos con el fin de tener un estado del arte sobre los distintos métodos propuestos para llevar a cabo la estimación de prevalencia con ajuste por covariables y corrección del sesgo por errores de clasificación inducidos por test imperfecto y que posteriormente pudieran ser comparados en términos de alguna medida de su desempeño. De acuerdo a la metodología propuesta se planteó:

### 4.2.1 Búsqueda y Selección de Artículos

Se diseñó bajo la estrategia PICO una pregunta de investigación a ser contestada mediante búsquedas en bases de datos de literatura científica que permitieran dar respuesta, la cual fue estructurada como:

**¿Cómo mejora la estimación de prevalencia de una condición el uso de modelos de regresión logística que incorporan ajuste por covariables y corrigen los errores de medición de pruebas diagnósticas?**, seguido se plantearon búsquedas exhaustivas y estructuradas en bases de datos académicas como PUBMED, SCOPUS, EMBASE y literatura gris mediante Google Académico.

### 4.2.2 Criterios de selección

Para seleccionar los estudios incluidos, se utilizaron los siguientes criterios:

- **Método estadístico:** ¿Utiliza el estudio modelos de regresión logística para estimar la prevalencia de enfermedades o condiciones?
- **Características metodológicas:** ¿Incorpora el estudio ajustes por covariables o métodos de corrección para errores de medición en pruebas diagnósticas (o ambos) en la estimación de la prevalencia?
- **Métricas de desempeño:** ¿Reporta el estudio medidas cuantitativas sobre la precisión de las estimaciones (por ejemplo, sesgo, precisión o cobertura del intervalo de confianza)?
- **Diseño del estudio:** ¿Incluye el estudio aplicaciones empíricas o evaluaciones por simulación de los métodos de estimación?
- **Enfoque del resultado:** ¿La estimación de prevalencia es el resultado principal del estudio?
- **Tipo de estudio:** ¿Es el estudio un artículo de investigación primaria o una revisión sistemática o meta-análisis sobre métodos de estimación de prevalencia?

La evaluación para la inclusión de cada artículo se realizó tomando en cuenta todos estos criterios conjuntamente, haciendo una valoración de cada uno de ellos.

### 4.2.3 Extracción de Datos

El investigador principal y dos revisores externos extrajeron para cada estudio las siguientes características, además de evaluar el cumplimiento de criterios de selección previamente descritos:

- **Tipo de Diseño del Estudio:** Los tipos posibles de estudio incluyeron:
  - Estudio transversal
  - Estudio de simulación
  - Estudio de modelado Bayesiano
  - Estudio de comparación metodológica

Cuando se identificaron múltiples elementos de diseño en los estudios incluidos, se procedió a listar todos los tipos relevantes de manera explícita. En los casos en que el diseño no estaba claramente especificado, se clasificó como "No claramente especificado" y se proporcionó una breve explicación del enfoque metodológico empleado, basada en la descripción disponible en la sección de métodos de cada estudio.

- **Técnica de regresión**

  Se extrajeron las técnicas específicas de modelado estadístico utilizadas en los estudios, a partir de la revisión detallada de las secciones de métodos y resultados. Se documentaron los tipos de modelos de regresión empleados (por ejemplo, modelos logísticos, de Poisson, de Cox o log-binomiales), así como las técnicas particulares aplicadas, incluyendo enfoques de ajuste Bayesiano y el uso de varianza robusta, entre otros. También se especificó el propósito del enfoque de modelado en cada caso (estimación de asociaciones, ajuste por confusión, evaluación de efectos modificadores, etc.). En los estudios que compararon múltiples enfoques, se consignaron todos los modelos considerados. Adicionalmente, se identificaron y registraron características particulares o modificaciones a técnicas estándar de modelado, cuando estas fueron reportadas por los autores.

- **Características de la Muestra:**

  Se extrajo información clave sobre la muestra incluida en cada estudio, incluyendo el tamaño total de la población analizada, el tipo de población (por ejemplo, adultos mayores, pacientes hospitalizados o población general).

- **Resultados de la Estimación de Prevalencia**

  Se identificaron y resumieron los hallazgos principales relacionados con la estimación de la prevalencia reportados en los estudios incluidos. En particular, se compararon las estimaciones obtenidas mediante diferentes enfoques de modelado estadístico, evaluando las discrepancias entre los modelos de regresión utilizados (por ejemplo, logística, Poisson, log-binomial). Se documentó la magnitud del sesgo observado en las estimaciones de de prevalencia, especialmente en presencia de errores de clasificación.

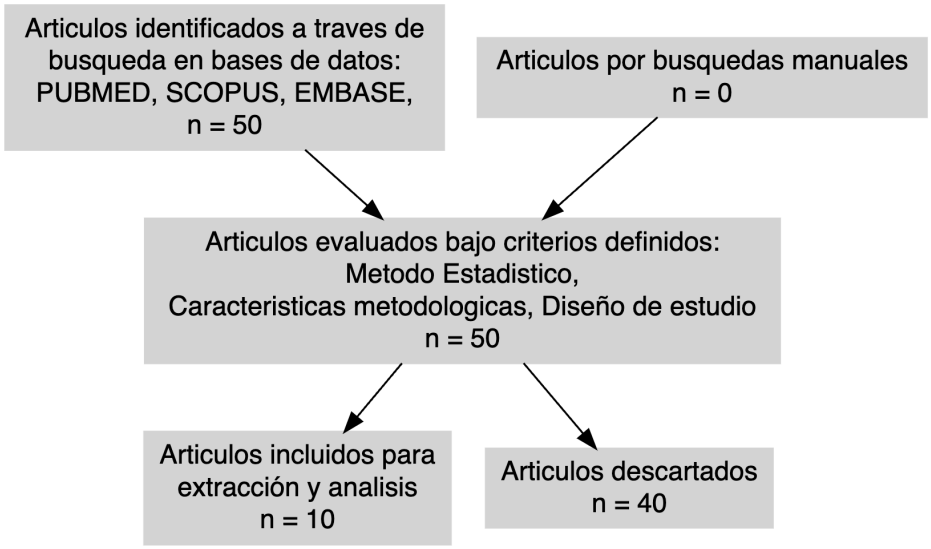


Figura 4.1: Diagrama de fliujo de selección de articulos bajo criterios establecidos

### 4.2.4 Resultados

La figura 4.1 muestra la selección de estudios acorde al cumplimiento de criterios de inclusión; once estudios cumplieron y están relacionados en la tabla 4.1

Tabla 4.1: Características de los Estudios Incluidos

| **Estudio** | **Diseño** | **Método Estadístico** | **Prueba Diagnóstica** | **Resultados** | **Rec.** |
|---|---|---|---|---|---|
| Coutinho (61) | Estudio transversal | Regresión de Cox, log-binomial, Poisson y logística | Diagnóstico de demencia, Cuestionario de Autoinforme (SRQ-20) | Estimación de razón de prevalencia | Sí |
| Liu et al. (59) | Estudio de simulación | Regresión logística modificada | Autoreporte de consumo de marihuana | Estimación de prevalencia | Sí |

Continúa en la siguiente página

**Tabla 4.1 (Continuación)**

| Autor | Diseño | Método | Test | Resultados | Rec. |
|---|---|---|---|---|---|
| Goette et al (62) | Estudio de simulación, Modelado Bayesiano, Comparación metodológica | Regresión logística bayesiana con ajuste de sesgo | Diagnóstico clínico de enfermedad de Alzheimer | Precisión en la predicción del diagnóstico real | No |
| Lewis et al. (63) | Estudio de simulación, Modelado Bayesiano, Comparación metodológica | Modelos de variables latentes, regresión estándar | Prueba serológica ELISA | Estimación de prevalencia considerando errores diagnósticos | Sí |
| Magder et al (57) | Comparación metodológica | Regresión logística con algoritmo EM (Expectation-Maximization) | No mencionado | Estimación no sesgada de razón de momios (odds ratio) | No |
| Qin et al. (64) | Estudio de simulación, Desarrollo metodológico para estudios de casos y controles | Regresión logística con información auxiliar | No mencionado | Mejora en la eficiencia de la regresión logística | No |
| Skove et al.(65) | Estudio de simulación, Comparación metodológica | Regresión log-binomial, ecuaciones de estimación generalizada (GEE), regresión de Cox | No aplica | Estimación de razón de proporción de prevalencia | No |
| Tu et al.(66) | Modelado Bayesiano, Estudio de simulación, Comparación metodológica | Inferencia bayesiana, enfoque de datos faltantes, técnica de variable latente | No mencionado | Estimación de prevalencia con sensibilidad/especificidad variable | No |

Continúa en la siguiente página

**Tabla 4.1 (Continuación)**

| Autor | Diseño | Método | Test | Resultados | Rec. |
|---|---|---|---|---|---|
| Valle et al.(60) | Estudio de simulación, Modelado Bayesiano, Comparación metodológica, Estudio transversal | Modelos Bayesianos, regresión logística estándar | Pruebas diagnósticas rápidas, microscopía, Reacción en Cadena de la Polimerasa (PCR) | Estimación de prevalencia considerando pruebas imperfectas | Sí |
| Vansteelandt et al. (67) | Comparación metodológica | Estimación de máxima verosimilitud, modelos lineales generalizados | Pruebas de suero agrupadas | Estimación de prevalencia con muestras agrupadas | No |

El diseño metodológico de comparación fue el más frecuente entre los estudios incluidos, apareciendo en 8 de los 10 estudios analizados. Le siguieron los estudios de simulación, presentes en 6 de los 10 estudios, y los estudios de modelado Bayesiano, utilizados en 4 de los 10 estudios. Los estudios transversales se emplearon en 3 de los 10 estudios, mientras que un estudio fue descrito como un desarrollo metodológico.

En cuanto a los métodos estadísticos, la regresión logística fue la técnica más utilizada, aplicada en 5 de los 10 estudios. Le siguieron la regresión de Cox y la regresión log-binomial, cada una presente en 3 de los 10 estudios. Los enfoques Bayesianos y los modelos de variables latentes se usaron en 2 de los 10 estudios. Se encontraron otros métodos, cada uno empleado en un estudio, incluyendo regresión de Poisson, regresión logística con ecuaciones de estimación generalizada (GEE), el algoritmo de Expectation-Maximization (EM), el enfoque de datos faltantes, la estimación de máxima verosimilitud y los modelos lineales generalizados.

### 4.2.5 Discusión

En los estudios examinados se evidenció que cada tipo de modelo presentó características particulares en términos de precisión. La regresión logística estándar tendió a sobrestimar las razones de prevalencia, mientras que tres modelos fueron descritos como mejoras en diversos aspectos de la estimación, ya sea en términos de eficiencia, precisión o calidad general de la estimación. Además, se identificó que dos modelos evitaron problemas

específicos, como la sobreestimación o inconvenientes metodológicos no especificados, y otros dos modelos fueron considerados más precisos o superiores a los enfoques estándar.

En cuanto a los requerimientos computacionales, se encontró información sobre los 11 modelos analizados. Cuatro modelos fueron descritos como altamente demandantes en términos de recursos computacionales, tres presentaron requerimientos moderados, dos fueron clasificados entre moderados y altos, y solo la regresión logística estándar fue identificada con bajos requerimientos computacionales. Esto sugiere que los modelos más avanzados, aunque potencialmente más precisos, pueden representar una barrera en su implementación debido a sus mayores demandas de procesamiento.

En relación con los desafíos de implementación, se identificaron dificultades específicas para cada tipo de modelo. Los problemas de convergencia fueron los más frecuentes, mencionados en dos modelos. Otras dificultades incluyeron la necesidad de tipos de datos específicos, la complejidad en la implementación y diversas cuestiones estadísticas. Estos hallazgos resaltan que, si bien ciertos modelos pueden ofrecer ventajas en términos de precisión, su aplicabilidad puede verse limitada por su viabilidad técnica y los recursos disponibles.

Con base en la información recopilada, parece existir un compromiso entre precisión, requerimientos computacionales y complejidad de implementación en los distintos enfoques de modelado para la estimación de prevalencia.

Dentro de esta revisión se identificaron dos modelos relevantes, para el tipo de estudio y el tipo de información que se obtiene en estudios de estimaciones de prevalencia usando test imperfectos binarios, para esto el uso de regresión logística es el modelo que más utilizado es y de amplia difusión entre la comunidad científica, de tal manera que se seleccionaron los modelos propuestos por Liu et al (59) y el publicado por (60), los cuales en las siguientes secciones son evaluados y comparados contra sus versiones estándar frecuentista y Bayesiana.

## 4.3 Estimación de la prevalencia mediante regresión logística estandar y Bayesiana

En muchos estudios epidemiológicos, la variable dependiente es una medición imperfecta del estado real de la enfermedad. Supongamos que queremos modelar la probabilidad de que un individuo tenga una enfermedad en función de un conjunto de covariables. Sin embargo, en lugar de observar directamente la presencia de la enfermedad ($D$),

observamos el resultado de una prueba diagnóstica ($t_1$), que puede contener errores de clasificación debido a su sensibilidad ($Se$) y especificidad ($Sp$).

Para este caso puede realizar una estimación de la proporción de $t_1$ resultados positivos por el test imperfecto y posteriormente llevar esta estimación a la corrección propuesta por Rogan y Galden (6). En las siguientes subsecciones se desarrollarán los modelos de regresión logística estándar con corrección de la prevalencia por Rogan-Gladen, regresión logística modificada propuesta por Liu (59), regresión logística bajo inferencia bayesiana con corrección externa por Rogan-Gladen y finalmente el modelo Bayesiano propuesto por Valle (60) con corrección a través de distribuciones previas incluidas sobre la sensibilidad y especificidad.

En esta sección se incluyeron los métodos propuestos en publicaciones que fueron encontrados dentro de la revisión sistemática de literatura realizada y descrita en la sección 4.2

## 4.4 Métodos de Estimación de Prevalencia mediante Regresión Logística

En esta sección se presentan los métodos utilizados para estimar la prevalencia cruda y ajustada mediante técnicas de regresión logística bajo diferentes estrategias de corrección por errores de clasificación, destacando la robustez de cada enfoque. Se utilizaron datos sobre seroprevalencia de infecciones de transmisión sexual (para más detalle, los datos son descritos en la sección

La prevalencia cruda de la infección se calculó utilizando el método de Rogan-Gladen (6), el cual ajusta la prevalencia observada considerando la sensibilidad ($Se$) y especificidad ($Sp$) de la prueba diagnóstica.

Se implementaron cuatro enfoques de regresión logística (dos de estos fueron encontrados mediante la revisión sistemática de la literatura descrita en 4.2) para estimar la prevalencia de dos infecciones de transmisión sexual: una de muy baja prevalencia (infección por Virus de la Inmunodeficiencia Humana, VIH) y otra de prevalencia moderadamente baja (infección por sífilis), bajo distintas estrategias de ajuste y corrección por errores de clasificación. A continuación se describen en detalle:

### 4.4.1 Regresión Logística Estándar (STD)

La regresión logística binaria modela la relación entre una variable de respuesta dicotómica ($Y$) (presencia o ausencia de enfermedad) y un conjunto de covariables ($X$). El modelo se define como:

$$P(Y = 1 \mid X) = \pi(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}, \tag{4.1}$$

donde:

- $\pi(X)$ es la probabilidad de que $Y = 1$, dado $X$.
- $\beta_0$ es el intercepto.
- $\beta_1, \ldots, \beta_p$ son coeficientes de regresión.

El modelo puede reescribirse en términos del logaritmo de los odds:

$$\log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p. \tag{4.2}$$

Los parámetros ($\beta$) se estiman por máxima verosimilitud, maximizando la función de verosimilitud:

$$L(\beta) = \prod_{i=1}^{n} \pi(X_i)^{Y_i} \cdot (1 - \pi(X_i))^{1-Y_i}. \tag{4.3}$$

El logaritmo de verosimilitud es:

$$\ell(\beta) = \sum_{i=1}^{n} \left[Y_i \log \pi(X_i) + (1 - Y_i) \log(1 - \pi(X_i))\right].$$

Las estimaciones de $\beta$ se obtienen resolviendo la ecuación:

$$\frac{\partial \ell(\beta)}{\partial \beta} = 0$$

Para la corrección de la prevalencia estimada y ajustada por covariables incluidas en el modelo, se aplicó el método propuesto por Rogan (6), corrigiendo extrínsecamente las prevalencias ajustadas según las características del modelo logístico.

### 4.4.2 Regresión logística modificada por Liu

Liu et al. (59) propuso un modelo de regresión logística que incorpora correcciones para la clasificación errónea en la variable de respuesta binaria. Siguiendo los planteamientos tradicionales sobre errores de clasificación no diferencial propuestos por Neuhaus (68) y Hausman et al. (69), el modelo asume que la probabilidad de mala clasificación es la misma para todos los sujetos.

Sea $(\tilde{Y})$ el verdadero estado de la variable de respuesta binaria. El modelo de regresión logística se define como:

$$\tilde{Y} \sim \text{Bernoulli}(F),$$

$$F = \frac{1}{1 + \exp(-g)},$$

$$g = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p,$$

donde los coeficientes $(\beta_1, \ldots, \beta_p)$ representan la asociación entre las covariables y la variable de resultado $(\tilde{Y})$.

En presencia de clasificación errónea, los datos observados $(y_i)$ pueden diferir de los valores verdaderos $(\tilde{y}_i)$. Se modela la probabilidad de misclasificación mediante la matriz de transición:

$$P(y_i = 1 \mid \tilde{y}_i = 0) = r_0, \tag{4.4}$$

$$P(y_i = 0 \mid \tilde{y}_i = 0) = 1 - r_0,$$

$$P(y_i = 0 \mid \tilde{y}_i = 1) = r_1, \tag{4.5}$$

$$P(y_i = 1 \mid \tilde{y}_i = 1) = 1 - r_1,$$

donde $(r_0)$ y $(r_1)$ representan las tasas de falsos positivos (FP) y falsos negativos (FN), respectivamente.

El modelo extendido de regresión logística, que incorpora estos parámetros de clasificación errónea, se define como:

$$\begin{aligned}
\pi_i &= \Pr(y_i = 1 \mid \boldsymbol{x}_i) \\
&= \Pr(y_i = 1 \mid \tilde{y}_i = 1, \boldsymbol{x}_i) \Pr(\tilde{y}_i = 1 \mid \boldsymbol{x}_i) + \Pr(y_i = 1 \mid \tilde{y}_i = 0, \boldsymbol{x}_i) \Pr(\tilde{y}_i = 0 \mid \boldsymbol{x}_i) \\
&= (1 - r_1) \Pr(\tilde{y}_i = 1 \mid \boldsymbol{x}_i) + r_0 [1 - \Pr(\tilde{y}_i = 1 \mid \boldsymbol{x}_i)] \\
&= r_0 + (1 - r_0 - r_1) \Pr(\tilde{y}_i = 1 \mid \boldsymbol{x}_i) \\
&= r_0 + (1 - r_0 - r_1) F_i.
\end{aligned}$$

$$P(y_i = 1 \mid x_i) = r_0 + (1 - r_0 - r_1) P(\tilde{y}_i = 1 \mid x_i), \tag{4.6}$$

donde la probabilidad de $(\tilde{y}_i = 1) \mid (x_i)$ sigue la estructura del modelo logístico convencional:

$$P(\tilde{y}_i = 1 \mid x_i) = \frac{1}{1 + \exp(-g_i)},$$

con $(g_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji})$.

Este modelo permite estimar simultáneamente los coeficientes de regresión ajustados y las tasas de error de clasificación.

**Estimación de coeficientes del modelo de Liu mediante el algoritmo de puntuación de Fisher**

La estimación de los parámetros en los modelos de regresión logística que incorporan parámetros de error de clasificación. Debido a la naturaleza no lineal del modelo y a la interacción entre los parámetros de clasificación y los coeficientes de regresión, no existen soluciones analíticas de forma cerrada para las estimaciones de máxima verosimilitud (ML). Por lo tanto, se recurre a métodos numéricos, específicamente al algoritmo de puntuación de Fisher.

El algoritmo se basa en las ecuaciones de estimación derivadas de la maximización de la verosimilitud. La probabilidad de observar $Y_i = y_i$ dado el vector de covariables $x_i$ se define como:

$$P(Y_i = y_i|x_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} = \exp y_i h_i - \log(1+\exp(h_i)) \tag{4.7}$$

donde $h_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ y $p_i = r_0 + (1-r_0-r_1)F_i$, con $F_i = \frac{\exp(g_i)}{1+\exp(g_i)}$ y $g_i = x_i^T\beta$.

Aquí, $r_0$ es la probabilidad de un falso positivo y $r_1$ es la probabilidad de un falso negativo. $\beta = (\beta_0, \beta_1, ..., \beta_p)^T$ son los coeficientes de regresión. Las funciones de puntuación (las primeras derivadas del logaritmo de la función de verosimilitud $l$ con respecto a los parámetros $r_0, r_1, \beta_0, ..., \beta_p$) se expresan como:

$$\frac{\partial l}{\partial c} = \sum_{i=1}^{n} \frac{y_i - p_i}{p_i(1-p_i)} \frac{\partial p_i}{\partial c} \tag{4.8}$$

donde $c$ representa cada uno de los parámetros. Las derivadas de $\pi_i$ con respecto a los parámetros son:

$\frac{\partial \pi_i}{\partial r_0} = 1 - F_i$

$\frac{\partial \pi_i}{\partial r_1} = -F_i$

$\frac{\partial \pi_i}{\partial \beta_j} = (1-r_0-r_1)F_i(1-F_i)x_{ij}$ (con $x_{i0} = 1$)

El algoritmo de puntuación de Fisher utiliza la matriz de información de Fisher $I(c)$, que es el negativo del valor esperado de la matriz Hessiana (segundas derivadas del logaritmo de la función de verosimilitud) o, equivalentemente, el valor esperado del producto exterior del vector de puntuación:

$$I(c) = -E\left[\frac{\partial^2 l}{\partial c \partial c^T}\right] = E\left[\left(\frac{\partial l}{\partial c}\right)\left(\frac{\partial l}{\partial c}\right)^T\right] \tag{4.9}$$

La matriz de información de Fisher se puede expresar de forma compacta como:

$$I(c) = D^T W D$$

donde $D$ es una matriz $n \times (p+3)$ con la $i$ -ésima fila siendo el gradiente de $\pi_i$ con respecto a los parámetros $(\frac{\partial \pi_i}{\partial r_0}, \frac{\partial \pi_i}{\partial r_1}, \frac{\partial \pi_i}{\partial \beta_0}, ..., \frac{\partial \pi_i}{\partial \beta_p})$, y $W$ es una matriz diagonal $n \times n$ con elementos diagonales $\frac{1}{\pi_i(1-\pi_i)}$.

El algoritmo iterativo de Fisher scoring actualiza las estimaciones de los parámetros $c^{(t)}$ en cada paso $t$ de la siguiente manera:

$$c^{(t+1)} = c^{(t)} + (I(c^{(t)}))^{-1}\frac{\partial l(c^{(t)})}{\partial c} = c^{(t)} + (D^{(t)T}W^{(t)}D^{(t)})^{-1}D^{(t)T}W^{(t)}(y - p^{(t)}) \quad (4.10)$$

donde todas las matrices y vectores se evalúan en las estimaciones de los parámetros en el paso actual $t$. El algoritmo itera hasta el cumplimiento del criterio de convergencia (Por ejemplo: $10^{-6}$). Bajo ciertas condiciones de regularidad, las estimaciones $\hat{c}$ obtenidas al converger el algoritmo son asintóticamente insesgadas y siguen una distribución normal con una matriz de covarianza estimada por la inversa de la matriz de información de Fisher evaluada en las estimaciones finales:

$$cov(\hat{c}) \approx I^{-1}(\hat{c}) = (\hat{D}^T\hat{W}\hat{D})^{-1} \quad (4.11)$$

Las desviaciones estándar de las estimaciones de los parámetros se obtienen como las raíces cuadradas de los elementos diagonales correspondientes de esta matriz de covarianza.

### 4.4.3 Regresión logística con estimación Bayesiana (BC)

La regresión logística Bayesiana extiende el modelo de regresión logística tradicional al incorporar un marco de inferencia probabilística. Las estimaciones puntuales para los parámetros del modelo ($\beta$) se hacen mediante la inferencia de distribuciones de probabilidad sobre estos parámetros.

Dado un conjunto de datos ($\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$), donde ($y_i$) es una variable de respuesta dicotómica ($y_i \in \{0, 1\}$) y ($x_i$) es un vector de covariables, asumiendo el modelo de regresión logística tradicional según la ecuación (4.1)

En el enfoque Bayesiano, los coeficientes ($\beta$) se modelan con distribuciones previas definidas por distintos métodos, sin embargo, recientemente Newman et al (70) proponen una metodología novedosa en el caso de estimación Bayesiana para regresión logística. Las distribuciones a previas para los coeficientes ($\beta_p$) pueden ser especificadas según la media y varianza esperada para una distribución logit y su aproximación a la normal. El desarrollo de este método se revisará en la sección 4.4.3. Las distribuciones previas de los coeficientes en el modelo Bayesiano según Newman serían:

$$\beta_i \sim \mathcal{N}\left(0, \frac{\sigma_\beta^2}{p+1}\right), \quad i = 0, \ldots, p. \tag{4.12}$$

Donde: $\mu_\beta = 0, \quad \sigma_\beta^2 = \frac{\pi^2}{3}$ y $p$ covariables

La distribución posterior de los coeficientes es obtenida con la función de verosimilitud según ecuación (4.3) y distribuciones previas como la expresión (4.12) mediante el teorema de Bayes.

La estimación puntual y por intervalo de la prevalencia obtenida por el modelo Bayesiano clásico pueden ser corregidas externamente por los errores de clasificación mediante la ecuación (4.6).

### Especificación de distribuciones previas en Modelos de Regresión Logística Bayesiana

Uno de los pasos más importantes durante la inferencia Bayesiana es la definición de las distribuciones previas a usar, las cuales consideran la información previa que se tiene sobre los parámetros, en este caso sobre los coeficientes a estimar. Esto es denominado por Ohagan et al. (71) como la elicitación de distribuciones que corresponde al proceso sistemático mediante el cual se traduce el conocimiento experto, creencias subjetivas o información externa previa al estudio en una distribución de probabilidad que represente la incertidumbre sobre un parámetro antes de observar los datos . Este proceso puede incluir técnicas cualitativas (entrevistas, cuestionarios) y cuantitativas, y para esta última forma existen varias metodologías propuestas por Ohagan et al. (72), para efectos de este estudio se revisó la publicada recientemente por (73) donde su planteamiento no solo incluye el comportamiento en la varianza de los coeficientes como información previa que se podría tener, sino también relación al número de variables incluidas en el modelo. En atención al modelo propuesto en la ecuación 4.2, denominamos a esta transformación logística como (logit($\theta$)) bajo la inferencia Bayesiana, se especifican distribuciones previas para los coeficientes $(\beta_0, \beta_1, \ldots, \beta_p)$, estas distribuciones previas inducen una distribución previa en $\theta$.

La elección de distribuciones previas para la inferencia Bayesiana y su efecto ha sido por Lesafre (74), sin embargo, el uso de distribuciones previas no informativas es de preocupación, sobre todo las que están por defecto preconfiguradas en varios software.

Una consideración metodológica relevante en la modelación Bayesiana jerárquica es que la elección de distribuciones prior supuestamente no informativas en niveles inferiores

del modelo puede inducir una estructura prior fuertemente informativa sobre parámetros de niveles superiores. Seaman et al. (75) documentó este comportamiento en modelos de regresión logística, donde distribuciones como una normal $N(0, 25^2)$ asignadas a los coeficientes de regresión pueden generar distribuciones inducidas sobre la probabilidad $\theta$ con formas altamente concentradas en los extremos del espacio paramétrico, particularmente en torno a 0 y 1 .

### Priorización inducida mediante transformación de parámetros

Cuando se desea especificar un prior particular sobre un parámetro transformado, por ejemplo, sobre la probabilidad $(\theta)$ en un modelo logístico, es posible derivar la distribución prior correspondiente sobre el parámetro original, en este caso $(\beta)$.

Un enfoque común consiste en definir la transformación logit inversa, $(\theta = \exp(\beta)/(1+\exp(\beta)))$, y calcular la densidad inducida para $(\beta)$ a partir de una distribución deseada para $(\theta)$. Por ejemplo, si se desea que $(\theta)$ siga una distribución uniforme en $(0, 1)$, la densidad inducida para $(\beta)$ resulta ser:

$$\pi_\beta(\beta) = \frac{\exp(\beta)}{(1 + \exp(\beta))^2}$$

la cual corresponde a una distribución logística estándar con media cero y varianza $(\pi^2/3)$. Este resultado establece que utilizar una distribución previa Logistic$(0, 1))$ para $(\beta)$ equivale a imponer un prior uniforme sobre $\theta$.

Este enfoque puede generalizarse para obtener distribuciones previas sobre $\beta$ que induzcan otras distribuciones deseadas sobre $\theta$, como una distribución beta.

### Distribución logística y aproximación normal

Como se ha demostrado, especificar un prior uniforme sobre la probabilidad $\theta$ en un modelo logístico induce una distribución sobre $\beta$ cuya función de densidad corresponde a una distribución logística estándar. Esta distribución, definida por un parámetro de localización $(\mu)$ y un parámetro de escala $(s)$, tiene la siguiente forma funcional:

$$f(x; \mu, s) = \frac{\exp\left(-\frac{x-\mu}{s}\right)}{s\left(1 + \exp\left(-\frac{x-\mu}{s}\right)\right)^2}.$$

Cuando $\mu = 0$ y $s = 1$, se obtiene la distribución Logistic$(0, 1)$, cuya esperanza es nula y cuya varianza es $\pi^2/3$. La simetría de esta densidad en torno a cero implica que

$\beta$ y $-\beta$ tienen la misma distribución. Por lo tanto, si se utiliza un prior Logistic$(0, 1)$ sobre $\beta$, el prior inducido sobre $\theta$ es uniforme en el intervalo $(0, 1)$.

Este resultado es útil para construir prior en modelos Bayesianos con interpretaciones probabilísticas directas, aunque en muchos contextos puede ser más apropiado especificar un prior distinto sobre $\theta$, como una distribución beta. En tales casos, se requiere una transformación inversa y el ajuste del prior sobre $\beta$ para inducir la forma deseada sobre $\theta$.

Dado que la distribución Logistic$(0, 1)$ es simétrica y unimodal, puede ser razonablemente aproximada por una distribución normal con media cero y varianza $\pi^2/3$, es decir, Normal$(0, \pi^2/3)$.

### Asignación de distribuciones previas en regresión logística multivariada

Ahora se considera el caso más general de la regresión logística, donde $\theta$ es una función de $\mathtt{p}$ covariables, si el prior para $\eta$ sigue una distribución Logistic$(0, 1)$, entonces el prior inducido para $\theta$ es Uniforme$(0, 1)$. Por lo tanto, requerimos que:

$$\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \sim \text{Logistic}(0, 1),$$

donde hemos supuesto que $x_0 = 1$.

Existen dos complicaciones en este contexto. En primer lugar, dado que hay $(\mathtt{p} + 1)$ parámetros $\beta$, existen $(\mathtt{p} + 1)$ distribuciones previas. Esto significa que hay un mapeo de muchos a uno desde los $\beta$ hacia el único $\theta$. En teoría, hay un número infinito de combinaciones posibles de distribuciones previas para los $\beta$ que inducen una distribución Logistic$(0, 1)$.

En segundo lugar, los valores de $x_0, x_1, \ldots, x_{\mathtt{p}}$ afectarán la distribución de $\eta$. Aquí se examinan soluciones aproximadas a estos dos problemas, con el objetivo de hacer coincidir la media y la varianza de $\eta$. En un escenario más realista, se consideran modelos logísticos lineales que incluyen covariables. Supondremos que dichas covariables $x_1, \ldots, x_{\mathtt{p}}$ han sido estandarizadas, es decir:

$$x_i = \frac{x_i^* - \bar{x}_i^*}{s_{x_i^*}},$$

donde $x_i^*$ son los valores originales, y por construcción $\mathbb{E}[x_i] = 0$, $\text{Var}(x_i) = 1$. Para evitar problemas de identificación asociados a múltiples combinaciones de parámetros que generan el mismo modelo, se asume que los coeficientes $\beta_0, \beta_1, \ldots, \beta_{\mathtt{p}}$ son independientes e idénticamente distribuidos (i.i.d.), con una distribución previa común $\pi_\beta(\beta)$.

Aplicando la técnica de correspondencia de momentos, se buscan condiciones sobre $\pi_\beta(\beta)$ que aseguren que la variable lineal $\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_{\mathtt{p}}$ tenga media cero y varianza $\pi^2/3$, como ocurre en el caso logístico estándar.

La condición sobre la media se cumple si $\mathbb{E}[\beta_0] = 0$ y $\mathbb{E}[\beta_i] = 0$, ya que las covariables estandarizadas tienen media cero. Así, se garantiza que:

$$\mathbb{E}[\eta] = \mathbb{E}[\beta_0] + \sum_{i=1}^{p} \mathbb{E}[\beta_i] \cdot \mathbb{E}[x_i] = 0.$$

Para la varianza, se calcula:

$$\mathrm{Var}[\eta] = \mathrm{Var}\left[\beta_0 + \sum_{i=1}^{p} \beta_i x_i\right] = \mathrm{Var}[\beta_0] + \sum_{i=1}^{p} \mathbb{E}[\beta_i^2] \cdot \mathrm{Var}[x_i] = \sigma_\beta^2(1+p),$$

asumiendo que $\mathrm{Var}[\beta_i] = \sigma_\beta^2$ y $\mathrm{Var}[x_i] = 1$ para todo $i$. Para que esta varianza coincida con la de la distribución logística estándar ($\pi^2/3$), se requiere que:

$$\sigma_\beta^2 = \frac{\pi^2}{3(p+1)}. \tag{4.13}$$

Este resultado proporciona una guía práctica para la elección de la varianza del prior sobre los coeficientes $\beta$ cuando se incorporan covariables estandarizadas, asegurando que la varianza inducida sobre el predictor lineal coincida con la de un modelo logístico base sin covariables.

### 4.4.4 Modelo Bayesiano con regresión logística con corrección por mala clasificación (BEC)

El Modelo propuesto por Valle et al. (60) plantea la estimación de la regresión logística en presencia de errores de clasificación en el diagnóstico, incorporando información externa sobre la sensibilidad ($Se$) y especificidad ($Sp$) del test diagnóstico mediante el uso de distribuciones previas informativas. Se asume que el estado de infección $y_i$ sigue un modelo de regresión logística:

$$\tilde{Y}_i \sim \text{Bernoulli}\left(\frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}\right), \tag{4.14}$$

donde:

- $\tilde{Y}_i$ es el verdadero estado de infección del individuo $i$.
- $\beta_0, \beta_1, \ldots, \beta_p$ son los coeficientes de regresión.
- $x_p$ son las covariables del modelo.

El resultado observado del test diagnóstico $D_i$ se modela como:

$$Y_i \sim \text{Bernoulli}(Se) \quad \text{si } \tilde{Y}_i = 1, \tag{4.15}$$

$$Y_i \sim \text{Bernoulli}(1 - Sp) \quad \text{si } \tilde{Y}_i = 0, \tag{4.16}$$

donde:

- $Se$ es la sensibilidad del test (probabilidad de clasificar correctamente un caso positivo).
- $Sp$ es la especificidad del test (probabilidad de clasificar correctamente un caso negativo).

## 4.5 Medidas para comparación o desempeño

### 4.5.1 Comparación de la Prevalencia Ajustada y Corregida vs. la Medida Cruda

De acuerdo a lo sugerido por Morris et al. (76) para comparar la estimación puntual de la prevalencia ajustada y corregida por sensibilidad y especificidad en cada modelo respecto a la prevalencia cruda ajustada por Rogan-Gladen, se utilizaron dos indicadores:

- **Cambio Relativo en la Prevalencia Estimada vs. la Cruda**

$$\text{Cambio} = \left( \frac{\hat{P}_{\text{modelo}} - \hat{P}_{\text{cruda}}}{\hat{P}_{\text{cruda}}} \right) \times 100,$$

donde $\hat{P}_{\text{modelo}}$ es la prevalencia ajustada y corregida en cada método, y $\hat{P}_{\text{cruda}}$ es la prevalencia estimada por Rogan-Gladen.

- **Amplitud del Intervalo de Confianza (IC)**

Se comparó la precisión de cada método evaluando el ancho de los intervalos de confianza:

$$\text{Amplitud IC} = \text{IC superior} - \text{IC inferior}$$

Modelos con menor amplitud de IC indicaron mayor precisión en la estimación de la prevalencia

### 4.5.2 Comparación de los coeficientes de regresión entre métodos que incluian errores de clasificación como parametros

Para evaluar las diferencias en la estimación de los coeficientes de regresión entre los métodos denominados **Liu** y **BEC**, se utilizaron las siguientes métricas:

$$\text{Cambio en Coeficiente} = \left( \frac{\hat{\beta}_{\text{Liu}} - \hat{\beta}_{\text{BEC}}}{\hat{\beta}_{\text{BEC}}} \right) \times 100.$$

Un valor positivo indica que el método Liu estima coeficientes mayores que el Bayesiano con corrección (BEC).

- **Incremento Relativo de Precisión**

$$\text{Incremento Relativo de Precisión} = \left( \frac{\text{Var}(\hat{\beta}_{\text{Liu}})}{\text{Var}(\hat{\beta}_{\text{BEC}})} - 1 \right).$$

Un valor positivo indica que el método Bayesiano con corrección (BEC) tiene menor varianza y, por tanto, es más preciso.

## 4.6 Datos y Variables

Se analizaron datos correspondientes a 11452 sujetos que fueron sometidos a un kit de pruebas diagnósticas rápidas para la detección de infecciones de transmisión sexual(ITS), específicamente dirigidas a identificar anticuerpos contra VIH-1, VIH-2, sífilis y hepatitis B, con el propósito de estimar la seroprevalencia de infección por VIH.

Como covariables, se incluyeron características demográficas relevantes como sexo y edad, así como los resultados de las pruebas diagnósticas para sífilis y hepatitis B, y el

Tabla 4.2: Asignación de coeficientes a las covariables del modelo

| **Covariable** | **Coeficiente asignado** |
|---|---|
| Intercepto | $\beta_0$ |
| Edad (años) | $\beta_1$ |
| Sexo | $\beta_2$ |
| Prueba sifilis* | $\beta_3$ |
| Prueba Hepatitis B | $\beta_4$ |
| HSH** | $\beta_5$ |
| Poblacion LGTBI | $\beta_6$ |
| Otras poblaciones | $\beta_7$ |
| Trabajador sexual | $\beta_8$ |

* El coeficiente $\beta_3$ para los modelos de prevalencia de sifilis, representa el resultado del test de VIH.
** HSH: Hombre que tiene sexo con otros hombres

tipo de población clave, el cual fue incorporado en los modelos mediante variables indicadoras (dummy variables). La asignación de coeficientes a cada covariable se presenta en la tabla 4.2

Las pruebas diagnósticas utilizadas incluyeron el test **Determine HIV Early Detect** para la detección de VIH, el cual reporta una sensibilidad de 0.97 y una especificidad de 0.99. Para la detección de sífilis se empleó la prueba **Bioline Syphilis 3.0**, con una sensibilidad de 0.964 y una especificidad de 0.974, según la información técnica disponible.

## 4.7 Resultados

### 4.7.1 Caracterización de la Muestra

La población de estudio incluyó un total de 11452 sujetos, cuyas características demográficas y serológicas se resumen en la Tabla 4.3. La edad media fue de 34 años (desviación estándar: 14). La distribución por sexo fue equilibrada, con 5759 hombres, lo que corresponde al 50 % de la muestra.

En cuanto a los resultados serológicos, 155 participantes (1.4%) presentaron reactividad para VIH, 905 (7.9%) para sífilis y 8 (<0.1%) para hepatitis B.

Respecto al tipo de población clave, la mayor proporción correspondió a población general con 6574 individuos (57 %), seguida de hombres que tienen sexo con hombres

Tabla 4.3: Características descriptivas de la población estudiada (n = 11452)

| **Característica** | **n(%)** |
|---|---|
| Edad (media [DE]) | 34 (14) |
| Sexo masculino | 5759 (50%) |
| Reactividad para VIH | 155 (1.4%) |
| Reactividad para sífilis | 905 (7.9%) |
| Reactividad para hepatitis B | 8 (<0.1%) |
| **Tipo de población** | |
| HSH | 3,248 (28%) |
| LGTBIQ | 224 (2.0%) |
| Otras poblaciones | 193 (1.7%) |
| Población general | 6574 (57%) |
| Trabajador sexual | 1213 (11%) |

Las variables continuas se presentan como media (desviación estándar), y las categóricas como n (porcentaje).

(HSH) con 3248 (28%) y trabajadores sexuales con 1213 (11 %). También se incluyeron personas identificadas como parte de la comunidad LGTBIQ (n = 224; 2,0 %) y otras poblaciones (n = 193; 1,7 %).

### 4.7.2 Comparación prevalencias ajustadas

La Tabla 4.4 presenta las estimaciones ajustadas de prevalencia para VIH y sífilis bajo distintos modelos de regresión con corrección por clasificación errónea. Se evidencian variaciones importantes en las estimaciones según el enfoque metodológico utilizado.

Para VIH, la prevalencia cruda fue de 1.39%. El modelo estándar (STD), que corrige externamente por sensibilidad y especificidad conocidas, redujo esta estimación a 0.73%, mientras que el modelo de Liu, que estima simultáneamente los parámetros de error diagnóstico, arrojó una prevalencia de 1.66%, con un aumento relativo del 127.69% frente al modelo STD. Por su parte, el modelo Bayesiano con corrección por clasificación errónea (BEC) estimó una prevalencia de 1.00%, con una amplitud del intervalo de confianza de 0.0049.

En el caso de sífilis, la prevalencia cruda fue de 5.67%. El modelo STD arrojó una estimación ajustada considerablemente menor (2.77%), mientras que el modelo de Liu estimó una prevalencia de 11.55%, más del doble de la cruda y con la mayor amplitud de intervalo (0.0187). El modelo BEC, en cambio, estimó una prevalencia de 4.14% con

Tabla 4.4: Estimaciones ajustadas de prevalencia para VIH y sífilis según distintos modelos de regresión

| Enfermedad | Modelo | P ajustada | Inferior | Superior | Cambio vs. Cruda (%) | Cambio vs. STD (%) | Amplitud IC |
|---|---|---|---|---|---|---|---|
| VIH | Cruda | 0.0139 | 0.0117 | 0.0160 | – | – | 0.0043 |
| | STD | 0.0073 | 0.0055 | 0.0097 | -47.42 | – | 0.0042 |
| | Liu | 0.0166 | 0.0131 | 0.0212 | 19.72 | 127.69 | 0.0081 |
| | BC | 0.0095 | 0.0078 | 0.0116 | -31.57 | 30.15 | 0.0037 |
| | BEC | 0.0100 | 0.0078 | 0.0127 | -28.24 | 36.47 | 0.0049 |
| Sífilis | Cruda | 0.0567 | 0.0514 | 0.0620 | – | – | 0.0106 |
| | STD | 0.0277 | 0.0223 | 0.0336 | -51.21 | – | 0.0113 |
| | Liu | 0.1155 | 0.1064 | 0.1252 | 103.64 | 317.39 | 0.0187 |
| | BC | 0.0340 | 0.0288 | 0.0391 | -40.08 | 22.82 | 0.0103 |
| | BEC | 0.0414 | 0.0379 | 0.0452 | -26.94 | 49.74 | 0.0073 |

*Nota:* STD = modelo estándar con corrección externa; Liu = modelo con estimación simultánea de parámetros de error; BC = modelo Bayesiano sin corrección interna; BEC = modelo Bayesiano con corrección por clasificación errónea.

Tabla 4.5: Comparación de coeficientes de regresión para VIH según modelo implementado

| Coeficiente | STD | Liu | Cambio Liu (%) | BC | BEC | Cambio BEC (%) |
|---|---|---|---|---|---|---|
| $\beta_0$ | -5.864 | -4.648 | -20.73 | -5.418 | -4.717 | -12.93 |
| $\beta_1$ | -0.001 | 0.002 | -325.76 | -0.001 | -0.015 | 2798.64 |
| $\beta_2$ | 0.991 | 0.356 | -64.09 | 0.749 | 0.512 | -31.68 |
| $\beta_3$ | 1.946 | 1.492 | -23.33 | 1.782 | 1.892 | 6.18 |
| $\beta_4$ | 1.547 | 1.545 | -0.14 | 0.433 | 0.454 | 4.86 |
| $\beta_5$ | 0.883 | 0.617 | -30.15 | 0.754 | 0.717 | -4.96 |
| $\beta_6$ | 1.214 | 0.604 | -50.23 | 0.664 | 0.513 | -22.74 |
| $\beta_7$ | 0.673 | 0.428 | -36.33 | 0.326 | 0.319 | -2.01 |
| $\beta_8$ | 0.516 | 0.060 | -88.44 | 0.191 | -0.138 | -171.98 |

la menor amplitud del intervalo (0.0073), indicando mayor precisión relativa entre los modelos comparados.

En general, los modelos Bayesianos mostraron un mejor compromiso entre ajuste por error de clasificación y precisión de las estimaciones, particularmente el modelo BEC, que ofreció valores intermedios de prevalencia con los intervalos de confianza más estrechos en ambos desenlaces.

### 4.7.3 Comparación de coeficientes de regresión para modelos de VIH

La Tabla 4.5 muestra los coeficientes estimados para el modelo de regresión logística de VIH bajo cuatro enfoques: el modelo estándar (STD), el modelo de Liu (ajuste simultáneo de parámetros de clasificación errónea), el modelo Bayesiano sin corrección interna (BC), y el modelo Bayesiano con corrección por clasificación errónea (BEC). Se incluyen también los porcentajes de cambio relativos con respecto al modelo estándar.

El intercepto mostró una reducción del 20.73% en el modelo de Liu y del 12.93% en el modelo Bayesiano con corrección, lo que sugiere diferencias en la estimación del nivel basal de prevalencia cuando se consideran los errores diagnósticos.

En términos generales, se observan diferencias sustanciales en la magnitud y dirección de algunos coeficientes según el método de ajuste utilizado. El modelo de Liu muestra reducciones notables frente al modelo estándar en la mayoría de los coeficientes, con un cambio de hasta -88.44% para el coeficiente asociado a la población trabajadora sexual ($\beta_8$), y de -64.09% para el sexo masculino ($\beta_2$), indicando un ajuste sustancial cuando se corrige simultáneamente por sensibilidad y especificidad del test.

Por el contrario, el modelo Bayesiano con corrección por error de clasificación (BEC) no solo ajusta en magnitud sino también en dirección algunos coeficientes. El más llamativo es el cambio en $\beta_8$, que pasa de un valor positivo en el modelo estándar (0.516) a negativo (-0.138) en BEC, con una variación relativa de disminución de 1.7 veces con relación al modelo BC.

Otro hallazgo relevante se observa en el coeficiente $\beta_1$ (edad agrupada), que cambia de -0.001 en el modelo estándar a -0.015 en BEC, con un incremento relativo de más aproximadamente 3 veces con relación al valor en el modelo BC. Aunque su magnitud sigue siendo pequeña, este cambio refleja sensibilidad del efecto de la edad a la forma en que se modela la clasificación errónea.

La covariable asociada a la reactividad para sífilis ($\beta_3$) mantiene una dirección positiva en todos los modelos, con un leve aumento en BEC (1.892) respecto a STD (1.946), sugiriendo estabilidad del efecto de esta covariable, aunque con ajuste por error de medición.

La Tabla 4.6 compara los errores estándar (SE) de los coeficientes estimados para el modelo de VIH entre el enfoque de Liu y el modelo Bayesiano con corrección por clasificación errónea (BEC). En general, se observan diferencias notables en la precisión de los estimadores, dependiendo del enfoque metodológico implementado.

Tabla 4.6: Comparación de errores estándar de los coeficientes del modelo de VIH entre Liu y modelo Bayesiano con corrección (BEC)

| **Coeficiente** | **SE (Liu)** | **SE (BEC)** | **Incremento relativo** |
|---|---|---|---|
| $\beta_0$ | 0.694 | 0.237 | 7.539 |
| $\beta_1$ | 0.004 | 0.006 | -0.639 |
| $\beta_2$ | 0.229 | 0.228 | 0.008 |
| $\beta_3$ | 0.311 | 0.168 | 2.424 |
| $\beta_4$ | 0.847 | 0.553 | 1.345 |
| $\beta_5$ | 0.230 | 0.200 | 0.327 |
| $\beta_6$ | 0.415 | 0.397 | 0.089 |
| $\beta_7$ | 0.387 | 0.402 | -0.073 |
| $\beta_8$ | 0.172 | 0.311 | -0.695 |

El modelo BEC presenta una reducción marcada del error estándar del intercepto (de 0.694 a 0.237), correspondiente a un incremento relativo en precisión de 7 veces más, lo que sugiere una mayor estabilidad en la estimación de la prevalencia base corregida. De forma similar, se observan reducciones en el error estándar de la covariable asociada a la reactividad para sífilis, con un incremento relativo en precisión de 2.42 veces respecto al modelo de Liu. También se observa una mejora sustancial para la covariable de hepatitis B.

En contraste, algunas covariables como edad agrupada y trabajador sexual mostraron un incremento en su error estándar bajo el modelo BEC, lo cual podría reflejar mayor incertidumbre asociada a su efecto, posiblemente debido a baja frecuencia o interacción con otras variables.

Estos hallazgos indican que el modelo BEC no solo ajusta las estimaciones puntuales de los coeficientes, sino que también mejora de forma diferencial la precisión de las estimaciones, especialmente en los componentes más relevantes para la inferencia poblacional. Esto refuerza su utilidad en escenarios donde el error diagnóstico puede comprometer tanto la validez como la precisión de los modelos tradicionales.

### 4.7.4 Comparación de coeficientes de regresión para modelos de Sífilis

La Tabla 4.7 presenta los coeficientes estimados para la regresión logística de sífilis bajo los distintos modelos evaluados. Se observan diferencias relevantes en magnitud y dirección entre los enfoques, particularmente al comparar los modelos corregidos por

Tabla 4.7: Comparación de coeficientes de regresión para sífilis según modelo implementado

| Coeficiente | STD | Liu | Cambio Liu (%) | BC | BEC | Cambio BEC (%) |
|---|---|---|---|---|---|---|
| $\beta_0$ | -4.890 | -3.164 | -35.30 | -4.766 | -5.322 | 11.67 |
| $\beta_1$ | 0.036 | 0.022 | -38.72 | 0.035 | 0.037 | 5.04 |
| $\beta_2$ | 0.615 | 0.159 | -74.15 | 0.575 | 0.728 | 26.62 |
| $\beta_3$ | 1.986 | 1.525 | -23.22 | 1.839 | 2.061 | 12.09 |
| $\beta_4$ | 2.140 | 1.743 | -18.55 | 0.747 | 0.719 | -3.71 |
| $\beta_5$ | 0.954 | 0.597 | -37.45 | 0.903 | 1.004 | 11.24 |
| $\beta_6$ | 1.473 | 0.672 | -54.39 | 1.222 | 1.371 | 12.19 |
| $\beta_7$ | 1.470 | 0.921 | -37.39 | 1.288 | 1.439 | 11.77 |
| $\beta_8$ | 1.825 | 0.886 | -51.43 | 1.691 | 1.990 | 17.67 |

error de clasificación con el modelo estándar.

El modelo de Liu mostró una reducción sistemática en la mayoría de los coeficientes, destacándose el sexo ($\beta_2$), con un cambio de -74.15%, y la categoría de trabajador sexual ($\beta_8$), con una disminución del 51.43%. Estos resultados podrían reflejar una sobreestimación de las asociaciones bajo el modelo estándar cuando no se corrige por mala clasificación.

Por su parte, el modelo Bayesiano con corrección (BEC) conservó la dirección positiva de la mayoría de los efectos, e incluso aumentó su magnitud en comparación con el modelo estándar. Por ejemplo, el coeficiente para trabajador sexual pasó de 1.825 a 1.990, con un incremento del 17.67%, mientras que el efecto del sexo aumentó en un 26.62%. Estos cambios indican que, tras corregir la clasificación errónea, la asociación entre estos factores y la probabilidad de infección por sífilis se refuerza, lo cual puede ser relevante desde el punto de vista epidemiológico y de intervención.

En contraste, el efecto del reactivo para hepatitis B ($\beta_4$) mostró una reducción en BEC respecto al modelo estándar, sugiriendo un posible ajuste hacia un efecto más conservador. La edad agrupada ($\beta_1$) mantuvo un efecto positivo y estable en todos los modelos.

La Tabla 4.8 se evidencian variaciones moderadas en la precisión de los estimadores entre ambos modelos, siendo el comportamiento distinto al observado en el análisis para VIH.

El modelo BEC mostró una mejora considerable en la precisión del intercepto, con una reducción del error estándar de 0.367 a 0.166, lo que representa un incremento relativo en precisión de más del 3.8 veces. Este hallazgo sugiere que la estimación de la

Tabla 4.8: Comparación de errores estándar en coeficientes del modelo de sífilis: modelo de Liu vs. Bayesiano con corrección (BEC)

| **Coeficiente** | **SE (Liu)** | **SE (BEC)** | **Incremento relativo** |
|---|---|---|---|
| $\beta_0$ | 0.367 | 0.166 | 3.884 |
| $\beta_1$ | 0.004 | 0.003 | 0.627 |
| $\beta_2$ | 0.062 | 0.165 | -0.857 |
| $\beta_3$ | 0.187 | 0.184 | 0.031 |
| $\beta_4$ | 0.675 | 0.516 | 0.711 |
| $\beta_5$ | 0.109 | 0.126 | -0.255 |
| $\beta_6$ | 0.193 | 0.285 | -0.544 |
| $\beta_7$ | 0.191 | 0.230 | -0.311 |
| $\beta_8$ | 0.168 | 0.164 | 0.047 |

prevalencia base para sífilis es más estable bajo un enfoque Bayesiano con corrección.

Sin embargo, en varias covariables —como sexo, tipo de población LGTBIQ y otras poblaciones— se observó un aumento del error estándar bajo el modelo BEC. Por ejemplo, el SE para sexo aumentó de 0.062 a 0.165, lo cual podría reflejar mayor incertidumbre en la estimación del efecto de esta covariable al incorporar explícitamente la incertidumbre diagnóstica. Un comportamiento similar se observó en las categorías poblacionales, con incrementos negativos del orden de -25% a -54%.

En contraste, variables como edad , reactividad a hepatitis B y trabajador sexual mostraron errores estándar similares o levemente más bajos bajo el modelo Bayesiano, lo que sugiere que para estas variables la precisión de la estimación se mantiene o incluso mejora.

En conjunto, estos resultados indican que el modelo Bayesiano con corrección no solo ajusta las estimaciones puntuales, sino que también modifica la precisión de manera variable según la covariable. La corrección explícita de errores de clasificación permite una estimación más robusta del intercepto (y por tanto de la prevalencia ajustada), aunque con posibles compromisos en la precisión de ciertos predictores individuales.

## 4.8 Discusión

Diversos estudios han demostrado que la estimación de la prevalencia de una enfermedad utilizando pruebas diagnósticas imperfectas puede conducir a sesgos sustanciales si no se corrige adecuadamente por la sensibilidad y especificidad de dichas pruebas. Esta necesidad de ajuste ha sido ampliamente documentada en la literatura, particularmente

en contextos donde la prevalencia varía entre poblaciones. En este sentido, Li y Fine (77) señalan que la sensibilidad, tradicionalmente considerada una propiedad intrínseca de la prueba, es un factor preponderante en las estimaciones de prevalencia, debido a factores como cambios en el espectro de severidad de los casos o variaciones en el contexto clínico del diagnóstico. En consecuencia, el ajuste por sensibilidad y especificidad se vuelve no solo metodológicamente relevante, sino también necesario para garantizar estimaciones válidas y comparables de prevalencia entre estudios o poblaciones heterogéneas.

La estimación precisa de la prevalencia ajustada por características individuales se vuelve particularmente relevante cuando el objetivo de un estudio no es solo obtener un valor promedio poblacional, sino también generar estimaciones estratificadas o específicas para subgrupos definidos por covariables demográficas o epidemiológicas. En este contexto, la implementación de modelos de regresión logística se posiciona como una herramienta óptima y flexible para abordar esta necesidad. Vansteelandt et al. (67) destacan que, cuando se dispone de información sobre covariables individuales, la extensión de los estimadores clásicos mediante modelos lineales generalizados permite incorporar directamente dichos factores en el análisis, mejorando así la precisión y validez de las estimaciones de prevalencia. Este enfoque no solo se adapta a distintas configuraciones de diseño muestral, sino que también resulta ventajoso en términos de eficiencia estadística y costo-beneficio al aprovechar de forma más efectiva la información recolectada junto con propuestas metodológicas que amplían su alcance en propósitos de estimaciones, según Villalobos (78).

Este estudio comparó distintos enfoques metodológicos para la estimación de prevalencias y asociaciones en presencia de errores de clasificación en el desenlace, aplicados a infecciones de transmisión sexual con prevalencias bajas y moderadas. Se evaluaron modelos de regresión logística estándar, el modelo propuesto por Liu et al. (59), y dos aproximaciones Bayesianas con y sin corrección explícita de mala clasificación de Valle et al. (60).

Es importante señalar que, si bien se observaron cambios relevantes en los coeficientes de algunas covariables entre los diferentes modelos, la magnitud y dirección de dichos cambios deben interpretarse en función de la distribución base de las variables en la población estudiada. Por ejemplo, coeficientes con valores bajos en el modelo estándar pueden mostrar grandes variaciones porcentuales en los modelos corregidos sin que esto necesariamente implique un cambio sustantivo en su efecto.

En cambio, el intercepto ($\beta_0$) mostró cambios consistentes entre modelos, lo que sugiere que las principales correcciones se reflejan en el componente basal del modelo,

es decir, en la estimación de la prevalencia cuando todas las covariables son cero. En este contexto, el intercepto corregido puede interpretarse como una aproximación a la prevalencia base ajustada por error de clasificación, proporcionando una medida más estable y robusta del desenlace en ausencia de factores de riesgo adicionales.

Estos hallazgos refuerzan la idea de que la corrección por error de clasificación impacta no solo en las asociaciones relativas, sino también en la estimación puntual de la prevalencia base. En particular, se observa que las estimaciones ajustadas difieren de manera sistemática respecto a las no ajustadas. Si bien no es posible afirmar con certeza que estas correcciones conduzcan necesariamente a un mayor acercamiento al verdadero valor de la prevalencia, sí es razonable considerar que representan una mejora metodológica significativa. En este sentido, los modelos corregidos pueden ofrecer estimaciones más coherentes con los supuestos epidemiológicos y con menor sesgo atribuible a errores de clasificación, lo cual resulta especialmente relevante en estudios poblacionales y sistemas de vigilancia epidemiológica.

Los resultados mostraron diferencias sustanciales en las estimaciones de prevalencia ajustada y en los coeficientes de regresión según el enfoque utilizado. En particular, el modelo Bayesiano con corrección por clasificación errónea (BEC) presentó un equilibrio favorable entre precisión y corrección del sesgo, produciendo estimaciones intermedias entre el modelo estándar y el modelo de Liu, con los intervalos de confianza más estrechos. Este hallazgo fue consistente tanto para VIH como para sífilis.

Sin embargo, deben considerarse cuidadosamente las implicaciones metodológicas y prácticas al seleccionar el enfoque estadístico. El modelo propuesto por Liu et al. (59), especialmente en su formulación más general, que estima simultáneamente las tasas de falsos positivos (FP) y falsos negativos (FN), incorpora un número sustancialmente mayor de parámetros que la regresión logística estándar. Esta complejidad se traduce en mayores requerimientos de tamaño muestral para lograr estabilidad en la estimación y garantizar la convergencia del algoritmo de puntuación de Fisher (Fisher scoring). Según las simulaciones presentadas en dicho trabajo, se requiere un tamaño de muestra mínimo del orden de 5000 observaciones para obtener estimaciones confiables de los parámetros, incluyendo las tasas de error de clasificación. Esta necesidad se vuelve aún más crítica cuando la prevalencia del desenlace es baja, ya que el número reducido de eventos informativos puede limitar severamente la capacidad del modelo para estimar de forma precisa los parámetros adicionales asociados a la clasificación errónea.

En este estudio, a pesar de contar con una muestra considerablemente grande ($n = 11452$), se observaron dificultades en términos de precisión, expresadas en amplios in-

tervalos de confianza y errores estándar elevados para varios coeficientes del modelo de Liu. Este fenómeno fue especialmente evidente en las covariables asociadas a condiciones de muy baja prevalencia, como la positividad al virus de la inmunodeficiencia humana (VIH), donde el número efectivo de casos fue relativamente limitado (1.4% del total). Estos hallazgos sugieren que, aun en contextos con tamaños muestrales grandes, el desempeño del modelo puede verse comprometido por la baja ocurrencia del evento, lo cual limita su aplicabilidad práctica en estudios de enfermedades infecciosas poco frecuentes.

Adicionalmente, el modelo de Liu mostró problemas de convergencia en escenarios con datos que no contienen ambos tipos de error de clasificación. Este aspecto ligado al planteamiento de la estructura del error puede comprometer la validez del modelo. Por el contrario, los modelos Bayesianos permiten incorporar información previa sobre la sensibilidad y especificidad, lo cual no solo mejora la estabilidad de las estimaciones en muestras más pequeñas, sino que permite una caracterización probabilística de la incertidumbre. En este estudio, esta ventaja se reflejó en mejoras sustanciales en la precisión del intercepto, directamente asociado a la prevalencia base corregida, así como en coeficientes clave.

Otra limitación del modelo de Liu radica en su supuesto de ausencia de error en los predictores, lo que puede ser restrictivo en aplicaciones reales. El marco Bayesiano, por su parte, ofrece flexibilidad para extender el modelo a escenarios más complejos, como el modelado jerárquico de tasas de error dependientes de covariables o la incorporación de errores de medición en los predictores.

Frente al modelamiento Bayesiano, cabe destacar la ventaja metodológica que representa la posibilidad de incorporar información previa sobre los coeficientes de regresión, lo cual permite regularizar las estimaciones y reducir la variabilidad asociada a muestras pequeñas. En este sentido, Newman et al. (70) propuso la formulación de distribuciones previas informativas basadas en la varianza de los coeficientes estandarizados y el número de covariables incluidas en el modelo, metodología implementada en este estudio. Esta aproximación resulta especialmente útil en diferentes contextos, sobre todo donde la estimación clásica tiende a producir coeficientes inestables y errores estándar excesivamente amplios. Sin embargo, el uso de distribuciones previas no informativas también tiene la posibilidad práctica de implementarse, como lo demuestran Gordóvil-Merino et al. (79), mediante un estudio de simulación, utilizando distribuciones previas débilmente informativas, las cuales también ofrecen una mayor estabilidad en la estimación de los parámetros en modelos de regresión logística aplicados a muestras pequeñas. Aunque los autores reconocen que este enfoque no resuelve completamente los problemas deriva-

dos de distribuciones asimétricas, subrayan que la inclusión de conocimiento previo, incluso mínimo, contribuye a mitigar inferencias extremas y a mejorar la regularización del modelo, lo cual es crucial en estudios con bajo número de observaciones o eventos raros. Así, el enfoque Bayesiano se posiciona como una alternativa metodológicamente sólida y flexible para situaciones donde las condiciones de los supuestos clásicos se ven comprometidas.

La estimación del sesgo asociado a los estimadores de prevalencia no pudo llevarse a cabo en este estudio, dado que las bases de datos utilizadas no contienen información sobre la prevalencia verdadera a nivel poblacional. Esta carencia imposibilita contrastar los valores estimados con el parámetro real, lo cual constituye un requisito fundamental para la evaluación rigurosa del sesgo.

En conjunto, los hallazgos respaldan el uso de modelos que consideren explícitamente la clasificación errónea en estudios de prevalencia y asociación basados en pruebas diagnósticas imperfectas. Si bien el modelo de Liu representa un avance sobre la regresión logística estándar al abordar el sesgo sistemático, sus limitaciones prácticas y teóricas lo hacen menos adecuado en ciertos contextos. Por el contrario, los modelos Bayesianos con corrección —como el evaluado en este estudio— se posicionan como una alternativa robusta y flexible, particularmente en escenarios de baja prevalencia, con errores diagnósticos conocidos o parcialmente conocidos, y necesidad de estimaciones precisas para la toma de decisiones en salud pública.

# Capítulo 5

# Conclusiones y Futuros Trabajos

## 5.1 Conclusiones

**Del Capitulo 2**

- Los métodos de estimación basados en la aplicación de un gold estándar a toda la muestra, como Agresti-Coull y Jeffreys, produjeron estimaciones similares ($0.81 \times 10^{-4}$ y $0.88 \times 10^{-4}$ respectivamente), con intervalos de confianza moderados, evidenciando buen desempeño en prevalencias bajas, como lo reporta la literatura, sin embargo pueden subestimar la incertidumbre si se ignora el esquema de muestro en fases cuando este se utilizó.

- El método de Lang, que integra la corrección de Rogan-Gladen en un enfoque frecuentista, presentó problemas de validez: la estimación puntual fue 0, y fue necesario truncar el intervalo inferior a 0 debido al incumplimiento de las condiciones de validez requeridas, específicamente $1 - \hat{Sp} \geq \widehat{AP}$. Esto resalta las limitaciones prácticas de este método en prevalencias bajas.

- El método Bayesiano de Flor, que incorpora explícitamente la incertidumbre en sensibilidad y especificidad, proporcionó estimaciones de prevalencia corregidas más coherentes ($0.16 \times 10^{-4}$) y con intervalos de credibilidad más estrechos, mejorando la precisión respecto a los métodos frecuentistas.

- Los métodos que respetaron el diseño de muestreo en dos fases, tanto en su versión frecuentista ($\text{Thomas}_{Fr}$, $0.94 \times 10^{-4}$) como bayesiana ($\text{Thomas}_{Bay}$, $1.04 \times 10^{-4}$), evidenciaron estimaciones más consistentes y plausibles, subrayando la importancia de modelar correctamente el proceso de selección en estudios de tamizaje.

- Se identificó que los métodos que no incorporan explícitamente la corrección por error de clasificación o el diseño de muestreo en fases tienden a subestimar o sobrestimar la prevalencia real, así como a generar intervalos de incertidumbre demasiado estrechos, proporcionando una falsa sensación de precisión.
- El enfoque Bayesiano mostró ventajas importantes sobre el enfoque frecuentista, particularmente en la restricción natural de la prevalencia al rango $[0, 1]$, en la incorporación de incertidumbre sobre los parámetros diagnósticos, y en la producción de intervalos de credibilidad más adecuados en presencia de prevalencias bajas.
- Finalmente, se resalta la necesidad de considerar en la práctica real las limitaciones inherentes a la implementación de métodos Bayesianos (requerimientos computacionales, necesidad de programación estadística avanzada) y los desafíos éticos y logísticos para garantizar muestras aleatorias, que motivan el uso de técnicas adicionales de ajuste como la post-estratificación.

**Del Capitulo 3**

- La aplicación de la post-estratificación mejoró notablemente las estimaciones de prevalencia en los tres escenarios estudiados, corrigiendo probablemente, en parte, el sesgo introducido por el muestreo no probabilístico y el test imperfecto.
- En el primer escenario (verificación completa con gold estándar), el ajuste por post-estratificación aumentó sustancialmente las estimaciones puntuales: de 0.0081% a 0.4621% en el método de Agresti–Coull y de 0.0088% a 0.2417% en el método Bayesiano. Esto evidencia que sin ajuste, las prevalencias resultaban severamente subestimadas debido a la falta de representatividad muestral.
- En el segundo escenario (uso de prueba diagnóstica imperfecta), los métodos frecuentistas como el de Lang subestimaron drásticamente la prevalencia (0%), mientras que el método Bayesiano de Flor corrigió esta subestimación, elevando la estimación a 0.2740% tras ajuste. Además, los intervalos sin ajuste fueron inusualmente cortos, generando una falsa sensación de precisión.
- El método frecuentista basado en el ajuste de Thomas, en el escenario de verificación parcial, redujo ligeramente la longitud del intervalo tras la post-estratificación (de 0.0149 a 0.0121), mientras que el enfoque Bayesiano mostró un incremento en la longitud del intervalo (de 0.0156 a 0.1025), reflejando su capacidad para capturar incertidumbre adicional proveniente del diseño de verificación parcial.

- El enfoque Bayesiano evidenció ventajas sustantivas: restringió naturalmente las estimaciones de prevalencia al rango permitido $[0, 1]$, modeló de forma explícita la incertidumbre asociada a las pruebas diagnósticas y a la post-estratificación, y generó inferencias más robustas frente a condiciones de prevalencia baja.
- Se confirmó que métodos frecuentistas basados en ajustes de máxima verosimilitud, aunque eficientes computacionalmente, pueden generar intervalos de confianza excesivamente optimistas (estrechos) si no se corrige adecuadamente por los sesgos de muestreo y errores de clasificación.
- A nivel de diseño, se corroboró que los métodos basados en la integración de post-estratificación son necesarios para garantizar la validez de las inferencias en muestras no probabilísticas, situación común en estudios de salud pública.
- Como limitaciones, se reconoce que la implementación de modelos Bayesianos depende críticamente de la disponibilidad y calidad de la información previa, y que su aplicabilidad puede verse restringida por necesidades computacionales avanzadas.

**Del Capitulo 4**

- Se evidenció que la corrección por clasificación errónea en la estimación de prevalencias es crítica para evitar sesgos sustanciales, particularmente en enfermedades infecciosas de baja prevalencia como VIH y sífilis.
- El modelo Bayesiano con corrección explícita por clasificación errónea (BEC) presentó un desempeño superior en términos de precisión y estabilidad, al producir estimaciones intermedias entre el modelo estándar (STD) y el modelo de Liu, pero con los intervalos de confianza más estrechos.
- La estimación de prevalencia para VIH varió de un 1.39% (cruda) a 1.00% (BEC), mientras que para sífilis la estimación se ajustó de 5.67% (cruda) a 4.14% (BEC), resultados que no solo reflejan la corrección por mala clasificación, sino que permitieron un ajuste por covariables incluidas en los modelos de estimación, representando una ventaja adicional al epidemiólogo ya que obtiene mediciones con mayor validez en su reporte.
- Los modelos corregidos modificaron sustancialmente los coeficientes de regresión asociados a variables claves. Por ejemplo, en el análisis de VIH, el coeficiente para

trabajador sexual ($\beta_8$) cambió de positivo a negativo en el modelo BEC, lo que refleja una sensibilidad importante a la corrección por error de clasificación.

- El modelo de Liu, aunque conceptualmente sólido al corregir simultáneamente sensibilidad y especificidad, presentó mayores dificultades prácticas: errores estándar elevados, amplios intervalos de confianza y problemas de convergencia en escenarios de baja prevalencia, aun contando con muestras de tamaño considerable ($n = 11452$).
- El enfoque Bayesiano mostró ventajas notables al incorporar información previa, mejorar la precisión de parámetros claves (como el intercepto, directamente relacionado con la prevalencia ajustada) y permitir la estabilidad de los modelos incluso en presencia de bajas tasas de eventos.
- Se destaca que en modelos Bayesianos, la correcta especificación de distribuciones previas, incluso débiles o parcialmente informativos, puede mejorar notablemente la regularización de las estimaciones, particularmente en muestras pequeñas o con desenlaces raros.
- Las comparaciones metodológicas realizadas confirman que en escenarios de clasificación imperfecta, baja prevalencia y muestras moderadas, los modelos Bayesianos con corrección implícita constituyen una alternativa más robusta y metodológicamente adecuada que los enfoques tradicionales.

## 5.2 Futuros Trabajos

Algunas líneas de trabajo interesantes para el futuro incluyen enfoques para resolver las limitaciones actuales de los métodos presentados y futuras extensiones de las metodologías desarrolladas. Parece obligatorio después del trabajo realizado, el llevar a cabo experimentos exhaustivos de simulación para obtener bases de datos en las que, partiendo de parámetros conocidos (sensibilidad, especificidad y prevalencia) se puedan estimar a fondo los sesgos de las estimaciones de la prevalencia en escenarios muy diferentes. Tales estudios de simulación permitirían obtener reglas automáticas de aplicación de los estimadores y de los intervalos en problemas específicos de aplicación.

**Idea 1:** La posible extensión de otros métodos de corrección para muestras no probabilísticas en estimaciones con test diagnostico imperfecto, en especial técnicas de machine learning que estiman la probabilidad de selección del sujeto.

**Idea 2:** Desarrollo de un modelo de regresión logística que incorpora parámetros desconocidos de sensibilidad y especificidad con el enfoque a un entorno de clasificación no paramétrica mediante Procesos Gaussianos (GP).

**Idea 3:** En el contexto enfermedades infecciosas, la implementación de la serovigilancia (Sero-encuestas) se lleva acabo la aplicación de paneles múltiples de enfermedades basado en un kit de detección por test rapidos (imperfectos), se hace necesario métodos de estimación que integren conjuntamente el uso de pruebas correlacionadas.

**Idea 4:** Evalaución del impacto de la corrección del error de clasificación y muestras no probabilísticas en contextos de estimación de la fuerza de infección a partir de seroprevalencias.

**Idea 5:** Desarrollo de Software que unifique las diferentes técnicas y escenarios para la estimación de prevalencias tanto con datos agrupados como individuales, que posea interoperatividad con otros paquetes del ecosistema Tidyverse en R Software.

**Idea 6:** Aplicar los estimadores con menos sesgo y más precisión a problemas concretos de estimación de la prevalencia a partir de tests diagnósticos falibles recogidos a partir de encuestas poblacionales

# Anexos A

# Apendice para Capitulo 2

## A.1 Código en R software y Stan

```
#### caluclo de Rogan gladen con intervalo de agresti-coull
   ajustado por Se y Sp conocidas

rg.agresti <- function(x, n, se, sp, conf = 0.95){
  z <- qnorm(1-(1-conf)/2)
  np = n+z^2
  ap = x/n
  AP = (n*ap+(z^2)/2)/(np)
  P_rg <- (AP+sp-1)/(se+sp-1)
  if(P_rg<0){
    P_rg <- 0
  }
  else(P_rg = P_rg)
  ee <- z*1/(se+sp-1)*sqrt(AP*(1-AP)/np)
  if(P_rg == 0){
    lower <- 0
    upper <- P_rg+ee
  }
  else(lower <-  P_rg-ee)
  upper <-  P_rg+ee
  r <- tibble(method = "Lang",x = x, n = n, mean = P_rg, lower =
      lower, upper = upper)
```

```
  return(r)
}
## Resultados Prevalencia de Quilomicronemia familiar
## Doctorado en Estad stica Matematica y Aplicada
## Departamento de Estad stica e Investigaci n Operativa
## Universidad de Granada

## Autor: Jorge Mario Estrada Alvarez

######################  resutlados tesis
pacman::p_load(tidyverse, rio, binom, rstan, coda)

dataset <- import("analisis_tesis/dataset.rds")

# casos por test falible
falible <- sum(dataset$pos_test)

# casos por gold estandar
mutados <- sum(dataset$pos_ge)

# muestra de estudio
n <- sum(dataset$n)

# esitmaciones asumiendo solo el resultado por Gold estandar

# Intervalo de Agresti-Coull (frecuentista)

tibble(binom.agresti.coull(x = mutados, n = n,conf.level = 0.95)
   )

# Intervalo Jeffrey (Bayesiano)

tibble(binom.bayes(x = mutados, n = n, conf.level = 0.95,
                   prior.shape1 = 0.5,prior.shape2 = 0.5, type =
                       "central")) %>%
  select(-shape1, -shape2, -sig)
```

```
# ESTIMACIONES DE PREVALENCIA CON UN SOLO TEST IMPERFECTO

# Intervalo de Lang et al. (Agresti-Coull modificado para
   corregir por Se y Sp) Frecuentista

source("analisis_tesis/Funciones/IC_agresti_se_sp.R")
se <- 0.88 # Sensibilidad
sp <- 0.85 # Especificidad
rg.agresti(x = falible, n = n, se = se, sp = sp)

# Intervalo Bayesiano (Flor et al)
# Cargar la biblioteca necesaria
library(rstan)

# Configuraci n para mejorar el rendimiento en RStan
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

# Especificar la ruta al archivo .stan
stan_file <- "analisis_tesis/Funciones/bayesiano_se_sp.stan"

# Datos
data <- list(
  n = 74145,          # Tamano de muestra
  x = 18,           # Casos positivos observados
  Se = 0.88,        # Sensibilidad conocida
  Sp = 0.85         # Especificidad conocida
)

# Compilar el modelo
stan_model <- stan_model(file = stan_file)

# Ajustar el modelo
fit <- sampling(stan_model, data = data, iter = 2000, chains =
   4, seed = 123)

# Resumen de los resultados
```

```
print(fit, pars = c("pi", "AP"))

# Visualizar los resultados
posterior_samples <- extract(fit)$pi
hist(posterior_samples, breaks = 20, main = "Distribuci n
   Posterior de   ", xlab = "  ")


tibble(method = "Bayesiano - Flor", x = falible, n = n, mean =
   mean(posterior_samples),
               lower = quantile(posterior_samples, probs =
                  0.0275), upper = quantile(posterior_samples,
                  probs = 0.975))

# ESTIMACIONES DE PREVALENCIA ASUMIENDO ESQUEMA EN FASES:

# FASE 1 - TEST IMPERFECTO
# FASE 2 - VERIFICACION POR GOLD ESTANDAR UNICAMENTE ENTRE
   POSITIVOS

# Intervalo frecuentista por Thomas et al.
source("analisis_tesis/Funciones/
   thomas_frecuentista_solo_positivos.R")

# N0 = numero de test negativos
# N11 = nuemro de verdaderos positivos
# N01 = numero de falsos positivos
# theta = Sensibilidad del test
# phi = Especificidad
# alpha = error tipo I

p <- p.scenario1(N0 = n-falible ,N01 = falible-mutados ,N11 =
   mutados,theta = 0.88, phi = 0.85, alpha = 0.05)

tibble(method = "Thomas-frecuentista", x = 6, n = 74145, mean =
   p$estimate, lower = p$CI[1], upper = p$CI[2])
```

```
# Intervalo bayesiano por Thomas et al.

### Estimaci n bayesiana
# Scenario 1
N0 <- 74127 # Number of test negatives
N1 <- 18 # Number of test positives
N01 <- 12 # Number of false positives
N11 <- N1-N01 # Number of true positives
n <- N0+N1 # Total number tested (fixed)

set.seed(134623)
# Sensitivity and specificity
# sensitivity
theta <- 47/53
# specificity
phi <- 44/52

# estimacion

shims_data <- list(n=n, N1=N1, N11=N11, theta = theta, phi = phi
   )
shims_fit1 <- stan(file="analisis_tesis/Funciones/bayes_thomas2.
   stan", data=shims_data, iter=2000, chains=5, warmup=500)

print(shims_fit1, pars=c("p"), digits_summary=10)
traceplot(shims_fit1, pars = c("p"))
posterior_samples <- extract(shims_fit1)$p
hist(posterior_samples, breaks = 30, main = "Distribuci n
   Posterior de p", xlab = "p")

prov1 <- tibble(method = "Thomas - bayesiano", x = 6, n = 74145,
    mean = median(posterior_samples),
                lower = quantile(posterior_samples, probs =
                   0.0275), upper = quantile(posterior_samples,
                   probs = 0.975))

# Instalar paquetes si es necesario
```

```
# install.packages(c("ggplot2", "extrafont"))

library(ggplot2)

# Datos
df <- data.frame(
  method = c("Agresti-Coull", "Jeffrey", "Lang", "Flor",
             "Thomas-frecuentista", "Thomas-bayesiano"),
  mean = c(8.09225e-5, 8.76649e-5, 0, 1.56934e-5, 9.42223e-5,
     1.04342e-4),
  lower = c(3.2434e-5, 2.76472e-5, 0, 3.90397e-7, 1.97449e-5,
     4.47865e-5),
  upper = c(1.8121e-4, 1.56006e-4, 1.6159e-4, 5.84583e-5, 1.687e
     -4, 2.00906e-4),
  type = c("Frecuentista", "Bayesiano", "Frecuentista", "
     Bayesiano", "Frecuentista", "Bayesiano")
)

# Ajustar escala (por 10,000)
df$mean <- df$mean * 10000
df$lower <- df$lower * 10000
df$upper <- df$upper * 10000

# Ordenar factores
df$method <- factor(df$method, levels = rev(df$method))

# Crear el grafico
p <- ggplot(df, aes(x = mean, y = method, color = type)) +
  geom_point(size = 2) +
  geom_errorbarh(aes(xmin = lower, xmax = upper), height = 0.2,
     linewidth = 0.6) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "
     gray50") +
  scale_color_manual(values = c("Frecuentista" = "blue", "
     Bayesiano" = "red")) +
  labs(x = "Prevalencia estimada (   10.000 hab)", y = NULL,
     color = "Tipo de m todo") +
```

```
  theme_classic(base_family = "serif") +
  theme(
    axis.text = element_text(size = 11, family = "serif"),
    axis.title = element_text(size = 13, family = "serif"),
    legend.text = element_text(size = 11, family = "serif"),
    legend.title = element_text(size = 12, family = "serif"),
    legend.position = "right",
    panel.grid.major.x = element_line(color = "gray90", linetype
        = "dashed"),
    panel.grid.minor.x = element_blank()
  )

# Mostrar en pantalla
print(p)

# Exportar como PDF de alta calidad
ggsave("estimaciones_prevalencia.pdf", plot = p, width = 7,
   height = 5, dpi = 600)
```

Listing A.1: Código R Estiamciones frecuentistas y Bayesianas para estudio de Prevalencia de Quilomicronemia

# Anexos B

# Apendice para Capitulo 3

## B.1 Codigo en R software y Stan

```
# Metodo Frecuentista sin post-estratificacion

pacman::p_load(tidyverse, rio, binom, rstan, coda)

dataset <- import("analisis_tesis/dataset.rds")

# casos por test falible
falible <- sum(dataset$pos_test)

# casos por gold estandar
mutados <- sum(dataset$pos_ge)

# muestra de estudio
n <- sum(dataset$n)

# poblacion >= 18 a os
N <- sum(dataset$CANTIDAD)

######## ESCENARIO 1: VERIFICACION GOLD ESTANDAR

# estimacion con gold estandar verificado frecuentista
```

```
results <- tibble(binom.agresti.coull(x = mutados, n = n,conf.
   level = 0.95))
# estimacion con gold estandar verificado bayesiano
results <- tibble(binom.agresti.coull(x = mutados, n = n,conf.
   level = 0.95))

bayes <- tibble(binom.bayes(x = mutados, n = n,conf.level =
   0.95)) %>%
  select(-shape1, -shape2, -sig)

results <- add_case(results,bayes)
results[,"Escenario"] <- c("Escenario 1","Escenario 1")
results <- results %>% relocate(Escenario,.before = method)


######### ESCENARIO 2: UN SOLO TEST FALIBLE SE Y SP CONOCIDAS

###### estimacion con test falible Se y sp conocidas
   frecuetntista
source("analisis_tesis/Funciones/IC_agresti_se_sp.R")
prov <- rg.agresti(x = 18, n = 74145, 0.88, 0.85)

###### estimacion con test falible Se y sp conocidas bayesiano
source("analisis_tesis/Funciones/estimacion_bayes_se_sp.R")

prov1 <- tibble(method = "Bayesiano - Flor", x = 18, n = 74145,
   mean = mean(posterior_samples),
                lower = quantile(posterior_samples, probs =
                   0.0275), upper = quantile(posterior_samples,
                   probs = 0.975))

falible <- add_case(prov, prov1)

falible[,"Escenario"] <- c("Escenario 2","Escenario 2")
falible <- falible %>%
  relocate(Escenario)
```

```
  results <- results %>%
  add_case(falible)

############### ESCENARIO 3: UN TEST FALIBLE Y VERIFICACION
   PARCIAL POR GOLD ESTANDAR (SOLO POSITIVOS)
  ### publicacion de Thomas.
  ### Estimacion frecuentista
  source("analisis_tesis/Funciones/
     thomas_frecuentista_solo_positivos.R")

  # N0 = number of test negatives
  # N11 = number of true positives
  # N01 = number of false positives
  # theta = known sensitivity of test
  # phi = known specificity of test
  # alpha = type 1 error rate (default=0.05)

p <- p.scenario1(N0 = 74127 ,N01 =12 ,N11 = 6,theta = 0.88, phi
   = 0.85, alpha = 0.05)

thomas <- tibble(method = "Thomas-frecuentista", x = 6, n =
   74145, mean = p$estimate, lower = p$CI[1], upper = p$CI[2])

### Estimaci n bayesiana
# Scenario 1
N0 <- 74127 # Number of test negatives
N1 <- 18 # Number of test positives
N01 <- 12 # Number of false positives
N11 <- N1-N01 # Number of true positives
n <- N0+N1 # Total number tested (fixed)

set.seed(134623)
# Sensitivity and specificity
# sensitivity
theta <- 47/53
# specificity
phi <- 44/52
```

```
# estimacion

shims_data <- list(n=n, N1=N1, N11=N11, theta = theta, phi = phi
   )
shims_fit1 <- stan(file="analisis_tesis/Funciones/bayes_thomas2.
   stan", data=shims_data, iter=2000, chains=5, warmup=500)

print(shims_fit1, pars=c("p"), digits_summary=10)
traceplot(shims_fit1, pars = c("p"))
posterior_samples <- extract(shims_fit1)$p
hist(posterior_samples, breaks = 30, main = "Distribucion
   Posterior de p", xlab = "p")

prov1 <- tibble(method = "Thomas - bayesiano", x = 6, n = 74145,
    mean = median(posterior_samples),
                lower = quantile(posterior_samples, probs =
                   0.0275), upper = quantile(posterior_samples,
                   probs = 0.975))

thomas <- thomas %>%
  add_case(prov1)

thomas[,"Escenario"] <- c("Escenario 3","Escenario 3")
thomas <- thomas %>%
  relocate(Escenario)

results <- results %>%
  add_case(thomas)

## resuyltados sin postestratificacion para todos los metodos
results

saveRDS(results,"analisis_tesis/results.rds")
```

Listing B.1: Código R Estiamciones frecuentistas sin Post-estratificación

```r
# Cargar la biblioteca necesaria
library(rstan)

# ConfiguraciOn para mejorar el rendimiento en RStan
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

# Especificar la ruta al archivo .stan
stan_file <- "analisis_tesis/Funciones/bayesiano_se_sp.stan"

# Datos
data <- list(
  n = 74145,          # Tamano de muestra
  x = 18,           # Casos positivos observados
  Se = 0.88,        # Sensibilidad conocida
  Sp = 0.85         # Especificidad conocida
)

# Compilar el modelo
stan_model <- stan_model(file = stan_file)

# Ajustar el modelo
fit <- sampling(stan_model, data = data, iter = 2000, chains =
   4, seed = 123)

# Resumen de los resultados
print(fit, pars = c("pi", "AP"))

# Visualizar los resultados
traceplot(fit, pars = c("pi"))
posterior_samples <- extract(fit)$pi
hist(posterior_samples, breaks = 20, main = "Distribucion
   Posterior de   ", xlab = "  ")
```

Listing B.2: Código R Estimaciones Bayesianas sin Post-estratificación

```
data {
  int<lower=0> n;       // Tama o de la muestra
```

```
  int<lower=0> x;        // Resultados positivos observados
  real<lower=0, upper=1> Se; // Sensibilidad conocida
  real<lower=0, upper=1> Sp; // Especificidad conocida
}
parameters {
  real<lower=0, upper=1> pi; // Prevalencia real
}
model {
  real AP; // Prevalencia aparente
  // Prior no informativo para pi
  pi ~ beta(1, 1);

  // Relacion entre AP, Se, Sp y pi
  AP = pi * Se + (1 - pi) * (1 - Sp);

  // Verosimilitud
  x ~ binomial(n, AP);
}
generated quantities {
  real AP; // Prevalencia aparente
  AP = pi * Se + (1 - pi) * (1 - Sp);
}
```

Listing B.3: Código de Stan para Estimaciones Bayesianas sin Post-estratificación

```
pacman::p_load(tidyverse, rio, binom, rstan)

# datos
dataset <- import("analisis_tesis/dataset.rds")

### creacion de weights
dataset <- dataset %>%
  mutate(w_h = CANTIDAD/sum(CANTIDAD))


######### ESCENARIO 1: VERIFICACION GOLD ESTANDAR

# post-estratificado con  agresti-coull
```

```r
# prevalecnia para cada estrato
casos_w <- dataset %>%
  pull(pos_ge)
# weights
w_h <- dataset %>%
  pull(w_h)
# n por estrato

n_h <- dataset %>%
  pull(n)

# calculo de p_level

source("analisis_tesis/Funciones/agresti_postestratificado.R")
results2 <- agresti_level(X_w = casos_w, W_h = w_h, n_h = n_h)

results2


### estimaci n postestratificada por bayesiano
# por el paquete binom
#binom.agresti.coull(x = casos_w, n = n,conf.level = 0.95)

# programado por Rstan
# Configuraci n para mejorar el rendimiento en RStan
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

# Especificar la ruta al archivo .stan
stan_file <- "analisis_tesis/Funciones/
   bayesiano_postestratificado.stan"

# Datos simulados
data <- list(
  n = 74145,          # Tama o de muestra
  x = 6           # Casos positivos observados
```

```
  )

# Compilar el modelo
stan_model <- stan_model(file = stan_file)
fit <- sampling(stan_model, data = data, iter = 2000, chains =
   4, seed = 123)

# Ajustar el modelo
posterior_samples <- data.frame(matrix(ncol = length(n_h), nrow
   = 4000))

for(i in 1:length(n_h)){
  fit <- sampling(stan_model,
                  data = list(n = n_h[i], x = casos_w[i]),
                  iter = 2000,
                  chains = 4,
                  seed = 123)
  posterior_samples[,i] <- extract(fit)$p*w_h[i]
}

# adicion de resultados bayesiano postestratificado
results2 <- results2 %>%
  add_case(tibble(adjust_method = "Bayes", mean = median(rowSums
     (posterior_samples)),
       lower = quantile(rowSums(posterior_samples), probs =
          0.0275),
       upper = quantile(rowSums(posterior_samples), probs =
          0.975)))
saveRDS(posterior_samples,"analisis_tesis/posterior_agresti.rds"
   )

######### ESCENARIO 2: UN SOLO TEST FALIBLE SE Y SP CONOCIDAS

#### agresti-coull interval
source("analisis_tesis/Funciones/Ic_agresti_se_sp_post.R")
positivos_w <- dataset %>%
  pull(pos_test)
```

```
sum(positivos_w)

escenario2 <- data.frame(matrix(ncol = 3, nrow = length(n_h),))

for(i in 1:length(n_h)){

  escenario2[i,] <- rg.agresti_post(x = positivos_w[i],
                                    n = n_h[i], se = 0.88, sp =
                                       0.85)
}

P <- sum(escenario2$X1*w_h)
var <- sum((w_h^2)*escenario2$X3)


results2 <- results2 %>%
  add_case(
tibble(adjust_method = "Lang",
       mean = P,
       lower = 0,
       upper = P + (1.96*sqrt(var))))

#### intervalo bayesiano con se y sp conocidas

# ConfiguraciOn para mejorar el rendimiento en RStan
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

# Especificar la ruta al archivo .stan
stan_file <- "analisis_tesis/Funciones/bayesiano_se_sp.stan"

# Compilar el modelo
stan_model <- stan_model(file = stan_file)
#fit <- sampling(stan_model, data = data, iter = 2000, chains =
   4, seed = 123)
```

```
# Ajustar el modelo
posterior_postestr <- data.frame(matrix(ncol = length(n_h), nrow
    = 4000))

for(i in 1:length(n_h)){
  fit <- sampling(stan_model,
                  data = list(n = n_h[i], x = positivos_w[i], Se
                      = 0.88, Sp = 0.85),
                  iter = 2000,
                  chains = 4,
                  seed = 123)
  posterior_postestr[,i] <- extract(fit)$pi*w_h[i]
}

# adiciOn de resultados bayesiano ajustado Se y Sp
   postestratificado
results2 <- results2 %>%
  add_case(
tibble(adjust_method = "Bayesiano - Flor", mean = median(rowSums
   (posterior_postestr)),
                  lower = quantile(rowSums(posterior_postestr),
                     probs = 0.0275),
        upper = quantile(rowSums(posterior_postestr), probs =
           0.975)))


saveRDS(posterior_postestr,"analisis_tesis/
   posterior_flor_estratificado.rds")

################ ESCENARIO 3: UN TEST FALIBLE Y VERIFICACION
   PARCIAL POR GOLD ESTANDAR (SOLO POSITIVOS)
### publicacion de Thomas.
### EstimaciOn frecuentista
source("analisis_tesis/Funciones/
   thomas_frecuentista_solo_positivos_estratificado.R")

# para estratificar
```

```
# N0 = number of test negatives
# N11 = number of true positives
# N01 = number of false positives

N0_h <- dataset$n-dataset$pos_test
N11_h <- casos_w
N01_h <- dataset$pos_test - casos_w

df <- data.frame(
  p_h = rep(NA, 136),    # Inicializar la columna 'p_h' con
     valores NA
  sigma2 = rep(NA, 136)  # Inicializar la columna 'sigma2' con
     valores NA
)

for(i in 1:length(N0_h)){

  p <- p.scenario_estrato(N0 = N0_h[i] ,N01 = N01_h[i] ,N11 =
     N11_h[i],theta = 0.88, phi = 0.85)

  df[i,"p_h"] <- p[1]
  df[i,"sigma2"] <- p[2]
}

P <- sum(df$p_h*w_h)

var <- sum((w_h^2) * df$sigma2)

results2 <- results2 %>%
  add_case(
tibble(adjust_method = "Thomas-frecuentista",
       mean = P,
       lower = P-1.96*sqrt(var),
       upper = P+1.96*sqrt(var))
)

### Estimaci n Bayesiana por thomas et al.
```

```
# Configuraci n para mejorar el rendimiento en RStan
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

# Datos estratificados para estiamcion bayesiana con
   verificacion solo en positivos
n_h # Fixed total number tested
positivos_w # Number of test positives
N11_h # Number of true positives

# ciclo para estratificacion de estimados bayesianos
# dataframe para guardar resultados de p

posterior_thomas_estrati <- data.frame(matrix(ncol = length(n_h)
   , nrow = 4000))

for(i in 1:length(n_h)){
  shims_fit1 <- stan(file="analisis_tesis/Funciones/
     bayes_thomas2.stan",
                     data=list(n = n_h[i], N1 = positivos_w[i],
                        N11 = N11_h[i], theta = 0.88, phi = 0.85)
                        ,
                     iter=2000,
                     chains=4, warmup=1000,seed = 123)
  posterior_thomas_estrati[,i] <- extract(shims_fit1)$p*w_h[i]
}

results2 <- results2 %>%
  add_case(
    tibble(adjust_method = "Thomas - bayesiano", mean = median(
       rowSums(posterior_thomas_estrati)),
           lower = quantile(rowSums(posterior_thomas_estrati),
              probs = 0.0275),
           upper = quantile(rowSums(posterior_thomas_estrati),
              probs = 0.975)))
```

```
saveRDS(posterior_thomas_estrati,"analisis_tesis/
   posterior_thomas_estratificado.rds") ## posterior de las
   estratificadas

# guardado de resultados finales de estratificados
saveRDS(results2,"analisis_tesis/results2.rds")

finales <- results %>%
  left_join(results2, by = c("method"="adjust_method"))

finales <- finales %>%
  mutate(Long_sin = (upper.x - lower.x)*100,
         Long_con = (upper.y - lower.y)*100)

# guardado de resultados definitivos consolidados

saveRDS(finales,"datos dx definitivos/analisis_tesis/finales.rds
   ")

export(finales,"datos dx definitivos/analisis_tesis/finales.xlsx
   ")
```

Listing B.4: Codigo de R y Stan para Estimaciones frecuentistas y Bayesianas con correccion por Post-estratificacion

```
###### modificaicon de thomas frecuentista para estimacion post-
   estratificada

# scenario 1 estimator
p.scenario_estrato <- function(N0,N01,N11,theta,phi,alpha=0.05){
  # N0 = number of test negatives
  # N11 = number of true positives
  # N01 = number of false positives
  # theta = known sensitivity of test
  # phi = known specificity of test
  # alpha = type 1 error rate (default=0.05)
  # If two.sided = TRUE, the point estimate and a two-sided
```

```
    confidence interval are returned. If two.sided = FALSE, a
    one-sided upper-confidence limit is returned, and theta and
    phi are interpreted as lower bounds on the sensitivity and
    specificity respectively.
  # The function returns a list containing the point estimate
    and either its (1-alpha/2)100% confidence interval (CI) or
    its (1-alpha)100% upper confidence limit (UCL).

  # Calculate the maximum likelihood estimate of the prevalence
  A <- -(1-theta-phi)*(N0+N01+N11)
  B <- (1-theta-phi)*(N0+N11)-phi*(N11+N01)
  C <- N11*phi
  p.est1 <- (-B-sqrt(B^2-4*A*C))/(2*A)
  # Calculate the asymptotic variance of the estimate
  I <- ((1-theta-phi)^2/((1-theta)*p.est1+phi*(1-p.est1))+theta/
     p.est1+(1-phi)/(1-p.est1))*(N0+N11+N01)
  sigma2 <- 1/I

  return(c(p.est1, sigma2))
}
```

Listing B.5: Codigo de R para estimaciones bayesianas en solo positivos acorde a thomas et al.

```
data {
  // Data
  int<lower=0> n; // Fixed total number tested
  int<lower=0, upper=n> N1; // Number of test positives
  int<lower=0, upper=N1> N11; // Number of true positives
  // Hyperparameters
  real<lower=0, upper=1> theta; // Sensibilidad conocida
  real<lower=0, upper=1> phi; // Especificidad conocida
  //real theta; // Previous estimate of psi_theta
  // real<lower=0> sigma_theta; // Standard error of previous
     estimate of psi_theta
  // real mu_phi; // Previous estimate of psi_phi
  // real<lower=0> sigma_phi; // Standard error of previous
```

```
    estimate of psi_phi
}
parameters {
  real<lower=0, upper=1> p; // Prevalence
  //real psi_theta; // Log odds of testing positive for true
    positives
  //real psi_phi; // Log odds of testing negative for true
    negatives
}
model {
  p ~ uniform(0, 1);
  N1 ~ binomial(n, p*theta + (1-p)*(1-phi));
  N11 ~ binomial(N1, p*theta/(p*theta + (1-p)*(1-phi)));
}
```

Listing B.6: Codigo de Stan para estimaciones bayesiana estratificado acorde a Thomas et al.

# Anexos C

# Apendice para Capitulo 4

## C.1 Codigo en R software

```
install.packages("pacman")

pacman::p_load(logistic4p, tidyverse, rio, binom, gtsummary,
   janitor,
               fastDummies, marginaleffects)


vih <- readRDS("vih.rds")
# combinaciones para marginalizacion estandarizada

new_data <- vih %>%
  select(edad_gr,sex,reactivo_sif1,reactivo_hepb,tip_pob_HSH,
         tip_pob_LGTBIQ, `tip_pob_OTRAS POBLACIONES`, `
            tip_pob_TRABAJADOR SEXUAL`) %>%
 datagrid(newdata = new_data)

######################### Se Sp de los test utilizados
   ##################

# PRUEBA VIH 1   (ANTICUERPOS Y ANTIGENO) Marca determine Early
   detect 4a generacion
# limite inferior del IC95%
# sensibilidad: 0.995
```

```
# especificidad:  0.994

# PRUEBA VIH 2   (ANTICUERPOS Y ANTIGENO) Bioline HIV 1/2 3.0  3
   a generacion

# sensibilidad: 0.98
# especificidad:  .989


####### estimacion de prevalencia de VIH aparente con ro y r1
   desconocido #######

# Calculo de la Se y Sp de los test combinados, teniendo que si
   Test A es negativo
# el test B no requiere ser aplicado.
# el primero debe dar positivo para aplicar el segundo (ambos
   positivos definen un caso)

# ####### Expresiones: ##############################
# sensitivity: (A)sen x (B)sen                      #
# specificity: (A)spec + [1 - (A)spec] x (B)spec    #
#####################################################

 se_1 = 0.995
 sp_1 = 0.994
 se_2 = 0.980
 sp_2 = 0.989

# sensibilidades y especificidad combinadas para test vih
se <- se_1*se_2
sp <- sp_1+(1 - sp_1)*sp_2
# calculo de r0 y r1

r0 <- 1-sp # fp
r1 <- 1-se # fn
```

```
##### metodo de Liu et cols. ESTIAMCION DE PREVALENCIA AJUSTADA
   Y AJUSTADA CORREGIDA

x <- vih[,c("edad_gr","sex","reactivo_sif1","reactivo_hepb",
            "tip_pob_HSH", "tip_pob_LGTBIQ", "tip_pob_OTRAS
               POBLACIONES","tip_pob_TRABAJADOR SEXUAL")] #
               predictors
y <- vih["vih_1"]      # outcome

# modelo Liu con fp = fp  SI converge
mod_vih1_eq <- logistic4p(x, y,model='equal', max.iter = 1000,
   detail = TRUE)

# fp y fn con valores iniciales no converge
mod_vih1_fpfn <- logistic4p(x, y, initial = c(r0,r1,
   model_vih1_std$coefficients),model='fp.fn', max.iter = 1000,
   detail = TRUE)


print(mod_vih1_eq) # trabajo con este para las estimaciones de
   prevalencia

fp <- 0  # por modelo liu fp = fn
fn <- 0

saveRDS(mod_vih1_eq, "Nuevos_resultados/Modelos/mod_vih1_eq.rds"
   )

# crear funcion de prediccion para el modelo de Liu
source("function_predict.R")

### obtener coeficientes del modelo de Liu

coef <- mod_vih1_eq$estimates[,1][-1]

# guarda coeficientes estiamdos
```

```
coef_models_vih <- data.frame(coef) %>%
  rename(modvih1_liu_eq = coef)

saveRDS(coef_models_vih, "Nuevos_resultados/Resultados/
   coef_models_vih.rds")

######### calculo de prevalencia global y varianza del estimador
    ########

# para el caso de modelo de Liu al no ser un objeto tipo modelo
   libreria
# marginal effects no lo acepta entonces hacemos un modelo dummy

# Ajustamos un modelo log stico dummy

modelo_dummy <- glm(vih_1 ~ edad_gr+sex+reactivo_sif1+
                      reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`
                         tip_pob_OTRAS POBLACIONES`+
                      `tip_pob_TRABAJADOR SEXUAL`, family = "
                         binomial", data = vih)

# Reemplazamos los coeficientes con los estimados con el Modelo
   de Liu

modelo_dummy$coefficients <- coef
prev_global <- predictions(modelo_dummy, newdata = new_data) %>%
  select(2,4:5)

#### intervalo de prevalencia aajustado por ro y r1

prevalencias <- tibble(Disease = "VIH 1 - Liu.eq",
                       prev_global,
       Se = 1-fn, Sp = 1-fp) %>%
  mutate(P_ajustada = (estimate-fp)/(1-fn-fp),
         lower_adj = (conf.low-fp)/(1-fn-fp),
         upper_adj = (conf.high-fp)/(1-fn-fp))
```

```
saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")

############## RESULTADOS POR  REGRESION LOGISITCA ESTANDAR
   ###############

mod_vih1_std <- glm(vih_1 ~ edad_gr+sex+reactivo_sif1+
      reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`tip_pob_OTRAS
         POBLACIONES`+
        `tip_pob_TRABAJADOR SEXUAL`, family = "binomial", data =
            vih)


summary(mod_vih1_std)

saveRDS(mod_vih1_std,"Nuevos_resultados/Modelos/mod_vih1_std.rds
   ")

### obtener coeficientes del Modelo estandar

coef <- mod_vih1_std$coefficients

# guarda coeficientes estiamdos

coef_models_vih <- add_column(coef_models_vih) %>%
data.frame(coef) %>%
  rename(modvih1_std = coef)

saveRDS(coef_models_vih, "Nuevos_resultados/Resultados/
   coef_models_vih.rds")

# calculo de prevalencia global con modelo regresion logistica

prev_global <- predictions(mod_vih1_std, newdata = new_data) %>%
  select(2,4:5)
```

```
#### intervalo de prevalencia aparente global y ajustada por Se
   y Sp REGRESION LOGISTICA
#### ESTANDAR

prevalencias <- prevalencias %>%
  add_case(tibble(Disease = "VIH 1-STD",
                  prev_global,
                  Se = se, Sp = sp) %>%
  mutate(P_ajustada = estimate-(1-Sp)/(1-(1-Se)-(1-Sp)),
         lower_adj = conf.low-(1-Sp)/(1-(1-Se)-(1-Sp)), #
            corregida por Se y Sp del test
         upper_adj = conf.high-(1-Sp)/(1-(1-Se)-(1-Sp)))) #
            orignal combinado

saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")
```

Listing C.1: Codigo de R Software para estimaciones frecuentista estandar y Modelo de Liu para datos de VIH

```
######## ESTIMACION BAYESIANA DE COEFICIENTES DE REGRESION
   ##########

pacman::p_load(tidyverse, rio, binom, marginaleffects, rstan)

vih <- import("vih.rds")
new_datavih <- import("Nuevos_resultados/Data/new_datavih.rds")
x <- vih[,c("edad_gr","sex","reactivo_sif1","reactivo_hepb",
            "tip_pob_HSH", "tip_pob_LGTBIQ", "tip_pob_OTRAS
               POBLACIONES","tip_pob_TRABAJADOR SEXUAL")]
N <- 11452  # N mero de observaciones
P <- 8    # N mero de covariables
X <- as.matrix.data.frame(x)
# Asignar sensibilidad y especificidad
Se <- 0.9751  # Sensibilidad
Sp <- 0.999934  # Especificidad
y_obs <- vih$vih_1 # variable dependiente con error de medida
```

```
# Preparar datos para Stan
data_list <- list(N = N, P = P, X = X, y_obs = y_obs, Se = Se,
   Sp = Sp)

#
stanc(file = "Nuevos_resultados/Modelo_logistico_Bayesiano/VIH1.
   stan")
# Especificar la ruta al archivo .stan
stan_file <- "Nuevos_resultados/Modelo_logistico_Bayesiano/VIH1.
   stan"
#COMPILAR MODELO
stan_model <- stan_model(file = stan_file)
# Configuraci n para mejorar el rendimiento en RStan
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
# Ajustar el modelo en Stan
mod_vih1_bayesiano <- sampling(stan_model, data = data_list,
   iter = 2000, chains = 4)

fit2 <- stan_glm(vih_1 ~ edad_gr+sex+reactivo_sif1+
           reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`
              tip_pob_OTRAS POBLACIONES`+
           `tip_pob_TRABAJADOR SEXUAL`, data = vih,
         family = binomial(link = "logit"),cores = 2, seed =
            12345,prior = normal(0,0.6),
         prior_intercept =  normal(0,0.6))

mod_vih1_bayesiano <- fit2
saveRDS(mod_vih1_bayesiano,"Nuevos_resultados/Modelos/
   mod_vih1_bayesiano.rds")
# Resumen de los resultados
# Visualizar los resultados

# trace plot de coeficientes B e intercepto
traceplot_vih1 <- traceplot(mod_vih1_bayesiano, pars = c("alpha"
   ,"beta"))
```

```
ggsave(plot = traceplot_vih1, filename = "traceplot_vih1.png",
   device = "png",
       path = "Nuevos_resultados/Modelos",dpi = 600, units = "cm
          ",width = 17 ,height = 12)

hist(posterior_samples,xlim = c(0,0.15) ,breaks = 100, main = "
   Distribuci n Posterior de intercepto", xlab = "alpha")

# extraido media del posterior para cada coeficiente
coef <- get_posterior_mean(mod_vih1_bayesiano)[1:9,5]
# re-arreglo coeficientes
coef <- mod_vih1_bayesiano$coefficients
# errores
se <- mod_vih1_bayesiano$ses

coef_models_vih <- add_column(coef_models_vih) %>%
  data.frame(coef) %>%
  rename(mod_vih1_bayes = coef)

coef_models_vih <- add_column(coef_models_vih) %>%
  data.frame(se) %>%
  rename(se_vih1_bayes = se)


saveRDS(coef_models_vih, "Nuevos_resultados/Resultados/
   coef_models_vih.rds")


prev_global <- predictions(mod_vih1_bayesiano, newdata =
   new_datavih) %>%
  select(2:4)

#### intervalo de prevalencia aparente global y ajustada por Se
   y Sp REGRESION LOGISTICA
#### ESTANDAR
```

```
prevalencias <- prevalencias %>%
  add_case(tibble(Disease = "VIH 1-Bayesiano",
                  prev_global,
                  Se = Se, Sp = Sp) %>%
             mutate(P_ajustada = (estimate-(1-Sp))/(1-(1-Se)-(1-
                Sp)),
                  lower_adj = (conf.low-(1-Sp))/(1-(1-Se)-(1-
                     Sp)), # corregida por Se y Sp del test
                  upper_adj = (conf.high-(1-Sp))/(1-(1-Se)-(1-
                     Sp)))) # orignal combinado

saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")
```

Listing C.2: Codigo de R Software para estimaciones Bayesianas clasicas para datos de VIH

```
pacman::p_load(rstan, rio, tidyverse)

# Cargar los datos

vih <- readRDS("vih.rds")
# Definir variables
N <- nrow(vih)
t1 <- vih$vih_1
x1 <- vih$edad_gr
x2 <- vih$sex
x3 <- vih$reactivo_sif1
x4 <- vih$reactivo_hepb
x5 <- vih$tip_pob_HSH
x6 <- vih$tip_pob_LGTBIQ
x7 <- vih$`tip_pob_OTRAS POBLACIONES`
x8 <- vih$`tip_pob_TRABAJADOR SEXUAL`

SN <- 0.9751
SP <- 0.999934
```

```
# Estimaci n inicial de betas con regresi n log stica
# zz <- glm(t1 ~ x1 + x2 + x3 + x4, data = data1, family = "
   binomial")
# betas_init <- as.numeric(zz$coefficients)

# Generar valores iniciales aleatorios para el estado latente d1
set.seed(123)
#d1_init <- rbinom(N, 1, 0.5)  # Supongamos que la mitad de la
   poblaci n est  enferma

# Estructurar los datos para Stan
stan_data <- list(
  N = N,
  x1 = x1,
  x2 = x2,
  x3 = x3,
  x4 = x4,
  x5 = x5,
  x6 = x6,
  x7 = x7,
  x8 = x8,
  t1 = t1,
  SN = SN,
  SP = SP
)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

stanc(file = "Nuevos_resultados/Modelo_logistico_Bayesiano/
   Model_bayesiano1.stan")
# Especificar la ruta al archivo .stan
stan_file <- "Nuevos_resultados/Modelo_logistico_Bayesiano/
   Model_bayesiano1.stan"
#COMPILAR MODELO
stan_model <- stan_model(file = stan_file, auto_write = TRUE)

vih1_bayesiano_ajust <- sampling(stan_model, data = stan_data,
```

```
   chains = 4,
          iter = 4000)




saveRDS(model_fit,"Nuevos_resultados/Modelos/
   vih1_bayesiano_ajust.rds")

# Ver resultados
print(vih1_bayesiano_ajust, pars = c("betas"))

traceplot_vih_ajust <- traceplot(vih1_bayesiano_ajust)
ggsave(plot = traceplot_vih_ajust, filename = "
   traceplot_vih_ajust.png",device = "png",
       path = "Nuevos_resultados/Modelos",dpi = 600, units = "cm
          ",width = 17 ,height = 12)


o <- summary(vih1_bayesiano_ajust)

# coeficientes
coef <- o$summary[,1][-10]

# errores estandar
se <- o$summary[,3][-10]

coef_models_vih <- coef_models_vih[,-8]

# adicion a tabla de coeficientes
coef_models_vih <- add_column(coef_models_vih) %>%
  data.frame(coef) %>%
  rename(mod_vih_bayes_ajust = coef)

coef_models_vih <- add_column(coef_models_vih) %>%
  data.frame(se) %>%
  rename(se_vih_bayes_ajust = se)
```

```
saveRDS(coef_models_vih, "Nuevos_resultados/Resultados/
   coef_models_vih.rds")

nom_coeficientes <- names(mod_vih1_std$coefficients)

names(coef) <- nom_coeficientes

coef

##### modelo dummy para predicciones
modelo_dummy <- glm(vih_1 ~ edad_gr+sex+reactivo_sif1+
                      reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`
                         tip_pob_OTRAS POBLACIONES`+
                      `tip_pob_TRABAJADOR SEXUAL`, family = "
                         binomial", data = vih)

# Reemplazamos los coeficientes con los estimados con el Modelo
   de Liu

modelo_dummy$coefficients <- coef

prev_global <- predictions(modelo_dummy, newdata = new_datavih)
   %>%
  select(2,5:6)

prevalencias[10,1] <- "VIH -Bayesiano_ajust"

prevalencias[10,"Se"] <- Se

prevalencias[10,"P_ajustada"] <- prev_global[1]

prevalencias[10,"lower_adj"] <- prev_global[2]

prevalencias[10,"upper_adj"] <- prev_global[3]
```

```
saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
    rds")

#############  codigo de Stan ###################

data {
  int<lower=1> N;        // N mero de observaciones
  vector[N] x1;          // Covariable 1
  vector[N] x2;
  vector[N] x3;
  vector[N] x4;
  vector[N] x5;
  vector[N] x6;
  vector[N] x7;
  vector[N] x8;
  int<lower=0,upper=1> t1[N]; // Resultado del test diagn stico
  real<lower=0,upper=1> SN;   // Sensibilidad fija
  real<lower=0,upper=1> SP;   // Especificidad fija
}

parameters {
  vector[9] betas; // Coeficientes de regresi n log stica
}

model {
  vector[N] logit_p1;
  vector[N] p1;
  vector[N] p2;

  // Priores para los betas
  betas ~ normal(0, sqrt(0.366));

  // Modelo de probabilidad de enfermedad
  for (i in 1:N) {
    logit_p1[i] = betas[1] + betas[2] * x1[i] + betas[3] * x2[i]
        + betas[4] * x3[i] + betas[5] * x4[i] + betas[6] * x5[i]
       + betas[7] * x6[i] + betas[8] * x7[i] + betas[9] * x8[i];
```

```
    p1[i] = inv_logit(logit_p1[i]); // Probabilidad de estar
       enfermo

    // Modelamos el estado latente `d1[i]` implicitamente como
       una variable oculta
    p2[i] = p1[i] * SN + (1 - p1[i]) * (1 - SP);

    // Modelo de observaci n: test diagn stico
    t1[i] ~ bernoulli(p2[i]);
  }
}
```

Listing C.3: Cogido de R y Stan para estimaciones Bayesianas por el modelo de Valle et al.

```
install.packages("pacman")

pacman::p_load(logistic4p, tidyverse, rio, binom,
               gtsummary, janitor, fastDummies,
               marginaleffects)
# carga de datos

vih <- readRDS("vih.rds")

# carga de datos de combinaciones para marginalizacion
   estandarizada

# new_datasif <- vih %>%
#   select(edad_gr,sex,vih_1,reactivo_hepb,tip_pob_HSH,
#          tip_pob_LGTBIQ, `tip_pob_OTRAS POBLACIONES`,
#          `tip_pob_TRABAJADOR SEXUAL`)
#
# new_datasif <- datagrid(newdata = new_datasif)

new_datasif <- readRDs("Nuevos_resultados/Data/new_datasif.rds")

# saveRDS(new_datasif,"Nuevos_resultados/Data/new_datasif.rds")
```

```
######################### Se Sp del test utilizado
   ##################

# PRUEBA SIFILIS   Marca Bioline Syphilis 3.0 seleccion de
   limite inferior de IC95%
# sensibilidad: 0.964
# especificidad: 0.974


# sensibilidades y especificidad combinadas para test Syphilis
   3.0
se <- 0.964
sp <- 0.974
# calculo de r0 (Falsos positivos, FP)y r1 (falsos negativos, FN
   )

r0 <- 1-sp  # FP
r1 <- 1-se  # FN

######## ESTIMACION DE PREVALENCIA APARENTE CON  con ro y r1
   desconocido ############
##### metodo de Liu et cols. ESTIAMCION DE PREVALENCIA AJUSTADA(
   Covariables) Y AJUSTADA CORREGIDA por Se y Sp

x <- vih[,c("edad_gr","sex","vih_1","reactivo_hepb",
            "tip_pob_HSH", "tip_pob_LGTBIQ", "tip_pob_OTRAS
               POBLACIONES","tip_pob_TRABAJADOR SEXUAL")] #
               predictors
y <- vih[,"reactivo_sif1"]      # outcome

mod_sif_fpfn <- logistic4p(x, y, model='fp.fn', max.iter = 1000,
    detail = TRUE)
mod_sif_eq <- logistic4p(x, y, model='equal', max.iter = 1000,
   detail = TRUE)

print(mod_sif_fpfn)
print(mod_sif_eq)
```

```
# estiamdos de FP y FN por Liu en sifilis

fn <- 0  # fp = fn = -0.06394875 p-value = 0.00567
fp <- 0

saveRDS(mod_sif_eq, "Nuevos_resultados/Modelos/mod_sif_eq.rds")
saveRDS(mod_sif_fpfn, "Nuevos_resultados/Modelos/mod_sif_fpfn.
   rds")

### obtener coeficientes del modelo de Liu eq

coef <- mod_sif_eq$estimates[-1,1]


# creamos modelo dummy con coeficionete del modelo de Liu
modelo_dummy <- glm(reactivo_sif1 ~ edad_gr+sex+vih_1+
                      reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`
                         tip_pob_OTRAS POBLACIONES`+
                      `tip_pob_TRABAJADOR SEXUAL`, family = "
                         binomial", data = vih)
# Reemplazamos los coeficientes
modelo_dummy$coefficients <- coef

# guarda coeficientes estiamdos

coef_models_sif <- data.frame(coef) %>%
  rename(coef_sif_eq = coef)

saveRDS(coef_models_sif, "Nuevos_resultados/Resultados/
   coef_models_sif.rds")

# calculo de prevalencia global y IC95%

prev_global <- predictions(modelo_dummy, newdata = new_datasif)
   %>%
  select(2,4:5)
```

```
#### intervalo de prevalencia aparente global y ajustada por Se
   y Sp

prevalencias <- prevalencias %>%
  add_case(
  tibble(Disease = "Sifilis - Liu.eq",
                       prev_global,
                       Se = 1-fn, Sp = 1-fp) %>%
  mutate(P_ajustada = (estimate-fp)/(1-fn-fp),
         lower_adj = (conf.low-fp)/(1-fn-fp),
         upper_adj = (conf.high-fp)/(1-fn-fp))
  )

saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")


############## RESULTADOS POR  REGRESION LOGISITCA ESTANDAR
   ###############

mod_sif_std <- glm(reactivo_sif1 ~ edad_gr+sex+vih_1+
      reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`tip_pob_OTRAS
         POBLACIONES`+
        `tip_pob_TRABAJADOR SEXUAL`, family = "binomial", data =
           vih)


summary(mod_sif_std)
saveRDS(mod_sif_std,"Nuevos_resultados/Modelos/mod_sif_std.rds")
### obtener coeficientes del modelo de Liu

coef <- mod_sif_std$coefficients

# guarda coeficientes estiamdos

coef_models_sif <- add_column(coef_models_sif) %>%
```

```
data.frame(coef) %>%
  rename(coef_sif_std = coef)

saveRDS(coef_models_sif, "Nuevos_resultados/Resultados/
   coef_models_sif.rds")

# calculo de prevalencia global con modelo regresion logistica

prev_global <- predictions(mod_sif_std, newdata = new_datasif)
   %>%
  select(2,4:5)

#### intervalo de prevalencia aparente global y ajustada por Se
   y Sp REGRESION LOGISTICA
#### ESTANDAR

prevalencias <- prevalencias %>%
  add_case(tibble(Disease = "Sifilis-STD",
                  prev_global,
                  Se = se, Sp = sp) %>%
             mutate(P_ajustada = estimate-(1-Sp)/(1-(1-Se)-(1-Sp
                ),
                    lower_adj = conf.low-(1-Sp)/(1-(1-Se)-(1-Sp)
                       ), # corregida por Se y Sp del test
                    upper_adj = conf.high-(1-Sp)/(1-(1-Se)-(1-Sp
                       )))) # orignal combinado

saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")
```

Listing C.4: Codigo R Software para estimaciones de modelo de Sifilis

```
install.packages("pacman")

pacman::p_load(tidyverse, rio, binom,
               gtsummary, janitor, fastDummies,
               marginaleffects, rstanarm, bayesplot)
```

```
# carga de datos

# carga de datos de combinaciones para marginalizacion
   estandarizada

new_datasif <- readRDS("Nuevos_resultados/Data/new_datasif.rds")

# saveRDS(new_datasif,"Nuevos_resultados/Data/new_datasif.rds")

######################### Se Sp del test utilizado
   ##################

# PRUEBA SIFILIS   Marca Bioline Syphilis 3.0 seleccion de
   limite inferior de IC95%
# sensibilidad: 0.964
# especificidad: 0.974


# sensibilidades y especificidad combinadas para test Syphilis
   3.0
se <- 0.964
sp <- 0.974


# MODELO DE REGRESION LOGISRICA CON STAN
mod_sif_bayesiano <- stan_glm(reactivo_sif1 ~ edad_gr+sex+vih_1+
                    reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`
                       tip_pob_OTRAS POBLACIONES`+
                    `tip_pob_TRABAJADOR SEXUAL`, data = vih,
                  family = binomial(link = "logit"),cores = 2,
                     seed = 12345,prior = normal(0,0.6),
                  prior_intercept =  normal(0,0.6))

saveRDS(mod_sif_bayesiano,"Nuevos_resultados/Modelos/
   mod_sif_bayesiano.rds")
```

```
posterior <- as.matrix(mod_sif_bayesiano)

plot_title <- ggtitle("Posterior distributions",
                      "with medians and 95% intervals")

# distribuciones posteriores
mcmc_areas(posterior,
           pars = c("reactivo_hepb","tip_pob_HSH","
              tip_pob_LGTBIQ"),
           prob = 0.95) + plot_title

mcmc_areas(posterior,
           pars = c("edad_gr","sex","vih_1"),
           prob = 0.95) + plot_title

# trace plot
plot_title <- ggtitle("Trace plot coefficients for syphilis
   Model")
traceplot_sifilis <- mcmc_trace(posterior, n_warmup = 300)+
   plot_title

ggsave(filename = "traceplot_sifilis.png", plot =
   traceplot_sifilis,device = "png",
       units = "cm",path = "Nuevos_resultados/Modelos",width =
          25,height = 10,dpi = 300)

# inetrvals
mcmc_intervals(
  posterior,
  pars = character(),
  regex_pars = character(),
  transformations = list(),
  prob = 0.5,
  prob_outer = 0.9,
  point_est = c("median"),
  outer_size = 0.5,
```

```
  inner_size = 2,
  point_size = 4,
  rhat = numeric()
)
# re-arreglo coeficientes
coef <- mod_sif_bayesiano$coefficients
# errores standar
se <- mod_sif_bayesiano$ses

coef_models_sif <- add_column(coef_models_sif) %>%
  data.frame(coef) %>%
  rename(mod_sif_bayes = coef)

coef_models_sif <- add_column(coef_models_sif) %>%
  data.frame(se) %>%
  rename(se_sif_bayes = se)

saveRDS(coef_models_sif, "Nuevos_resultados/Resultados/
   coef_models_sif.rds")

# Resumen de los resultados
prev_global <- predictions(mod_sif_bayesiano, newdata =
   new_datasif) %>%
  select(2:4)

#### intervalo de prevalencia aparente global y ajustada por Se
   y Sp REGRESION LOGISTICA
#### ESTANDAR

prevalencias <- prevalencias %>%
  add_case(tibble(Disease = "Sifilis - Bayesiano",
                  prev_global,
                  Se = se, Sp = sp) %>%
             mutate(P_ajustada = (estimate-(1-Sp))/(1-(1-Se)-(1-
                Sp)),
                    lower_adj = (conf.low-(1-Sp))/(1-(1-Se)-(1-
                       Sp)), # corregida por Se y Sp del test
```

```
                    upper_adj = (conf.high-(1-Sp))/(1-(1-Se)-(1-
                       Sp)))) # orignal combinado

saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")
```

Listing C.5: software R y Stan para estimaciones bayesianas clasicas en el modelo de Sifilis

```
pacman::p_load(rstan, rio, tidyverse)

# Cargar los datos

vih <- readRDS("vih.rds")
# Definir variables
N <- nrow(vih)
t1 <- vih$reactivo_sif1
x1 <- vih$edad_gr
x2 <- vih$sex
x3 <- vih$vih_1
x4 <- vih$reactivo_hepb
x5 <- vih$tip_pob_HSH
x6 <- vih$tip_pob_LGTBIQ
x7 <- vih$`tip_pob_OTRAS POBLACIONES`
x8 <- vih$`tip_pob_TRABAJADOR SEXUAL`

SN <- 0.964
SP <- 0.974

# Estimaci n inicial de betas con regresi n log stica
# zz <- glm(t1 ~ x1 + x2 + x3 + x4, data = data1, family = "
   binomial")
# betas_init <- as.numeric(zz$coefficients)

# Generar valores iniciales aleatorios para el estado latente d1
set.seed(123)
#d1_init <- rbinom(N, 1, 0.5)  # Supongamos que la mitad de la
```

```
   poblaci n  est   enferma

# Estructurar los datos para Stan
stan_data <- list(
  N = N,
  x1 = x1,
  x2 = x2,
  x3 = x3,
  x4 = x4,
  x5 = x5,
  x6 = x6,
  x7 = x7,
  x8 = x8,
  t1 = t1,
  SN = SN,
  SP = SP
)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

stanc(file = "Nuevos_resultados/Modelo_logistico_Bayesiano/
   Model_bayesiano1.stan")
# Especificar la ruta al archivo .stan
stan_file <- "Nuevos_resultados/Modelo_logistico_Bayesiano/
   Model_bayesiano1.stan"
#COMPILAR MODELO
stan_model <- stan_model(file = stan_file, auto_write = TRUE)

sif_bayesiano_ajust <- sampling(stan_model, data = stan_data,
   chains = 4,
                                  iter = 4000)

saveRDS(sif_bayesiano_ajust,"Nuevos_resultados/Modelos/
   sif_bayesiano_ajust.rds")

# Ver resultados
print(sif_bayesiano_ajust, pars = c("betas"))
```

```
traceplot_sif_ajust <- traceplot(sif_bayesiano_ajust)
ggsave(plot = traceplot_sif_ajust, filename = "
   traceplot_sif_ajust.png",device = "png",
       path = "Nuevos_resultados/Modelos",dpi = 600, units = "cm
          ",width = 17 ,height = 12)


o <- summary(sif_bayesiano_ajust)

# coeficientes
coef <- o$summary[,1][-10]

# errores estandar
se <- o$summary[,3][-10]

coef_models_sif <- coef_models_sif[,-8]
# adicion a tabla de coeficientes
coef_models_sif <- add_column(coef_models_sif) %>%
  data.frame(coef) %>%
  rename(sif_bayesiano_ajust = coef)

coef_models_sif <- add_column(coef_models_sif) %>%
  data.frame(se) %>%
  rename(se_sif_bayesiano_ajust = se)


saveRDS(coef_models_sif, "Nuevos_resultados/Resultados/
   coef_models_sif.rds")

nom_coeficientes <- names(mod_sif_std$coefficients)

names(coef) <- nom_coeficientes

coef

##### modelo dummy para predicciones
```

```
modelo_dummy <- glm(reactivo_sif1 ~ edad_gr+sex+vih_1+
                      reactivo_hepb+tip_pob_HSH+tip_pob_LGTBIQ+`
                        tip_pob_OTRAS POBLACIONES`+
                      `tip_pob_TRABAJADOR SEXUAL`, family = "
                        binomial", data = vih)

# Reemplazamos los coeficientes con los estimados con el Modelo
   de Liu

modelo_dummy$coefficients <- coef

prev_global <- predictions(modelo_dummy, newdata = new_datasif)
   %>%
  select(2,5:6)

prevalencias[11,1] <- "Sifilis -Bayesiano_ajust"

prevalencias[11,"Se"] <- SN

prevalencias[11,"Sp"] <- SP

prevalencias[11,"P_ajustada"] <- prev_global[1]

prevalencias[11,"lower_adj"] <- prev_global[2]

prevalencias[11,"upper_adj"] <- prev_global[3]

saveRDS(prevalencias,"Nuevos_resultados/Resultados/prevalencias.
   rds")

############ Codigo de Stan ######################
data {
  int<lower=1> N;        // N mero de observaciones
  vector[N] x1;          // Covariable 1
  vector[N] x2;
  vector[N] x3;
  vector[N] x4;
```

```
  vector[N] x5;
  vector[N] x6;
  vector[N] x7;
  vector[N] x8;
  int<lower=0,upper=1> t1[N]; // Resultado del test diagn stico
  real<lower=0,upper=1> SN;   // Sensibilidad fija
  real<lower=0,upper=1> SP;   // Especificidad fija
}

parameters {
  vector[9] betas; // Coeficientes de regresi n log stica
}

model {
  vector[N] logit_p1;
  vector[N] p1;
  vector[N] p2;

  // Priores para los betas
  betas ~ normal(0, sqrt(0.366));

  // Modelo de probabilidad de enfermedad
  for (i in 1:N) {
    logit_p1[i] = betas[1] + betas[2] * x1[i] + betas[3] * x2[i]
        + betas[4] * x3[i] + betas[5] * x4[i] + betas[6] * x5[i]
       + betas[7] * x6[i] + betas[8] * x7[i] + betas[9] * x8[i];
    p1[i] = inv_logit(logit_p1[i]); // Probabilidad de estar
       enfermo

    // Modelamos el estado latente `d1[i]` implicitamente como
       una variable oculta
    p2[i] = p1[i] * SN + (1 - p1[i]) * (1 - SP);

    // Modelo de observacion: test diagnostico
    t1[i] ~ bernoulli(p2[i]);
  }
}
```

Listing C.6: Codigo R y Stan para estimaciones Bayesianas mediante el modelo de Valle et. al para Sifilis

# Anexos D

# Publicaciones

## D.1 Articulos publicados

- Estrada Alvarez JM, Luna del Castillo JdD, Montero-Alonso M Point and Interval Estimation of Population Prevalence Using a Fallible Test and a Non-Probabilistic Sample: Post-Stratification Correction. Mathematics. 2025;13(5). Available from: https://www.mdpi.com/2227-7390/13/5/805

- Rodriguez* FH, Estrada* JM, Quintero HMA, Nogueira JP, Porras-Hurtado GL. Analyses of familial chylomicronemia syndrome in Pereira, Colombia 2010–2020: a cross-sectional study. Lipids in Health and Disease. 2023;22:43. Available from: https://doi.org/10.1186/s12944-022-01768-x
  **Se comparte primera autoria*

- Alvarez JME, Garcia HF, Ángel Montero-Alonso M, de Dios Luna del Castillo J. Prevalence estimation in infectious diseases with imperfect tests: A comparison of Frequentist and Bayesian Logistic Regression methods with misclassification correction; 2025. Available from: https://arxiv.org/abs/2504.15150

# Point and interval estimation of population prevalence using a fallible test and a non-probabilistic Sample: post-stratification correction

Jorge Mario Estrada A. Juan de Dios Luna del Castillo

Miguel Ángel Montero Alonso

## Abstract

Accurate prevalence estimation is crucial for public health planning, particularly for rare diseases or low-prevalence conditions. This study evaluated frequentist and Bayesian methods for estimating prevalence, addressing challenges such as imperfect diagnostic tests, partial disease status verification, and non-probabilistic samples. Post-stratification, applied as a novel method, was used to improve representativeness and correct biases.

Three scenarios were analyzed: (1) full verification using a gold standard, (2) estimation with a diagnostic test of known sensitivity and specificity, and (3) partial verification of disease status limited to test positives. In all scenarios, post-stratification adjustments increased prevalence estimates and interval lengths, highlighting the importance of accounting for population variability. Bayesian methods demonstrated advantages in integrating prior information and modeling uncertainty, particularly under high variability and low-prevalence conditions.

Key findings included the flexibility of Bayesian approaches to maintain estimates within plausible ranges and the effectiveness of post-stratification in correcting biases in non-probabilistic samples. Frequentist methods provided narrower intervals but were limited in addressing inherent uncertainties. This study underscores the need for methodological adjustments in epidemiological studies, offering robust solutions for real-world challenges. These results have significant implications for improving public health decision-making and the design of prevalence studies in resource-constrained or non-probabilistic contexts.

## Introduction

The estimation of prevalence in epidemiological studies is a fundamental component for public health planning, particularly for pathological conditions with low population prevalence. However, this process faces numerous methodological challenges, such as the need for large sample sizes and the implementation of complex designs for probabilistic sample selection[1];[2]. Additionally, the use of diagnostic tests with inaccuracies in the absence of a gold standard, or the limited application of such tests to only a fraction of the sample, adds complexity to the analysis and may introduce bias. Furthermore, it is crucial to employ advanced interval estimation methods

that overcome the inherent limitations of extremely low prevalence rates, particularly those below 0.1[3].

Various authors have addressed the methodological challenges associated with estimating prevalence in the context of low prevalence conditions and imperfect diagnostic tests. For instance, Rogan and Gladen[4] proposed a classical approach for inference and bias correction in estimated prevalences when using tests with sensitivities and specificities below 1.0. This approach has been revisited and expanded more recently by Izbicki[5]. Similarly, Reiczigel[6] proposed an exact method for constructing confidence intervals when employing tests with known sensitivity and specificity, a contribution that was later developed further by Lang et al.[7].

Another methodological challenge in the estimation of low prevalence lies in the combined use of a fallible diagnostic test within a two-phase cross-sectional design, where verification via a gold standard is performed only on positive cases. This issue was addressed by Thomas et al.[8], who developed both frequentist and bayesian estimators to correct the bias introduced by this verification scheme.

On the other hand, while probabilistic samples enable robust population-level inference, their implementation often involves high costs and significant limitations in resource-constrained settings. In such contexts, convenience and non-probabilistic samples present a practical alternative. Various methodological approaches have been proposed for estimation with non-probabilistic samples[9];[10]. A notable example is the use of post-stratification estimators, as proposed by Smith[11], which leverage auxiliary population-level covariate distributions to correct prevalence estimates. Originally designed to mitigate nonresponse bias in survey sampling, this technique serves as an effective tool for adjusting bias in non-probabilistic samples through the estimation of selection probabilities.

The purpose of this study is to present and evaluate some of the previously described methods for estimating the prevalence of rare or orphan diseases, incorporating adjustments proposed by various authors. These methods address contexts involving the use of fallible diagnostic tests and partial verification schemes applied exclusively to individuals who test positive. Additionally, a novel post-stratification approach is integrated as a strategy to correct prevalence estimates derived from non-probabilistic samples. The methods and adjustments are applied and analyzed in different scenarios using real-world data.

## Methods

Three scenarios were established for the use of different interval estimation methods, utilizing two main approaches: the frequentist approach and the bayesian approach. For each of these approaches, a version of the method with post-stratification adjustment was applied.

### Estimation of a proportion in non-probabilistic samples with post-Stratification adjustment

Post-stratification is a weighting adjustment technique used in non-probabilistic samples. It was initially proposed to correct various biases in probabilistic samples, such as undercoverage, overcoverage, and nonresponse bias. This technique involves adjusting the weights of sample units to reflect the known proportions of a target population divided into strata. However, these strata are constructed after the sample has been selected. In the ideal case, using a gold-standard test, the prevalence estimate adjusted for post-stratification, as proposed by Holt

et al.[12], in standard notation, assumes that the population comprises $N$ units, which can be uniquely divided into $H$ strata of sizes $N_1, N_2, \ldots, N_H$, such that $\sum_{h=1}^{H} N_h = N$.

Let $y_i$ be a variable taking values $y_{hi}$, where $h = 1, \ldots, H$ and $i = 1, \ldots, N_h$. The selected sample is of fixed size $n$, which, after selection, is distributed among the strata according to the vector $n = (n_1, n_2, \ldots, n_H)$, where $\sum_{h=1}^{H} n_h = n$. The components of $n$ are not known until after the sample is drawn.

The prevalence estimator for the population in each stratum $h$ is defined as:

$$\hat{p}_h = \frac{\sum_{i=1}^{n_h} y_{ih}}{n_h}$$

where $y_i$ is an indicator variable, such that $y_i = 1$ if the individual in the sample tests positive, and $y_i = 0$ otherwise. Additionally, $x_h = \sum_{i=1}^{n_h} y_{ih}$ represents the total number of cases identified in stratum $h$.

Thus, the population prevalence estimator adjusted for post-stratification, $\hat{p}_w$, and its variance, $\hat{V}(\hat{p}_w)$, are expressed as:

$$\hat{p}_w = \sum_{h=1}^{H} \frac{N_h}{N} \hat{p}_h$$

$$\hat{V}(\hat{p}_w) = \sum_{h=1}^{H} \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1}$$

## Scenario 1: Prevalence Estimation with Full Verification Using a Gold Standard

### Frequentist approach

Let $P$ denote the unknown population prevalence to be estimated from a probabilistic sample of size $n$. Following the interval proposed by Agresti and Coull[3], an adjustment referred to as the "add 2 successes and 2 failures" method is used. This adjustment modifies both the sample size $n$ and the estimator $\hat{p}$. The corresponding expressions are:

$$\tilde{n} \equiv n + z_{\alpha}^2$$

and

$$\tilde{p} = \frac{1}{\tilde{n}} \left(x + \frac{z_{\alpha/2}^2}{2}\right)$$

where $z_{\alpha/2}$ represents the critical value corresponding to the random variable from a standard normal distribution, and $x$ denotes the number of observed cases. Based on these definitions, an interval for the population prevalence $P$ is expressed as:

$$P \approx \tilde{p} \pm z_{\alpha} \sqrt{\frac{\tilde{p}}{\tilde{n}}(1 - \tilde{p})}$$

**Post-Stratification adjustment:** Let $\tilde{P}$ denote the adjusted prevalence, where $w_h$ represents the weight of stratum $h$ in the population, and $\hat{p}_h$ is the estimated prevalence within stratum $h$.

The adjustment for the sample size per stratum, $\tilde{n}_h$, is given by:

$$\tilde{n}_h = n_h + z^2_{\alpha/2}$$

- Adjustment in weighted prevalence: The estimator $\tilde{p}$ must also be adjusted using the weights and prevalences of each stratum, and it would be expressed as:

$$\tilde{p}_h = \frac{x_h + \frac{z^2_\alpha}{2}}{\tilde{n}_h}$$

For the weighted prevalence, the formula is:

$$\tilde{P} = \sum_{h=1}^{H} w_h \tilde{p}_h$$

The weighted variance is given by:

$$\text{Var}(\tilde{P}) = \sum_{h=1}^{H} w_h^2 \frac{\tilde{p}_h(1-\tilde{p}_h)}{\tilde{n}_h}$$

The confidence interval is:

$$P \approx \tilde{P} \pm z_\alpha \sqrt{\text{Var}(\tilde{P})}$$

### Bayesian approach

From a bayesian perspective, an interval for the parameter $P$ can be formulated under a beta-binomial model, known for its favorable frequentist properties[13], and with an adjustment for post-stratification.

**Notation** $H$: Number of strata in the population.

$n_h$: Sample size in stratum $h$.

$x_h$: Number of positive outcomes observed in stratum $h$.

$w_h$: Proportion of the total population belonging to stratum $h$.

$P_h$: Prevalence in stratum $h$.

$\tilde{P}$: Overall prevalence adjusted for post-stratification.

For the prevalence in each stratum, we use a beta distribution as the prior: $P_h \sim \text{Beta}(\alpha_h, \beta_h)$. In this specific case, we use a non-informative prior: $P_h \sim \text{Beta}(1, 1)$.

The likelihood of observing $x_h$ positives in a sample of size $n_h$ in stratum $h$ is given by the binomial distribution: $x_h \sim \text{Binomial}(n_h, P_h)$

This means:

$\mathbf{P}(x_h|n_h, P_h) = \binom{n_h}{x_h} P_h^{x_h}(1-P_h)^{n_h-x_h}$

posterior distribution for $P_h$ using Bayes' theorem, we combine the prior and the likelihood to obtain the posterior distribution for $P_h$:

$\mathbf{P}(P_h|x_h, n_h) \propto \mathbf{P}(x_h|n_h, P_h) \cdot \mathbf{P}(P_h)$

substituting the terms:

$\mathbf{P}(P_h|x_h, n_h) \propto P_h^{x_h}(1-P_h)^{n_h-x_h} \cdot P_h^{\alpha_h-1}(1-P_h)^{\beta_h-1}$

for $\alpha_h = 1$ y $\beta_h = 1$: $\mathbf{P}(P_h|x_h, n_h) \propto P_h^{x_h}(1-P_h)^{n_h-x_h}$

the adjusted prevalence, $(\tilde{P})$, is a weighted combination of subgroup prevalences, where the weights are the proportions of each stratum in the total population. During the bayesian inference process, samples are drawn from the posterior distribution of $P_h$ for each stratum $h$, which we denote as $P_h^s$, where $s$ indicates a specific sample from the posterior distribution.

$$\tilde{P}^s = \sum_{h=1}^{H} w_h P_h^s$$

Based on the above, an interval for the adjusted prevalence $\tilde{P}$ can be defined as:

$$\text{ICr}_{95\%}(\tilde{P}) = \left[\tilde{P}_{2.5}, \tilde{P}_{97.5}\right]$$

Where:

$\tilde{P}_{2.5}$ is the 2.5th percentile of the posterior samples of $\tilde{P}^{(s)}$. $\tilde{P}_{97.5}$ is the 97.5th percentile of the posterior samples of $\tilde{P}^{(s)}$.

## Scenario 2: Estimation with a single diagnostic test with known sensitivity and specificity

In most practical situations, the use of a gold-standard test is not feasible, either due to the complexity of its application or the high associated costs. This limitation is further exacerbated when, due to sampling schemes—whether simple or complex—the gold standard is not available for all selected individuals. Additionally, in some cases, its use may be contraindicated due to specific indications.

### Frequentist approach

Reiczigel[6] proposes the estimation of prevalence using an exact method that corrects the bias introduced by an imperfect diagnostic test, assuming its sensitivity and specificity are known. Later, Lang et al.[7] extended this proposal by combining the formula of Rogan and Gladen with an adjustment based on the method proposed by Agresti and Coull.

Let $Se$ and $Sp$ denote the sensitivity and specificity, respectively, obtained from an independent study of an imperfect diagnostic test used to estimate the prevalence $P$ of a condition or disease. The apparent prevalence $(AP)$ is defined as the proportion of positive results obtained from the test in the target population. According to Rogan and Gladen, the point estimator of $P$ is calculated as:

$$\widehat{p} = \frac{\widehat{AP} + \widehat{Sp} - 1}{\widehat{Se} + \widehat{Sp} - 1}$$

if the adjustment is applied to the interval proposed by Agresti and Coull:
on $\widehat{AP}$:

$$n' = n + Z_\alpha^2$$

$$\widehat{AP'} = \frac{n \cdot \widehat{AP} + Z_\alpha^2/2}{n'}$$

Thus, both the corrected point estimate of $P$ and its confidence interval are constructed as:

$$\widehat{p}' = \frac{\widehat{AP'} + Sp - 1}{Se + Sp - 1}$$

$$\hat{p}' \pm Z_{\alpha/2} \frac{1}{(Se + Sp - 1)} \left( \frac{\widehat{AP'}(1 - \widehat{AP'})}{n'} \right)^{1/2}$$

**Post-Stratification Adjustment** $H$: Number of strata in the population

$n_h$: Sample size in stratum $h$

$x_h$: Number of positive results in stratum $h$ using a fallible test

$w_h$: Proportion of the total population belonging to stratum $h$

$Se$: Sensitivity of the test

$Sp$: Specificity of the test

Applying the adjustment proposed by Agresti and Coull for each stratum $h$:

$\widehat{AP'_h} = \frac{x_h + Z^2_{\alpha/2}/2}{n_h + Z^2_{\alpha/2}}$

$\hat{p}'_h = \frac{\widehat{AP'_h} + Sp - 1}{Se + Sp - 1}$

The estimator for the prevalence adjusted by post-stratification and the known sensitivity and specificity of the test, $\tilde{P}$, is:

$\tilde{P} = \sum_{h=1}^{H} w_h \hat{p}'_h$

The variances for each stratum:

$$\mathrm{Var}(\hat{p}'_h) = \frac{1}{(Se + Sp - 1)^2} \cdot \frac{\widehat{AP'_h}(1 - \widehat{AP'_h})}{n'_h}$$

The combined total variance is:

$$\text{Var}(\tilde{P}) = \sum_{h=1}^{H} w_h^2 \cdot \text{Var}(\hat{p}'_h)$$

Finally, the confidence interval adjusted for post-stratification is:

$$\tilde{P} \pm Z_{\alpha/2} \cdot \sqrt{\text{Var}(\tilde{P})}$$

### Bayesian approach

Flor et al.[14] proposed a comparative evaluation of point and interval estimation methods, utilizing both frequentist and bayesian approaches, for the estimation of prevalence with an imperfect diagnostic test. In this context, the bayesian approach offers the advantage of incorporating prior information about the test's sensitivity and specificity, which can improve the precision of the estimates.

Under a bayesian framework, the true prevalence ($P$) is considered a random parameter with a prior distribution. The sensitivity ($Se$) and specificity ($Sp$) are known parameters, obtained previously from independent studies, and are defined as follows:

The true prevalence $P$, which is the parameter of interest in this model, is estimated using a non-informative prior, reflecting a lack of prior knowledge about its value before observing the data.

$$P \sim \text{Beta}(1, 1) \quad \text{(The prior distribution is uniform over the interval [0,1])}$$

This non-informative prior reflects that any value of $P$ between 0 and 1 is equally likely a priori.

The likelihood of observing ($x$) positive cases using the test in a sample of size $n$, with a probability of success given by the apparent prevalence ($AP$)—defined as the proportion of positive results—follows a binomial distribution:

$x \sim \text{Binomial}(n, AP)$

The apparent prevalence $AP$ is related to the true prevalence $P$, the sensitivity $Se$, and the specificity $Sp$ of the diagnostic test through the following equation:

$$AP = P \cdot Se + (1 - P) \cdot (1 - Sp)$$

Here, $Se$ and $Sp$ are known and do not need to be estimated. The likelihood function, that is, the probability of observing $x$ positives in the sample, is:

$$\mathbf{P}(x|n, \pi, Se, Sp) = \binom{n}{x} \cdot AP^x \cdot (1 - AP)^{n-x}$$

The posterior distribution for the prevalence $P$, using Bayes' theorem, can be computed for $P$ given the observed data $x$, the sample size $n$, and the known values of $Se$ and $Sp$:

$$\mathbf{P}(P|x, n, Se, Sp) \propto P(x|n, \pi, Se, Sp) \cdot \mathbf{P}(P)$$

Substituting the terms:

$$\mathbf{P}(P|x,n,Se,Sp) \propto \binom{n}{x} \cdot (P \cdot Se + (1-P) \cdot (1-Sp))^x \cdot (1-(P \cdot Se + (1-P) \cdot (1-Sp)))^{n-x} \cdot \text{Beta}(1,1)$$

Since the prior $P$ is a uniform distribution Beta$(1,1)$, the posterior calculation can be simplified as:

$$\mathbf{P}(P|x,n,Se,Sp) \propto (P \cdot Se + (1-P) \cdot (1-Sp))^x \cdot (1-(P \cdot Se + (1-P) \cdot (1-Sp)))^{n-x}$$

To obtain the posterior distribution of $P$, numerical methods such as MCMC (Markov Chain Monte Carlo) can be used. This approach allows generating samples from the posterior distribution of $P$ and performing inferences about the true prevalence.

**Post-Stratification adjustment** The previous bayesian estimation can be complemented with a post-stratification adjustment, a technique that approximates an adjustment in the context of a non-probabilistic sample. In this case, as previously defined:

$H$: Number of strata in the population.

$n_h$: Sample size in stratum $h$.

$x_h$: Number of positive results in stratum $h$ using a fallible test.

$w_h$: Proportion of the total population belonging to stratum $h$.

$Se$: Sensitivity of the test.

$Sp$: Specificity of the test.

For $H$ strata, each with a prevalence $P_h$, and each stratum having a proportion $w_h$ of the total population, the total prevalence adjusted for post-stratification can be calculated as:

$$\mathbf{P}(P_h|x_h,n_h,Se,Sp) \propto (P_h \cdot Se + (1-P_h) \cdot (1-Sp))^{x_h} \cdot (1-(P_h \cdot Se + (1-P_h) \cdot (1-Sp)))^{n_h-x_h}$$

$$\tilde{P} = \sum_{h=1}^{H} w_h \cdot P_h$$

However, when estimating the prevalence $P_h$ for each stratum $h$ under a bayesian approach, we do not obtain a single point estimate but rather a posterior distribution for $P_h$. Thus, this post-stratification must be based on posterior distributions.

Suppose we have obtained $S$ samples from the posterior distribution of $P_h$ for $H$ subgroups. We denote these samples for each subgroup $h$ as:

$$P_h^{(1)}, P_h^{(2)}, \ldots, P_h^{(S)}$$

For each iteration $s$ in the posterior, we weight the prevalence $P_h^{(s)}$ by the proportion $w_h$ of stratum $h$ in the total population. In this way, the adjusted prevalence in iteration $s$ is:

$$\tilde{P}^{(s)} = \sum_{h=1}^{H} w_h \cdot P_h^{(s)}$$

This process is repeated for each sample $s = 1, \ldots, S$, resulting in a combined posterior distribution for the total prevalence adjusted by post-stratification:

$$\tilde{P}^{(1)}, \tilde{P}^{(2)}, \ldots, \tilde{P}^{(S)}$$

From the samples $\tilde{P}^{(s)}$, it is possible to calculate any credible interval. For example, the 2.5th and 97.5th percentiles of the posterior distribution of $\tilde{P}$ can be used to define a 95% credible interval. Similarly, the point estimate is obtained as the median of the posterior distribution of $\tilde{P}$.

## Scenario 3: Estimation when disease status is verified only among test positives

### Frequentist Approach

This scenario is common during diagnostic processes or public health surveys, where the verification of the true disease status is performed only on individuals who, in an initial evaluation, test positive using a fallible or screening test. For reasons of cost or ethics, verification is typically limited to those with a high probability of disease based on the screening test results.

In this context, Thomas et al.[8] derived maximum likelihood estimators for the prevalence of the condition or disease under the two approaches discussed in this article: frequentist and bayesian. These estimates assume that the sensitivity and specificity of the screening test are known and present no uncertainty. The details of these approaches are presented below.

The likelihood function proposed was:

$$\mathcal{L}(P) = [(1 - Se)P + Sp(1 - P)]^{N_{00}} (Se \cdot P)^{N_{11}} ((1 - Sp)(1 - P))^{N_{01}}$$

Where:

True positives are $(N_{11})$, false positives among those who tested positive are $N_{01}$, total negatives according to the test are $(N_{\cdot 0})$, prevalence is $(P)$ and Sensitivity and specificity, respectively $(Se, Sp)$.

The likelihood function for the data is constructed based on these observed values and the probability of test results given the true health conditions of the individuals (for more details, see[8]). This allows the derivation of the maximum likelihood estimator (MLE) for the prevalence, which is given by the following expression:

$$\hat{p} = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

where the terms $A$, $B$, and $C$ are defined as:

$$A = -(1 - Se - Sp)n$$

$$B = (1 - Se - Sp)(N_{\cdot 0} + N_{11}) - Sp(N_{11} + N_{01})$$

$$C = N_{11}Sp$$

Here, $n$ represents the total sample size (the sum of individuals who tested positive and negative in the screening test).

The estimator $\hat{p}$ provides an accurate estimate of the prevalence, and its asymptotic variance is given by:

$$\hat{\sigma}^2 = \left[ \frac{Se}{\hat{p}} + \frac{1-Sp}{1-\hat{p}} + \frac{(1-Se-Sp)^2}{(1-Se)\hat{p} + Sp(1-\hat{p})} \right]^{-1}$$

Using this variance, an asymptotic $100(1-\alpha)$ confidence interval can be constructed as:

$$\hat{p} \pm z_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

However, in daily practice, screening programs are not typically based on random samples from the population. Therefore, disease prevalence estimates require adjustments to reduce the bias associated with this practice. In this context, post-stratification emerges as a suitable methodological option to achieve such adjustments.

**Post-Stratification adjustment** The prevalence for each stratum is calculated using the method proposed by[8]. A $\hat{p_h}$ is defined for each of the $H$ strata in the sample:

$$\hat{p_h} = \frac{-B_h - \sqrt{B_h^2 - 4A_hC_h}}{2A_h}$$

where $A_h$, $B_h$, and $C_h$ are defined as previously. The variance for each stratum is:

$$\hat{\sigma}_h^2 = \left[ \frac{Se}{\hat{p_h}} + \frac{1-Sp}{1-\hat{p_h}} + \frac{(1-Se-Sp)^2}{(1-Se)\hat{p_h} + Sp(1-\hat{p_h})} \right]^{-1}$$

The overall prevalence adjusted by post-stratification is:

$$\tilde{p} = \sum_{h=1}^{H} w_h \cdot \hat{p}_h$$

The weighted variance is given by:

$$Var(\tilde{p}) = \sum_h \left( \frac{N_h}{N} \right)^2 \hat{\sigma}_h^2$$

A confidence interval for $\tilde{p}$ is:

$$IC(\tilde{p}) = \tilde{p} \pm z_{1-\alpha/2} \sqrt{Var(\tilde{p})}$$

### Bayesian approach

In the same work published by Thomas et al.[8], the bayesian version for prevalence estimation is presented for cases where verification using a gold standard is performed only on individuals who test positive in a screening test.

Starting from the previously defined likelihood, and assuming that $Se$ and $Sp$ are known and fixed for the screening test, the prevalence is estimated as the parameter of interest under a non-informative prior, defined as a distribution: $P \sim \text{Beta}(1, 1)$, Thus, the posterior distribution for $P$ is:

$$\mathbf{P}\left(P \mid \mathbf{D}\right) \propto \mathbf{P}\left(\mathbf{D} \mid P\right) \cdot \mathbf{P}(P)$$

Where $\mathbf{D}$ corresponds to true positives ($n_{11}$), false positives among those who tested positive in the test ($n_{01}$), total negatives according to the test ($n_{.0}$), Sensitivity and specificity of the test ($Se, Sp$).

$$\mathbf{P}\left(P \mid \mathbf{D}\right) = [(1 - Se)P + Sp(1 - P)]^{n_{00}}(Se \cdot P)^{n_{11}}((1 - Sp)(1 - P))^{n_{01}} \cdot \text{Beta}(P)$$

Finally, to obtain the posterior distribution of $P$, MCMC (Markov Chain Monte Carlo) methods can be implemented, with the calculation of a credible interval for $P$ derived from the estimated posterior distribution.

**Post-Stratification adjustment** The previous bayesian estimation can also include a post-stratification adjustment. In this case, as previously defined, for the $H$ strata formed, $P_h$ is estimated by stratifying the posterior of $P$:

$$\mathbf{P}\left(P_h \mid \mathbf{D}\right) = [(1 - Se)P_h + Sp(1 - P_h)]^{n_{00_h}}(Se \cdot P_h)^{n_{11_h}}((1 - Sp)(1 - P_h))^{n_{01_h}} \cdot \text{Beta}(P_h)$$

$$\tilde{P} = \sum_{h=1}^{H} w_h \cdot P_h$$

As previously defined, when estimating the prevalence $P_h$ for each stratum $h$ under a bayesian approach, a single point estimate is not obtained; instead, a posterior distribution for $P_h$ is generated. In this context, post-stratification must be based on these posterior distributions, allowing the incorporation of uncertainty into the stratum-adjusted estimates.

We have obtained $S$ samples from the posterior distribution of $P_h$ for $H$ strata. These samples for each stratum $h$ are denoted as:

$$P_h^{(1)}, P_h^{(2)}, \ldots, P_h^{(S)}$$

For each iteration $s$ in the posterior, we weight the prevalence $P_h^{(s)}$ by the proportion $w_h$ of stratum $h$ in the total population. In this way, the adjusted prevalence in iteration $s$ is:

$$\tilde{P}^{(s)} = \sum_{h=1}^{H} w_h \cdot P_h^{(s)}$$

This process is repeated for each sample $s = 1, \ldots, S$, resulting in a combined posterior distribution for the total prevalence adjusted by post-stratification:

$$\tilde{P}^{(1)}, \tilde{P}^{(2)}, \ldots, \tilde{P}^{(S)}$$

From the samples $\tilde{P}^{(s)}$, we can calculate any credible interval.

### Application of prevalence estimation for mutation in Familial Chylomicronemia with post-stratification adjustment

To evaluate the applicability of the proposed post-stratification adjustment, data from a two-phase descriptive cross-sectional study, in which one of the authors (JMEA) participated, was used[15]. The study began with an initial non-probabilistic sample of $n = 74,145$ adult subjects from the city of Pereira, Colombia. The progressive selection of patients was conducted in two phases. In the first phase, positive cases were identified through the application of the FCS (Familial Chylomicronemia Syndrome) clinical score, considered a fallible diagnostic test. This score, with a cutoff point $\geq 8$, has a reported sensitivity of 0.88 and a specificity of 0.85 for detecting mutations associated with familial chylomicronemia syndrome[16]. As a result of this stage, $n = 18$ positive subjects were identified.

In the second phase, these 18 individuals underwent a gold-standard test through molecular sequencing of the coding region of the genome (exome: $20,000$ genes), with greater than 98 coverage and a minimum depth of $20X$. This analysis included the genes APOA5, APOC2, GPIHBP1, LMF1, and LPL, covering all pathogenic variants in exonic regions, splicing sites, as well as small insertions and deletions. As a final result, $n = 6$ subjects were confirmed to have mutations associated with familial chylomicronemia syndrome.

The strata used for the post-stratification adjustment were based on sex and single-year age (ranging from 18 to 85 years or older), obtained from the official population projections published by the National Administrative Department of Statistics (DANE) of Colombia.

Frequentist estimates were performed using the R software, while bayesian estimates were conducted using Stan via the RStan library. In the MCMC simulation, the convergence of the Markov chains toward a stationary distribution was verified. A total of 1,000 iterations were performed across 4 chains, ensuring that $\hat{R} \leq 1.1$ as the convergence criterion.

## Results

Table **??** presents the results obtained for the estimation with and without post-stratification adjustment, corresponding to each scenario. Table 2 shows the lengths of each interval with and without post-stratification adjustment. According to each scenario, the following was found:

#### Scenario 1: Prevalence sstimation with full Verification using a gold standard

In the first scenario, the Agresti-Coull method showed a point estimate of prevalence of 0.0081% without post-stratification adjustment, with a confidence interval of $[0.0032\%, 0.0181\%]$. After applying the post-stratification adjustment, the estimate increased to $P = 0.4621\%$. This result highlights a significant increase in the prevalence estimate due to the correction for biases in the sample.

Table 1: Point and Interval Estimates with and without Post-Stratification Adjustment

| Scenario | Method | Unadjusted (%) | | | Adjusted (%) | | |
|---|---|---|---|---|---|---|---|
| | | P | Lower | Upper | P | Lower | Upper |
| Scenario 1 | Agresti-Coull | 0.0081 | 0.0032 | 0.0181 | 0.4621 | 0.4015 | 0.5226 |
| | Bayesian (beta-binomial) | 0.0088 | 0.0028 | 0.0156 | 0.2417 | 0.2014 | 0.2886 |
| Scenario 2 | Lang | 0.0000 | 0.0000 | 0.0162 | 0.0000 | 0.0000 | 0.0836 |
| | Bayesian-Flor | 0.0016 | 0.0000 | 0.0058 | 0.2740 | 0.2280 | 0.3286 |
| Scenario 3 | Frequentist-Thomas | 0.0094 | 0.0020 | 0.0169 | 0.0075 | 0.0014 | 0.0135 |
| | Bayesian-Thomas | 0.0104 | 0.0045 | 0.0201 | 0.2797 | 0.2328 | 0.3352 |

The length of the confidence interval without post-stratification was 0.0149, while it increased to 0.1211 after the adjustment, reflecting a notable increment. This result aligns with the objective of post-stratification: to account for variability previously unconsidered due to the use of a non-probabilistic sample.

Similarly, the bayesian approach initially presented a prevalence of $P = 0.0088\%$, with a 95% credible interval ($CIr_{95\%}$) of $[0.0028\%, 0.0156\%]$. With the adjustment, the estimate increased to $P = 0.2417\%$, with an adjusted interval of $[0.2014\%, 0.2886\%]$. The interval length increased from 0.0128 without post-stratification to 0.0872 with the adjustment, reflecting the same trend observed with the frequentist method.

| Method | Without Adjustment | With Adjustment |
|---|---|---|
| Agresti-Coull | 0.0149 | 0.1212 |
| Bayesian (Beta-Binomial) | 0.0128 | 0.0872 |
| Lang | 0.0162 | 0.0836 |
| Flor - Bayesian | 0.0058 | 0.1005 |
| Frequentist - Thomas | 0.0149 | 0.0121 |
| Bayesian - Thomas | 0.0156 | 0.1025 |

Table 2: Comparison of Interval Lengths by Estimation Method

#### Scenario 2: Estimation with a diagnostic test of Known sensitivity and specificity

In this scenario, the Lang method presented an initial prevalence of $P = 0.000\%$ without post-stratification adjustment, indicating severe underestimation due to the limited sensitivity of the method. After applying post-stratification, the estimator increased to 0.0836%, highlighting the importance of adjustment to correct bias in designs with imperfect diagnostic tests. The interval length without post-stratification was 0.0162, while with post-stratification, it increased to 0.0836.

The bayesian method proposed by Flor showed an initial prevalence of $P = 0.0016\%$, with a 95% credible interval ($CIr_{95\%}$) of $[0.0000\%, 0.0058\%]$. After adjustment, the estimator increased significantly, as shown in Table **??**; this result emphasizes the flexibility of the bayesian approach in addressing the limitations of fallible diagnostic tests and adjusting prevalence estimates. In this case, the interval length increased from 0.0058 to 0.1006. This significant increase is due to the bayesian method explicitly incorporating the additional uncertainty associated with the non-probabilistic sample.

### Scenario 3: Estimation when disease status is verified only among positives

In this final scenario, the frequentist method of Thomas estimated an initial prevalence of $P = 0.0094\%$, with a 95% confidence interval ($CI_{95\%}$) of $[0.0020\%, 0.0169\%]$. After applying the correction by the inverse of inclusion in the sample, the estimator decreased, resulting in a narrower interval. The interval length was 0.0149 without post-stratification and slightly decreased to 0.0121 with the adjustment. This result is unique to this scenario, as post-stratification appears to reduce variability by addressing the specific limitations of partial verification. This change reflects the adjustment's ability to improve precision in partial verification designs, a situation more representative of daily practice.

The bayesian approach by Thomas et al. initially estimated a prevalence of $P = 0.0104\%$, with a 95% credible interval ($CIr_{95\%}$) of $[0.0045\%, 0.0201\%]$. The adjusted estimator using the post-stratification method increased, demonstrating the ability of bayesian estimates to integrate additional information and adjust prevalence estimates in partial verification scenarios. The length of the credible interval increased from 0.0156 to 0.1024 with the adjustment. This considerable increase is consistent with the bayesian approach's ability to model the additional uncertainty introduced by post-stratification.

## Discussion

Accurate prevalence estimation in epidemiological studies is essential for public health planning, particularly for rare diseases or conditions with low prevalence. The methods evaluated in this study address key challenges such as the use of imperfect diagnostic tests, partial verification of disease status, and reliance on non-probabilistic samples. These limitations, if not corrected, can lead to biased estimates and inadequate public health decisions.

The findings of the study highlight the need to implement robust methodological adjustments to improve the accuracy of prevalence estimates. The comparison of frequentist and bayesian approaches revealed significant differences in their ability to address inherent limitations in study designs and diagnostic tests. These observations align with McNamee et al.[17], who noted that two-phase designs, despite their complexity, can optimize precision and reduce costs in prevalence studies.

In the first scenario, with full verification using a gold standard, both methodological approaches successfully corrected initial biases. However, post-stratification adjustment had a significant impact by incorporating additional information on the distribution of population variables. This result is consistent with Holt et al.[12], who reported the effectiveness of post-stratification in addressing the lack of representativeness in non-probabilistic samples. Furthermore, the bayesian approach adjusted with post-stratification generated wider confidence intervals compared to the Agresti-Coull method, as observed in the study by Flor et al.[14]. This behavior reflects the bayesian approach's ability to explicitly model uncertainty, making it a particularly useful alternative in contexts of high variability.

Scenario 2, which uses a diagnostic test with known sensitivity and specificity, underscores the importance of considering the quality of initial tests in study design. As noted by Shrout and Newman[18], the limited sensitivity of screening methods can significantly underestimate prevalence, a finding consistent with the unadjusted results obtained in this research. However, post-stratification adjustments and bayesian estimates successfully corrected these limitations, aligning with previous studies that highlight the flexibility of the bayesian approach in incorporating prior information and improving precision.

Regarding interval lengths, unadjusted intervals were observed to be shorter, which could create false confidence in underrepresented estimates. In contrast, post-stratification adjustment, while increasing interval lengths, corrected the inherent bias in the data and improved population representativeness.

A particular situation, well-documented in the literature, became evident in the interval proposed by Lang and was also observed in this study. When using the adjusted prevalence estimator proposed by Rogan and Gladen, there is a possibility that the adjusted estimate may fall outside the allowed range $[0, 1]$, especially for negative values when the true prevalence is very low. This situation, previously studied by authors like Viana[19] and Hasselt[20], was reflected in the present study, where the reported prevalence (Table **??**) was zero, consistently underestimating the true prevalence in low-prevalence scenarios.

In contrast, the bayesian approach proved beneficial, as evidenced by the results of this study. By employing a prior distribution $P \sim \text{Beta}(\alpha, \beta)$, this approach ensures that the estimated prevalence remains within the allowed range $[0, 1]$, even in situations of extremely low prevalence. This advantage was also reflected in the bayesian approach adjusted by post-stratification used in Scenario 2, where a more robust and representative estimate was achieved, effectively addressing the limitations observed with frequentist methods.

In Scenario 3, where disease status is verified only among positives, the results confirm the advantages of combining maximum likelihood methods and post-stratification adjustments. The frequentist approach, while traditional, showed a reduction in interval length after adjustment, highlighting its ability to address specific biases in such designs. On the other hand, the bayesian approach provided wider intervals but with better representation of uncertainty, consistent with the observations of Flor et al.[14] in partial verification scenarios.

Regarding interval lengths, unadjusted intervals were shorter, potentially leading to overconfidence in underrepresented estimates. In contrast, post-stratification adjustment, while increasing interval lengths, corrected inherent bias in the data and improved population representativeness.

A different pattern in confidence interval lengths was observed in this scenario. The frequentist approach showed a reduction in interval length after adjustment, while bayesian approaches presented wider intervals post-adjustment, reflecting their ability to incorporate the inherent uncertainty of partial verification designs using a gold standard test. These results underscore the importance of selecting the most appropriate method based on the study context, as suggested in previous research[17].

Finally, Bayer et al.[21] propose an innovative approach for confidence interval estimation in prevalence derived from complex sampling surveys, based on melding methods that use gamma and beta distributions. While their method is particularly robust in addressing bias introduced by imperfect sensitivity and specificity of diagnostic tests, its context is limited to probabilistic samples, ensuring population representativeness from the sampling design but not directly addressing bias in non-probabilistic samples or post-stratification adjustments.

In contrast, the approach in this study focuses on contexts where samples are non-probabilistic, presenting an additional challenge. The integration of bayesian methods with post-stratification adjustments allows for correction of biases related to population representativeness, a dimension not explicitly addressed in Bayer's method. This methodological difference makes the present work complementary to that of Bayer et al., offering a solution adaptable to scenarios where practical limitations, such as the inability to use a probabilistic sampling design or partial verification of diagnostic tests, require methods that combine representativeness and flexibility.

Despite the favorable findings, this study has some limitations. The reliance on known metrics, such as the sensitivity and specificity of diagnostic tests, may restrict the applicability of the methods in contexts where these parameters are not well established. Additionally, while robust, the bayesian approach critically depends on the quality of prior information. In the absence of reliable data, the use of non-informative priors may lead to less precise estimates.

Another important aspect for future improvement in this work is the evaluation of coverage for intervals adjusted by post-stratification, as well as the assessment of bias associated with point estimators obtained in different scenarios. These evaluations are fundamental to validating the robustness of the proposed methods, especially in comparison to alternative methodologies.

The results have important implications for public health practice and epidemiological research. Adjusting prevalence estimates through post-stratification significantly improves representativeness in contexts with non-probabilistic samples and imperfect diagnostic tests. Bayesian methods have proven to be valuable tools in scenarios with high uncertainty, particularly when incorporating prior information.

## References


[1] Arya, R.; Antonisamy, B.; Kumar, S. Sample Size Estimation in Prevalence Studies. *Springer Science+Business Media*, 2012, *79*, 1482–1488. https://doi.org/10.1007/s12098-012-0763-3.

[2] Lewis, F.; Torgerson, P. R. A Tutorial in Estimating the Prevalence of Disease in Humans and Animals in the Absence of a Gold Standard Diagnostic. *BioMed Central*, 2012, *9*. https://doi.org/10.1186/1742-7622-9-9.

[3] Agresti, A.; Coull, B. A. Approximate Is Better Than "Exact" for Interval Estimation of Binomial Proportions. *Taylor & Francis*, **1998**, *52* (2), 119–126. https://doi.org/https://doi.org/10.1080/00031305.1998.10480550.

[4] Rogan, W. J.; Gladen, B. Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, **1978**, *107* (1), 71–76. https://doi.org/10.1093/oxfordjournals.aje.a112510.

[5] Izbicki, R.; Diniz, M. A.; Bastos, L. S. Sensitivity and Specificity in Prevalence Studies: The Importance of Considering Uncertainty. *Elsevier BV*, 2020, *75*, e2449–e2449. https://doi.org/10.6061/clinics/2020/e2449.

[6] Reiczigel, J.; Földi, J.; Ózsvári, L. Exact Confidence Limits for Prevalence of a Disease with an Imperfect Diagnostic Test. *Cambridge University Press*, **2010**, *138* (11), 1674–1678. https://doi.org/https://doi.org/10.1017/s0950268810000385.

[7] Láng, Z.; Reiczigel, J. Confidence Limits for Prevalence of Disease Adjusted for Estimated Sensitivity and Specificity. *Elsevier BV*, **2014**, *113* (1), 13–22. https://doi.org/https://doi.org/10.1016/j.prevetmed.2013.09.015.

[8] Thomas, E. F.; Peskoe, S. B.; Spiegelman, D. Prevalence Estimation When Disease Status Is Verified Only Among Test Positives: Applications in HIV Screening Programs. *Wiley*, **2017**, *37* (7), 1101–1114. https://doi.org/https://doi.org/10.1002/sim.7568.

[9] Elliott, M. R.; Valliant, R. Inference for Nonprobability Samples. *Statistical Science*, **2017**, *32*, 249–264. https://doi.org/10.1214/16-STS598.

[10] Lohr, S. L. *Sampling: Design and Analysis: Design and Analysis*; 2019.

[11] Smith, T. M. F. Post-Stratification, 1990.

[12] Holt, D.; Smith, T. M. F. Post Stratification. *Journal of the Royal Statistical Society. Series A (General)*, **1979**, *142*, 33–46. https://doi.org/10.2307/2344652.

[13] Cai, T. T. One-Sided Confidence Intervals in Discrete Distributions. *Elsevier BV*, **2005**, *131* (1), 63–88. https://doi.org/https://doi.org/10.1016/j.jspi.2004.01.005.

[14] Flor, M.; Weiß, M.; Selhorst, T.; Müller-Graf, C.; Greiner, M. Comparison of Bayesian and Frequentist Methods for Prevalence Estimation Under Misclassification. *BioMed Central*, **2020**, *20* (1). https://doi.org/https://doi.org/10.1186/s12889-020-09177-4.

[15] Rodriguez, F. H.; Estrada, J. M.; Quintero, H. M. A.; Nogueira, J. P.; Porras-Hurtado, G. L. Analyses of Familial Chylomicronemia Syndrome in Pereira, Colombia 2010–2020: A Cross-Sectional Study. *Lipids in Health and Disease*, **2023**, *22*, 43. https://doi.org/10.1186/s12944-022-01768-x.

[16] Moulin, P.; Dufour, R.; Averna, M.; Arca, M.; Cefalù, A. B.; Noto, D.; D'Erasmo, L.; Costanzo, A. D.; Marçais, C.; Walther, L. A. A.-S.; et al. Identification and Diagnosis of Patients with Familial Chylomicronaemia Syndrome (FCS): Expert Panel Recommendations and Proposal of an "FCS Score." *Atherosclerosis*, **2018**, *275*, 265–272. https://doi.org/10.1016/j.atherosclerosis.2018.06.814.

[17] McNamee, R. Two-Phase Sampling for Simultaneous Prevalence Estimation and Case Detection. *Biometrics*, **2004**, *60* (3), 783–792.

[18] Shrout, P. E.; Newman, S. C. Design of Two-Phase Prevalence Surveys of Rare Disorders. *Biometrics*, **1989**, *45* (2), 549–555.

[19] M. A. G. Viana, V. R.; Levy, P. S. Bayesian Analysis of Prevalence from the Results of Small Screening Samples. *Communications in Statistics - Theory and Methods*, **1993**, *22* (2), 575–585. https://doi.org/10.1080/03610929308831038.

[20] Hasselt, M. van; Bollinger, C. R.; Bray, J. W. A Bayesian Approach to Account for Misclassification in Prevalence and Trend Estimation. *Journal of Applied Econometrics*, **2022**, *37* (2), 351–367. https://doi.org/https://doi.org/10.1002/jae.2879.

[21] Bayer, D. M.; Fay, M. P.; Graubard, B. I. Confidence Intervals for Prevalence Estimates from Complex Surveys with Imperfect Assays. *Statistics in Medicine*, **2023**, *42* (11), 1822–1867. https://doi.org/https://doi.org/10.1002/sim.9701.

# PREVALENCE ESTIMATION IN INFECTIOUS DISEASES WITH IMPERFECT TESTS: A COMPARISON OF FREQUENTIST AND BAYESIAN LOGISTIC REGRESSION METHODS WITH MISCLASSIFICATION CORRECTION

**Jorge Mario Estrada Alvarez**
Caja de Compensacion Familiar de Risaralda
Salud - Comfamiliar
Pereira, PA 66003
jestradaa@comfamiliar.com

**Henan F. Garcia**
SISTEMIC Research Group
Faculty of engineering
University of Antioquia
Medellin, PA 15213
hernanf.garcia@udea.edu.co

**Miguel Ángel Montero-Alonso**
Department of Statistics and Operational Research
University of Granada
Granada, Spain PA 18071
mmontero@ugr.edu

**Juan de Dios Luna del Castillo**
Department of Statistics and Operational Research
University of Granada
Granada, Spain PA 18071
jdluna@ugr.edu

April 22, 2025

**ABSTRACT**

Accurate estimation of disease prevalence is critical for informing public health strategies. Imperfect diagnostic tests can lead to misclassification errors, such as false positives and false negatives, which may skew estimates if not properly addressed. This study compared four statistical methods for estimating the prevalence of sexually transmitted infections (STIs) and the factors associated with them, while incorporating corrections for misclassification. The methods examined were: (i) standard logistic regression with external correction using known sensitivity and specificity; (ii) the Liu et al. model, which jointly estimates false positive and false negative rates; (iii) Bayesian logistic regression with external correction; and (iv) a Bayesian model with internal correction that utilizes informative priors on diagnostic accuracy. Data were obtained from 11,452 individuals participating in a voluntary screening campaign for HIV, syphilis, and hepatitis B from 2020 to 2024. Prevalence estimates and regression coefficients were compared across models using metrics of relative change from crude estimates, confidence interval width, and coefficient variability. The Liu model yielded higher prevalence estimates but exhibited wider intervals and convergence issues for low-prevalence conditions. In contrast, the Bayesian model with internal correction (BEC) provided intermediate estimates with the narrowest confidence intervals and more stable intercepts, reflecting an improved estimation of baseline prevalence. Prior distributions either informative or weakly informative contributed to regularization, particularly in small-sample or rare-event settings. Furthermore, accounting for misclassification impacts both prevalence estimates and covariate associations. Although the

Liu model presents theoretical advantages compared to standard regression, its practical limitations, particularly in cases with sparse data, diminish its effectiveness. In contrast, Bayesian models that incorporate misclassification correction provide a robust and flexible alternative in low-prevalence situations, proving especially useful when accurate estimates are necessary despite diagnostic uncertainties.

***K*eywords** Prevalence estimation · Missclassification · Logistic regression, Infectious disease

## 1 Introduction

Accurate estimation of the prevalence of infectious diseases is essential for guiding the planning, implementation, and evaluation of public health strategies. Such estimates are typically based on the large-scale application of diagnostic tests, whose performance is rarely perfect. As a result, misclassification errors such as false positives and false negatives occur, introducing bias and reducing the precision of the resulting estimates [1].

Accounting for possible misclassification in the binary outcome variable is crucial, as ignoring it can lead to substantial bias in parameter estimates. Suppose misclassification is suspected, and information on the misclassification rates (sensitivity and specificity) is available or can be reasonably assumed. In that case, methods that explicitly incorporate this information should be employed to correct the estimated prevalence and avoid bias toward the null value in the resulting estimates [1].

Logistic regression is a commonly used tool for prevalence estimation that allows for covariate adjustment. Moreover, correcting for measurement error in logistic regression models improves the estimation of the prevalence of a condition while allowing the direct inclusion of relevant demographic and epidemiological covariates [2]. Nonetheless, methods that integrate both elements are required to adequately address misclassification's challenges and benefit from the flexibility to include covariates of interest. This combination enables the derivation of prevalence estimates with greater validity, particularly when multiple infectious diseases are epidemiologically correlated.

Various strategies have been developed to handle diagnostic errors in prevalence estimation, ranging from post-hoc adjustments via marginal standardization of predicted probabilities [3], which provides an extrinsic adjustment based on the model's predictions. In this same line of work, Liu et al. [4] propose a logistic regression model with correction for misclassification in the binary outcome variable. The method incorporates parameters for the false positive (FP) and false negative (FN) rates into the standard logistic regression model, thereby enabling the estimation of coefficients adjusted for these error rates. The model proposed by Liu et al. includes different variants depending on whether one or both types of error need to be estimated, assuming these errors may be either equal or distinct.

Valle et al. [5] propose a Bayesian inference approach to logistic regression for correcting bias in prevalence estimates when using imperfect tests. However, their results primarily focus on the bias induced in the estimation of risk factor associations rather than on the method for adjusting prevalence estimates in the presence of covariates.

This study directly addresses the prevalence estimation of diseases using imperfect tests while accounting for the need to control for epidemiological and/or demographic covariates. The objective was to compare four statistical methodologies for prevalence estimation under a practical framework for adjusted and corrected point estimates, emphasizing the importance of an integrated approach that considers diagnostic uncertainty and covariate adjustment. The proposed methodological approaches (frequentist and Bayesian) were: (i) standard logistic regression for prevalence estimation adjusted for covariates, with correction based on known test sensitivity and specificity, (ii) modified logistic regression for simultaneous estimation of test errors as proposed by Liu et al. [4], (iii) standard logistic regression with Bayesian inference and extrinsic correction using known sensitivity and specificity, and (iv) a Bayesian logistic regression model with intrinsic correction of misclassification error via prior distribution elicitation for classification errors [5]. The methods are compared using a representative sample of 11452 individuals simultaneously evaluated for HIV and syphilis seroprevalence.

## 2 Methods

This study aims to estimate the prevalence of low-prevalence sexually transmitted infections (STIs) while correcting for diagnostic misclassification. We implemented and compared four logistic regression models that differ in their handling of misclassification and statistical inference frameworks.

### 2.1 Crude Prevalence Estimation

Crude prevalence was estimated using the Rogan–Gladen method [6], which adjusts the observed proportion of positive tests based on known sensitivity $(Se)$ and specificity $(Sp)$ of the diagnostic test:

$$\hat{P}_{\text{adj}} = \frac{\hat{P}_{\text{obs}} + Sp - 1}{Se + Sp - 1}. \tag{1}$$

This estimator assumes fixed, known values of Se and Sp and yields an unbiased prevalence estimate under nondifferential misclassification.

### 2.2 Adjusted Prevalence Estimation via Regression Models

Four logistic regression models were applied to estimate adjusted prevalence for HIV and syphilis, accounting for both covariates and test misclassification. Covariates included age, sex, test results for other infections, and key population indicators. Table 1 defines the regression coefficients associated with each covariate.

Table 1: Coefficient assignment for covariates in the models

| Covariate | Assigned Coefficient |
|---|---|
| Intercept | $\beta_0$ |
| Age (years) | $\beta_1$ |
| Sex | $\beta_2$ |
| Syphilis test* | $\beta_3$ |
| Hepatitis B test | $\beta_4$ |
| MSM** | $\beta_5$ |
| LGTBI population | $\beta_6$ |
| Other populations | $\beta_7$ |
| Sex worker | $\beta_8$ |

* The coefficient $\beta_3$ for the syphilis prevalence models represents the result of the HIV test.
** MSM: Men who have sex with men.

#### 2.2.1 Standard Logistic Regression (STD)

The standard model estimates the probability of disease given covariates using the logistic function:

$$P(Y_i = 1 \mid X_i) = \pi(X_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}, \tag{2}$$

$$\eta_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij}. \tag{3}$$

Parameters are estimated via maximum likelihood. Adjusted prevalence is computed by correcting the predicted values using the Rogan–Gladen formula.

#### 2.2.2 Bayesian Logistic Regression with External Correction (BC)

In the Bayesian extension, regression coefficients are assigned prior distributions. Following Newman et al. [7], priors are:

$$\beta_j \sim \mathcal{N}\left(0, \frac{\pi^2}{3(p+1)}\right), \quad j = 0, \ldots, p. \tag{4}$$

Posterior inference is conducted via Markov chain Monte Carlo (MCMC). Point estimates are then corrected post hoc using known Se and Sp.

#### 2.2.3 Logistic Regression with Simultaneous Estimation of Misclassification (Liu)

The Liu model [4] introduces latent variables $\tilde{Y}_i$ for true disease status, linking them to observed outcomes $Y_i$ via misclassification probabilities, following traditional assumptions on non-differential misclassification [8, 9]:

$$P(Y_i = 1 \mid \tilde{Y}_i = 0) = r_0, \tag{5}$$
$$P(Y_i = 0 \mid \tilde{Y}_i = 1) = r_1. \tag{6}$$

The conditional probability is then:

$$P(Y_i = 1 \mid X_i) = r_0 + (1 - r_0 - r_1) \cdot \frac{1}{1 + \exp(-\eta_i)}. \tag{7}$$

Model parameters $(\beta, r_0, r_1)$ are jointly estimated via maximum likelihood.

#### 2.2.4 Bayesian Logistic Regression with Internal Correction (BEC)

This hierarchical model [5] assumes the true outcome $\tilde{Y}_i$ follows a logistic model:

$$\tilde{Y}_i \sim \text{Bernoulli}(\pi_i), \tag{8}$$
$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}. \tag{9}$$

The observed test result $Y_i$ depends on $\tilde{Y}_i$ and the test characteristics:

$$P(Y_i = 1 \mid \tilde{Y}_i = 1) = Se, \tag{10}$$
$$P(Y_i = 1 \mid \tilde{Y}_i = 0) = 1 - Sp. \tag{11}$$

Priors are specified as:

$$Y_i \sim \text{Bernoulli}(Se) \quad \text{if } \tilde{Y}_i = 1, \tag{12}$$

$$Y_i \sim \text{Bernoulli}(1 - Sp) \quad \text{if } \tilde{Y}_i = 0, \tag{13}$$

Inference is conducted via MCMC to obtain posterior distributions of prevalence and covariate effects.

### 2.3 Evaluation Criteria

Models were compared using the following metrics, according to the approach suggested by Morris et al., [10]:

- **Relative change in adjusted prevalence** with respect to crude estimates.
- **Interval width** of confidence/credible intervals.
- **Stability of regression coefficients**, measured by standard errors or posterior variance.

Table 2: Descriptive characteristics of the study population (n = 11,452)

| **Characteristic** | |
|---|---|
| Age (mean [SD]) | 34 (14) |
| Male sex | 5759 (50%) |
| HIV reactivity | 155 (1.4%) |
| Syphilis reactivity | 905 (7.9%) |
| Hepatitis B reactivity | 8 (<0.1%) |
| **Key population type** | |
| MSM | 3248 (28%) |
| LGTBIQ | 224 (2.0%) |
| Other populations | 193 (1.7%) |
| General population | 6574 (57%) |
| Sex worker | 1213 (11%) |

Continuous variables are presented as mean (standard deviation), and categorical variables as n (percentage).

### 2.4 Data Source and Variables

Data were obtained from 11,452 individuals screened between 2020–2024. Covariates included age, sex, syphilis and hepatitis B status, and population group classification (e.g., MSM, sex worker). Tests used were the Determine HIV Early Detect (Se = 0.975, Sp = 0.999) and Bioline Syphilis 3.0 (Se = 0.964, Sp = 0.974).

### 2.5 Software

All analyses were performed in `R` version 4.4.3 [11]. Frequentist models used base `glm()` and `optim()` routines; Bayesian models were estimated using `rstan` and `rstanarm` [12, 13].

## 3 Results

### 3.1 Sample Characteristics

The study population included 11452 individuals, whose demographic and serological characteristics are summarized in Table 2. The mean age was 34 years (standard deviation: 14). The sex distribution was balanced, with 5759 men corresponding to 50% of the sample.

Regarding serological results, 155 participants (1.4%) tested reactive for HIV, 905 (7.9%) for syphilis, and 8 (<0.1%) for hepatitis B.

Concerning key population categories, the most significant proportion corresponded to the general population, with 6574 individuals (57%), followed by men who have sex with men (MSM), with 3248 (28%), and sex workers, with 1213 (11%). Additionally, individuals identified as part of the LGTBIQ community (n = 224; 2.0%) and other populations (n = 193; 1.7%) were included.

### 3.2 Comparison of Adjusted Prevalence Estimates

Table 3 presents the adjusted prevalence estimates for HIV and syphilis under different regression models with misclassification correction. Notable variations in the estimates were observed depending on the methodological approach used.

For HIV, the crude prevalence was 1.39%. The standard model (STD), which applies an external correction based on known sensitivity and specificity, reduced this estimate to 0.73%. In contrast, the Liu model, which simultaneously estimates diagnostic error parameters, yielded a prevalence of 1.66%, representing a relative increase of 127.69%

Table 3: Adjusted prevalence estimates for HIV and syphilis according to different regression models

| Disease | Model | Adjusted P | Lower | Upper | Change vs. Crude (%) | Change vs. STD (%) | CI Width |
|---|---|---|---|---|---|---|---|
| HIV | Crude | 0.0139 | 0.0117 | 0.0160 | – | – | 0.0043 |
| | STD | 0.0073 | 0.0055 | 0.0097 | -47.42 | – | 0.0042 |
| | Liu | 0.0166 | 0.0131 | 0.0212 | 19.72 | 127.69 | 0.0081 |
| | BC | 0.0095 | 0.0078 | 0.0116 | -31.57 | 30.15 | 0.0037 |
| | BEC | 0.0100 | 0.0078 | 0.0127 | -28.24 | 36.47 | 0.0049 |
| Syphilis | Crude | 0.0567 | 0.0514 | 0.0620 | – | – | 0.0106 |
| | STD | 0.0277 | 0.0223 | 0.0336 | -51.21 | – | 0.0113 |
| | Liu | 0.1155 | 0.1064 | 0.1252 | 103.64 | 317.39 | 0.0187 |
| | BC | 0.0340 | 0.0288 | 0.0391 | -40.08 | 22.82 | 0.0103 |
| | BEC | 0.0414 | 0.0379 | 0.0452 | -26.94 | 49.74 | 0.0073 |

*Note:* STD = standard model with external correction; Liu = model with simultaneous estimation of error parameters; BC = Bayesian model without internal correction; BEC = Bayesian model with misclassification correction.

Table 4: Comparison of regression coefficients for HIV according to the implemented model

| Coefficient | STD | Liu | Liu Change (%) | BC | BEC | BEC Change (%) |
|---|---|---|---|---|---|---|
| $\beta_0$ | -5.864 | -4.648 | -20.73 | -5.418 | -4.717 | -12.93 |
| $\beta_1$ | -0.001 | 0.002 | -325.76 | -0.001 | -0.015 | 2798.64 |
| $\beta_2$ | 0.991 | 0.356 | -64.09 | 0.749 | 0.512 | -31.68 |
| $\beta_3$ | 1.946 | 1.492 | -23.33 | 1.782 | 1.892 | 6.18 |
| $\beta_4$ | 1.547 | 1.545 | -0.14 | 0.433 | 0.454 | 4.86 |
| $\beta_5$ | 0.883 | 0.617 | -30.15 | 0.754 | 0.717 | -4.96 |
| $\beta_6$ | 1.214 | 0.604 | -50.23 | 0.664 | 0.513 | -22.74 |
| $\beta_7$ | 0.673 | 0.428 | -36.33 | 0.326 | 0.319 | -2.01 |
| $\beta_8$ | 0.516 | 0.060 | -88.44 | 0.191 | -0.138 | -171.98 |

compared to the STD model. Meanwhile, the Bayesian model with misclassification correction (BEC) estimated a prevalence of 1.00%, with a confidence interval width of 0.0049.

In the case of syphilis, the crude prevalence was 5.67%. The STD model produced a considerably lower adjusted estimate (2.77%), while the Liu model estimated a prevalence of 11.55%—more than twice the crude value and with the widest confidence interval (0.0187). In contrast, the BEC model estimated a prevalence of 4.14%, with the narrowest confidence interval (0.0073), indicating the highest relative precision among the models compared.

Overall, the Bayesian models demonstrated a better balance between misclassification error adjustment and estimate precision—particularly the BEC model, which yielded intermediate prevalence values and the narrowest confidence intervals for both outcomes.

### 3.3 Comparison of Regression Coefficients for HIV Models

Table 4 displays the estimated coefficients for the logistic regression model for HIV across four approaches: the standard model (STD), the Liu model (simultaneous adjustment of misclassification parameters), the Bayesian model without internal correction (BC), and the Bayesian model with misclassification correction (BEC). Relative percentage changes concerning the standard model are also included.

The intercept showed a reduction of 20.73% in the Liu model and 12.93% in the Bayesian model with correction, suggesting differences in the estimation of the baseline prevalence level when diagnostic error is accounted for.

Overall, substantial differences were observed in the magnitude and direction of some coefficients depending on the adjustment method used. The Liu model showed notable reductions compared to the standard model across most coefficients, with a change of up to -88.44% for the coefficient associated with the sex worker population ($\beta_8$) and

Table 5: Comparison of standard errors for HIV model coefficients between Liu and Bayesian model with correction (BEC)

| Coefficient | SE (Liu) | SE (BEC) | Relative Change |
|---|---|---|---|
| $\beta_0$ | 0.694 | 0.237 | 7.539 |
| $\beta_1$ | 0.004 | 0.006 | -0.639 |
| $\beta_2$ | 0.229 | 0.228 | 0.008 |
| $\beta_3$ | 0.311 | 0.168 | 2.424 |
| $\beta_4$ | 0.847 | 0.553 | 1.345 |
| $\beta_5$ | 0.230 | 0.200 | 0.327 |
| $\beta_6$ | 0.415 | 0.397 | 0.089 |
| $\beta_7$ | 0.387 | 0.402 | -0.073 |
| $\beta_8$ | 0.172 | 0.311 | -0.695 |

-64.09% for male sex ($\beta_2$), indicating significant adjustment when simultaneously correcting for test sensitivity and specificity.

In contrast, the Bayesian model with misclassification correction (BEC) adjusted the magnitude and direction of specific coefficients. The most striking case is $\beta_8$, which changed from a positive value in the standard model (0.516) to negative (-0.138) in BEC, representing a relative change of approximately 1.7 times compared to the BC model.

Another relevant finding is observed in the coefficient for age group ($\beta_1$), which changed from -0.001 in the standard model to -0.015 in BEC, representing a relative increase of approximately 3 times compared to its value in the BC model. Although its magnitude remains small, this change reflects the sensitivity of the age effect to how misclassification is modeled.

The covariate associated with syphilis reactivity ($\beta_3$) maintained a positive direction in all models, with a slight increase in BEC (1.892) compared to STD (1.946), suggesting the stability of this effect despite adjustment for measurement error.

Table 5 compares the standard errors (SE) of the estimated coefficients for the HIV model between the Liu approach and the Bayesian model with misclassification correction (BEC). Overall, notable differences were observed in estimator precision depending on the methodological approach.

The BEC model showed a marked reduction in the standard error of the intercept (from 0.694 to 0.237), corresponding to a relative precision gain of more than sevenfold, suggesting increased stability in estimating the corrected baseline prevalence. Similarly, reductions were observed in the standard errors of the covariate associated with syphilis reactivity, with a 2.42-fold relative precision gain compared to the Liu model. A substantial improvement was also observed for the hepatitis B covariate.

In contrast, some covariates—such as age group and sex worker—showed increased standard errors under the BEC model, which may reflect greater uncertainty associated with their effects, possibly due to low frequency or interactions with other variables.

These findings indicate that the BEC model not only adjusts point estimates of the coefficients but also differentially improves their precision, particularly for components most relevant to population-level inference. This reinforces its usefulness in scenarios where diagnostic error may compromise both the validity and precision of traditional models.

### 3.4 Comparison of Regression Coefficients for Syphilis Models

Table 6 presents the estimated coefficients for the logistic regression model for syphilis under the different models evaluated. Differences in magnitude and direction are observed across approaches, particularly when comparing models corrected for misclassification with the standard model.

Table 6: Comparison of regression coefficients for syphilis according to the implemented model

| Coefficient | STD | Liu | Liu Change (%) | BC | BEC | BEC Change (%) |
|---|---|---|---|---|---|---|
| $\beta_0$ | -4.890 | -3.164 | -35.30 | -4.766 | -5.322 | 11.67 |
| $\beta_1$ | 0.036 | 0.022 | -38.72 | 0.035 | 0.037 | 5.04 |
| $\beta_2$ | 0.615 | 0.159 | -74.15 | 0.575 | 0.728 | 26.62 |
| $\beta_3$ | 1.986 | 1.525 | -23.22 | 1.839 | 2.061 | 12.09 |
| $\beta_4$ | 2.140 | 1.743 | -18.55 | 0.747 | 0.719 | -3.71 |
| $\beta_5$ | 0.954 | 0.597 | -37.45 | 0.903 | 1.004 | 11.24 |
| $\beta_6$ | 1.473 | 0.672 | -54.39 | 1.222 | 1.371 | 12.19 |
| $\beta_7$ | 1.470 | 0.921 | -37.39 | 1.288 | 1.439 | 11.77 |
| $\beta_8$ | 1.825 | 0.886 | -51.43 | 1.691 | 1.990 | 17.67 |

Table 7: Comparison of standard errors for syphilis model coefficients: Liu model vs. Bayesian model with correction (BEC)

| Coefficient | SE (Liu) | SE (BEC) | Relative Change |
|---|---|---|---|
| $\beta_0$ | 0.367 | 0.166 | 3.884 |
| $\beta_1$ | 0.004 | 0.003 | 0.627 |
| $\beta_2$ | 0.062 | 0.165 | -0.857 |
| $\beta_3$ | 0.187 | 0.184 | 0.031 |
| $\beta_4$ | 0.675 | 0.516 | 0.711 |
| $\beta_5$ | 0.109 | 0.126 | -0.255 |
| $\beta_6$ | 0.193 | 0.285 | -0.544 |
| $\beta_7$ | 0.191 | 0.230 | -0.311 |
| $\beta_8$ | 0.168 | 0.164 | 0.047 |

The Liu model showed a systematic reduction in most coefficients, with the most notable decreases observed for male sex ($\beta_2$), with a change of -74.15%, and the sex worker category ($\beta_8$), with a decrease of 51.43%. These results may reflect an overestimation of associations in the standard model when misclassification is not accounted for.

In contrast, the Bayesian model with misclassification correction (BEC) preserved the positive direction of most effects and even increased their magnitude compared to the standard model. For example, the coefficient for sex workers increased from 1.825 to 1.990, representing a 17.67% rise, while the effect of male sex increased by 26.62%. These changes suggest that, once misclassification is corrected, the association between these factors and syphilis infection strengthens—an insight that may be epidemiologically and programmatically significant.

Conversely, the effect of hepatitis B reactivity ($\beta_4$) showed a decrease in the BEC model relative to the standard model, suggesting a potential adjustment toward a more conservative effect. The age group variable ($\beta_1$) maintained a positive and stable effect across all models.

Table 7 shows moderate variations in the precision of the estimators between the two models, with a pattern differing from that observed in the analysis for HIV.

The BEC model demonstrated a substantial improvement in the intercept's precision, with a reduction in the standard error from 0.367 to 0.166, representing a relative precision gain of more than 3.8-fold. This finding suggests that the estimation of the baseline prevalence for syphilis is more stable under a Bayesian approach with misclassification correction.

However, for several covariates—such as sex, LGTBIQ population type, and other population groups—an increase in the standard error was observed under the BEC model. For instance, the SE for sex increased from 0.062 to 0.165, which may reflect more significant uncertainty in estimating this covariate's effect when diagnostic uncertainty is explicitly incorporated. A similar trend was noted for the population group categories, with relative decreases in precision ranging from -25% to -54%.

In contrast, variables such as age, hepatitis B reactivity, and sex worker status showed comparable or slightly reduced standard errors under the Bayesian model, suggesting that the estimates' precision was maintained or even improved.

Overall, these results indicate that the Bayesian model with correction adjusts point estimates and alters precision in a covariate-specific manner. The explicit correction of misclassification errors enables a more robust estimation of the intercept (and thus of the adjusted prevalence). However, it may come with trade-offs in the precision of specific individual predictors.

## 4 Discussion

Several studies have demonstrated that estimating disease prevalence using imperfect diagnostic tests can lead to substantial bias if sensitivity and specificity are not adequately accounted for. The need for such correction has been widely documented in the literature, particularly in settings where prevalence varies across populations. In this regard, Li and Fine [14] point out that sensitivity—traditionally considered an intrinsic property of the test—is a significant factor influencing prevalence estimates due to factors such as changes in the severity spectrum of cases or variations in the clinical context of diagnosis. Consequently, adjusting for sensitivity and specificity becomes methodologically relevant and necessary to ensure valid and comparable prevalence estimates across studies or heterogeneous populations.

Accurate estimation of prevalence adjusted for individual characteristics becomes especially relevant when the need is not only to obtain an overall population average but also to generate stratified or subgroup-specific estimates defined by demographic or epidemiological covariates. In this context, logistic regression models emerge as an optimal and flexible tool to address this need. Vansteelandt et al. [15] emphasize that when information on individual-level covariates is available, the extension of classical estimators through generalized linear models allows for the direct inclusion of these factors in the analysis, thereby improving the precision and validity of prevalence estimates. This approach is not only adaptable to different sampling design configurations. However, it is also advantageous in terms of statistical efficiency and cost-effectiveness, as it makes more effective use of the collected information—especially when combined with methodological proposals that expand its scope for estimation purposes [16].

This study compared different methodological approaches for estimating prevalence and associations in the presence of misclassification in the outcome, applied to sexually transmitted infections with low and moderate prevalence. Standard logistic regression models were evaluated, the model proposed by Liu et al. [4], and two Bayesian approaches—with and without explicit correction for misclassification [5].

It is important to note that although relevant changes were observed in the coefficients of some covariates across the different models, the magnitude and direction of these changes should be interpreted in light of the underlying distribution of variables in the study population. For instance, coefficients with low absolute values in the standard model may exhibit significant relative variations in the corrected models without necessarily implying a substantive change in their effect.

In contrast, the intercept ($\beta_0$) showed consistent changes across models, suggesting that the primary corrections are reflected in the baseline component of the model—that is, in the prevalence estimation when all covariates are set to zero. In this context, the corrected intercept can be interpreted as approximating the baseline prevalence adjusted for misclassification error, providing a more stable and robust outcome measure without additional risk factors.

These findings reinforce the notion that correcting for misclassification affects relative associations and the point estimate of the baseline prevalence. In particular, it was observed that adjusted estimates differ systematically from unadjusted ones. While it cannot be asserted with certainty that these corrections necessarily yield values closer to the true prevalence, it is reasonable to consider that they represent a significant methodological improvement. In this sense, corrected models may offer estimates more consistent with epidemiological assumptions and less biased due

to misclassification error, which is especially relevant in population-based studies and epidemiological surveillance systems.

The results revealed substantial differences in adjusted prevalence estimates and regression coefficients depending on the approach used. In particular, the Bayesian model with misclassification correction (BEC) demonstrated a favorable balance between precision and bias correction, producing intermediate estimates between the standard model and the Liu model while yielding the narrowest confidence intervals. This finding was consistent for both HIV and syphilis.

However, methodological and practical implications must be carefully considered when selecting the statistical approach. The model proposed by Liu et al. [4], especially in its more general formulation that simultaneously estimates the false positive (FP) and false negative (FN) rates, incorporates a substantially more significant number of parameters than standard logistic regression. This increased complexity translates into higher sample size requirements to achieve estimation stability and to ensure convergence of the Fisher scoring algorithm. According to the simulations presented in their work, a minimum sample size of approximately $5{,}000$ observations is required to obtain reliable parameter estimates, including the misclassification rates. This requirement becomes even more critical when the outcome prevalence is low, as the limited number of informative events may severely constrain the model's ability to estimate the additional parameters associated with classification error accurately.

In this study, despite having a considerably large sample size ($n = 11452$), challenges in terms of precision were observed, expressed as wide confidence intervals and high standard errors for several coefficients in the Liu model. This phenomenon was particularly evident in covariates associated with very low-prevalence conditions, such as HIV positivity, where the effective number of cases was relatively limited ($1.4\%$ of the total sample). These findings suggest that even in settings with large sample sizes, the model's performance may be compromised by the low occurrence of the event, thereby limiting its practical applicability in studies of rare infectious diseases.

Additionally, the Liu model exhibited convergence issues in scenarios where the data did not include both misclassification errors. This aspect, linked to the structural assumptions of the error model, may compromise the validity of the approach. In contrast, Bayesian models allow for the incorporation of prior information on sensitivity and specificity, improving estimation stability in smaller samples and enabling a probabilistic characterization of uncertainty. In this study, that advantage was reflected in substantial improvements in the precision of the intercept—directly related to the corrected baseline prevalence—as well as in key coefficients.

Concerning Bayesian modeling, an important methodological advantage lies in incorporating prior information on the regression coefficients, allowing for the regularization of estimates and reducing variability associated with small samples. In this regard, Newman et al. [7] proposed the formulation of informative prior distributions based on the variance of standardized coefficients and the number of covariates included in the model—a methodology implemented in this study. This approach is beneficial in contexts where classical estimation yields unstable coefficients and extensive standard errors.

However, the use of non-informative priors also holds practical applicability, as demonstrated by Gordóvil-Merino et al. [17] through a simulation study. They employed weakly informative priors, providing excellent stability in estimating parameters in logistic regression models applied to small samples. Although the authors acknowledged that this approach does not fully resolve issues arising from asymmetric distributions, they emphasized that including prior knowledge—even minimal—helps mitigate extreme inferences and improve model regularization. It is crucial in studies with small numbers of observations or rare events. Thus, the Bayesian approach emerges as a methodologically sound and flexible alternative for situations where the assumptions of classical models are compromised.

Our findings support using models that explicitly account for misclassification in prevalence and association studies based on imperfect diagnostic tests. While the Liu model represents an improvement over standard logistic regression by addressing systematic bias, its practical and theoretical limitations make it less suitable in specific contexts. In contrast, Bayesian models with misclassification correction—such as the one evaluated in this study—stand out as a

robust and flexible alternative, particularly in low-prevalence settings, when diagnostic errors are known or partially known, and when precise estimates are needed to support public health decision-making.

## References


[1] Magder L. and P Hughes James. Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*, 1997.

[2] I. Lewis Fraser, Sanchez-Vazquez M., and R. Torgerson P. Association between covariates and disease occurrence in the presence of diagnostic error. *Epidemiology and Infection*, 2011.

[3] Clemma J Muller and Richard F MacLehose. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *International Journal of Epidemiology*, 43(3):962–970, 03 2014.

[4] Haiyan Liu and Zhiyong Zhang. Logistic regression with misclassification in binary outcome variables: a method and software. *Behaviormetrika*, 44:447–476, 2017.

[5] Valle D., M. Tucker Lima Joanna, Millar J., Amratia P., and Haque Ubydul. Bias in logistic regression due to imperfect diagnostic test results and practical correction approaches. *Malaria Journal*, 2015.

[6] Walter J. Rogan and Beth Gladen. Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, 107(1):71–76, 01 1978.

[7] Ken B. Newman, Cristiano Villa, and Ruth King. Logistic regression models: practical induced prior specification, 2025.

[8] JM Neuhaus. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86(4):843–855, 12 1999.

[9] J.A. Hausman, Jason Abrevaya, and F.M. Scott-Morton. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, 87(2):239–269, 1998.

[10] Tim P. Morris, Ian R. White, and Michael J. Crowther. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102, 2019.

[11] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025.

[12] Stan Development Team. RStan: the R interface to Stan, 2025. R package version 2.32.7.

[13] Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. rstanarm: Bayesian applied regression modeling via Stan., 2024. R package version 2.32.1.

[14] Jialiang Li and Jason P. Fine. Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis. *Biostatistics*, 12(4):710–722, 04 2011.

[15] S. Vansteelandt, E. Goetghebeur, and T. Verstraeten. Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics*, 56(4):1126–1133, 05 2004.

[16] Mario Villalobos-Arias. Estimation of population infected by covid-19 using regression generalized logistics and optimization heuristics, 2020.

[17] Amalia Gordóvil Merino, Joan Guàrdia Olmos, and Maribel Peró Cebollero. Estimation of logistic regression models in small samples. a simulation study using a weakly informative default prior distribution. *Psicológica: Revista de metodología y psicología experimental*, 33(2):345–361, 2012.

This figure "test.png" is available in "png"  format from:

http://arxiv.org/ps/2504.15150v1

# Referencias


[1] Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press; 2003. Available from: https://doi.org/10.1093/oso/9780198509844.001.0001. (pages 1 and 36)

[2] Rodriguez* FH, Estrada* JM, Quintero HMA, Nogueira JP, Porras-Hurtado GL. Analyses of familial chylomicronemia syndrome in Pereira, Colombia 2010–2020: a cross-sectional study. Lipids in Health and Disease. 2023;22:43. Available from: https://doi.org/10.1186/s12944-022-01768-x.

[3] Estrada Alvarez JM, Luna del Castillo JdD, Montero-Alonso M Point and Interval Estimation of Population Prevalence Using a Fallible Test and a Non-Probabilistic Sample: Post-Stratification Correction. Mathematics. 2025;13(5). Available from: https://www.mdpi.com/2227-7390/13/5/805.

[4] Alvarez JME, Garcia HF, Ángel Montero-Alonso M, de Dios Luna del Castillo J. Prevalence estimation in infectious diseases with imperfect tests: A comparison of Frequentist and Bayesian Logistic Regression methods with misclassification correction; 2025. Available from: https://arxiv.org/abs/2504.15150.

[5] Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. Statistical Science. 2001;16(2):101-17. Available from: http://www.jstor.org/stable/2676784. (pages 6, 9, 14, 16, 23, 46, and 48)

[6] Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. American Journal of Epidemiology. 1978 01;107(1):71-6. Available from: https://doi.org/10.1093/oxfordjournals.aje.a112510. (pages 7, 23, 26, 27, 52, 83, and 84)

[7] Agresti A, Coull BA. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. The American Statistician. 1998 05;52(2):119-26. (pages 11, 15, 31, 46, 51, and 58)

[8] Vollset SE. Confidence intervals for a binomial proportion. Statistics in Medicine. 1993;12(9):809-24. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780120902. (pages 11 and 15)

[9] Agresti A. Categorical Data Analysis. Wiley Series in Probability and Statistics. Wiley; 2012. Available from: https://books.google.com.co/books?id=UOrr47-2oisC. (pages 11 and 15)

[10] Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. Wiley Series in Probability and Statistics. Wiley; 2003. Available from: https://books.google.com.co/books?id=PNn2nQEACAAJ. (page 16)

[11] Somerville MC, Brown RS. Exact likelihood ratio and score confidence intervals for the binomial proportion. Pharmaceutical Statistics. 2013;12(3):120-8. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/pst.1560. (page 16)

[12] Thulin M. Coverage-adjusted Confidence Intervals for a Binomial Proportion. Scandinavian Journal of Statistics. 2014;41(2):291-300. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12021. (pages 17 and 19)

[13] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Chapman & Hall/CRC Texts in Statistical Science. CRC Press; 2013. Available from: https://books.google.com.co/books?id=eSHSBQAAQBAJ. (page 20)

[14] Dalitz C. Construction of Confidence Intervals; 2018. Available from: https://arxiv.org/abs/1807.03582. (page 23)

[15] Cepeda-Cuervo E, Aguilar W, Cervantes V, Corrales M, Díaz I, Rodriguez D. Intervalos de confianza e intervalos de credibilidad para una proporción. Revista Colombiana de Estadistica. 2008 12;31:211 228. Available from: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-17512008000200006&nrm=iso. (page 23)

[16] Estrada Álvarez J. El índice de Youden y su aplicación a la determinación del punto de corte en un test cuantitativo [Trabajo Fin de Máster]. Granada, España: Universidad de Granada; 2016. Máster en Estadística Aplicada. Tutor: Juan de Dios Luna Del Castillo. (page 24)

[17] Andrés AM, de Dios Luna del Castillo J. Bioestadistica +: Para las ciencias de la salud. Textos Universitarios. Norma-Capitel; 2004. Available from: https://books.google.com.co/books?id=kZ5NoA2BwjEC. (page 26)

[18] Reiczigel J, Foldi J, Ózsvári L. Exact confidence limits for prevalence of a disease with an imperfect diagnostic test. Epidemiology and Infection. 2010;138(11):1674–1678. (pages 29, 52, and 60)

[19] Blaker H. Confidence curves and improved exact confidence intervals for discrete distributions. Canadian Journal of Statistics. 2000;28(4):783-98. Available from: https://onlinelibrary.wiley.com/doi/abs/10.2307/3315916. (page 29)

[20] Sterne TE. Some Remarks on Confidence or Fiducial Limits. Biometrika. 1954;41(1/2):275-8. Available from: http://www.jstor.org/stable/2333026. (page 29)

[21] Lang Z, Reiczigel J. Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. Preventive Veterinary Medicine. 2014;113(1):13-22. Available from: https://www.sciencedirect.com/science/article/pii/S0167587713002936. (pages 31, 46, 49, 52, and 61)

[22] Lew RA, Levy PS. Estimation of prevalence on the basis of screening tests. Statistics in Medicine. 1989;8(10):1225-30. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780081006. (page 32)

[23] Joseph L, Gyorkos TW, Coupal L. Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard. American Journal of Epidemiology. 1995 02;141(3):263-72. Available from: https://doi.org/10.1093/oxfordjournals.aje.a117428. (page 32)

[24] Lewis FI, Torgerson PR. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. Emerging Themes in Epidemiology. 2012;9:9. Available from: https://doi.org/10.1186/1742-7622-9-9. (pages 32 and 51)

[25] Speybroeck N, Devleesschauwer B, Joseph L, Berkvens D. Misclassification errors in prevalence estimation: Bayesian handling with care. International Journal of Public Health. 2013;58(5):791-5. Available from: https://doi.org/10.1007/s00038-012-0439-9. (page 32)

[26] Flor M, Weiß M, Selhorst T, Müller-Graf C, Greiner M. Comparison of Bayesian and frequentist methods for prevalence estimation under misclassification. BMC Public Health. 2020;20:1135. Available from: https://doi.org/10.1186/s12889-020-09177-4. (pages 36, 46, 70, and 71)

[27] McNamee R. Two-Phase Sampling for Simultaneous Prevalence Estimation and Case Detection. Biometrics. 2004;60(3):783-92. Available from: http://www.jstor.org/stable/3695401. (pages 37, 45, 69, and 71)

[28] Pickles A, Dunn G, Vázquez-Barquero JL. Screening for stratification in two-phase ('two- stage') epidemiological surveys. Statistical Methods in Medical Research. 1995;4(1):73-89. PMID: 7613639. Available from: https://doi.org/10.1177/096228029500400106. (page 37)

[29] Shrout PE, Newman SC. Design of Two-Phase Prevalence Surveys of Rare Disorders. Biometrics. 1989;45(2):549-55. Available from: http://www.jstor.org/stable/2531496. (pages 37 and 45)

[30] G Thomas E, B Peskoe S, Spiegelman D. Prevalence estimation when disease status is verified only among test positives: Applications in HIV screening programs. Statistics in Medicine. 2018;37(7):1101-14. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7568. (pages 39, 46, 52, 63, 64, and 65)

[31] Cepeda LL, Ramos-Garibay JA, Calderón DP, Reyes M. Xantomas eruptivos como manifestación inicial de diabetes mellitus e hipertrigliceridemia severa. Revista Centro Dermatológico Pascua. 2010;19:4. (page 44)

[32] Tamez-Pérez HE, Sáenz-Gallegos R, Hernández-Rodríguez K, Forsbach-Sánchez G, Gómez-de Ossio MD, Fernández-Garza N, et al. Terapia con insulina en pacientes con hipertrigliceridemia severa. Revista Médica del Instituto Mexicano del Seguro Social. 2006;44(3):235-7. (page 44)

[33] Hegele RA, Ginsberg HN, Chapman MJ, Nordestgaard BG, Kuivenhoven JA, Averna M, et al. The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. The Lancet Diabetes & Endocrinology. 2014;2(8):655-66. (page 44)

[34] Virani SS, Morris PB, Agarwala A, Ballantyne CM, Birtcher KK, Kris-Etherton PM, et al. 2021 ACC Expert Consensus Decision Pathway on the Management of ASCVD Risk Reduction in Patients With Persistent Hypertriglyceridemia. Journal of the American College of Cardiology. 2021;78(9):960-93. (page 44)

[35] Herrera Del Águila DD, Garavito Rentería J, Linarez Medina K, Lizarzaburu Rodríguez V. Pancreatitis aguda por hipertrigliceridemia severa: reporte de caso y revisión de la literatura. Revista de Gastroenterología del Perú. 2015;35(2):159-64. (page 44)

[36] Davidson M, Stevenson M, Hsieh A, Ahmad Z, Roeters van Lennep J, Crowson C, et al. The burden of familial chylomicronemia syndrome: results from the global IN-FOCUS study. Journal of Clinical Lipidology. 2018;12(4):898-907.e2. (page 44)

[37] Bchetnia M, Bouchard L, Mathieu J, Campeau PM, Morin C, Brisson D, et al. Genetic burden linked to founder effects in Saguenay–Lac-Saint-Jean illustrates the importance of genetic screening test availability. Journal of Medical Genetics. 2021;58(10):653-65. (page 44)

[38] Gaudet D, Stevenson M, Komari N, Trentin G, Crowson C, Hadker N, et al. The burden of familial chylomicronemia syndrome in Canadian patients. Lipids in Health and Disease. 2020;19(1):120. (page 44)

[39] Moulin P, Dufour R, Averna M, Arca M, Cefalù AB, Noto D, et al. Identification and diagnosis of patients with familial chylomicronaemia syndrome (FCS): Expert panel recommendations and proposal of an "FCS score". Atherosclerosis. 2018 8;275:265-72. (page 45)

[40] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2025. Available from: https://www.R-project.org/. (page 46)

[41] Dorai-Raj S. binom: Binomial Confidence Intervals For Several Parameterizations; 2021. R package version 1.1-1. Available from: https://CRAN.R-project.org/package=binom. (page 46)

[42] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017;76(1):1-32. Available from: https://www.jstatsoft.org/article/view/v076i01. (page 46)

[43] Stan Development Team. RStan: the R interface to Stan; 2025. R package version 2.32.7. Available from: https://mc-stan.org/. (page 46)

[44] McNamee R. Efficiency of two-phase designs for prevalence estimation. International Journal of Epidemiology. 2003 12;32(6):1072-8. Available from: https://doi.org/10.1093/ije/dyg230. (page 50)

[45] Arya R, Antonisamy B, Kumar S. Sample Size Estimation in Prevalence Studies. The Indian Journal of Pediatrics. 2012;79:1482-8. Available from: https://doi.org/10.1007/s12098-012-0763-3. (page 51)

[46] Izbicki R, Diniz MA, Bastos LS. Sensitivity and specificity in prevalence studies: The importance of considering uncertainty. Clinics. 2020;75:e2449. Available from: https://www.sciencedirect.com/science/article/pii/S1807593222004574. (page 52)

[47] Elliott MR, Valliant R. Inference for nonprobability samples. Statistical Science. 2017 5;32:249-64. (page 52)

[48] Lohr SL. Sampling: Design and Analysis. 2nd ed. Boston, MA: Cengage Learning; 2010. (pages 52 and 54)

[49] Smith TMF. Post-Stratification. Journal of the Royal Statistical Society Series D (The Statistician). 1991;40(3):315-23. Available from: http://www.jstor.org/stable/2348284. (page 52)

[50] Holt D, Smith TMF. Post Stratification. Journal of the Royal Statistical Society Series A (General). 1979;142:33-46. Available from: http://www.jstor.org/stable/2344652. (pages 57 and 70)

[51] Tony Cai T. One-sided confidence intervals in discrete distributions. Journal of Statistical Planning and Inference. 2005;131(1):63-88. Available from: https://www.sciencedirect.com/science/article/pii/S0378375804000679. (page 59)

[52] Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science. 1992;7(4):457 472. Available from: https://doi.org/10.1214/ss/1177011136. (page 66)

[53] Shrout PE, Newman SC. Design of Two-Phase Prevalence Surveys of Rare Disorders. Biometrics. 1989;45(2):549-55. Available from: http://www.jstor.org/stable/2531496. (page 70)

[54] M A G Viana VR, Levy PS. Bayesian analysis of prevalence from the results of small screening samples. Communications in Statistics - Theory and Methods. 1993;22(2):575-85. Available from: https://doi.org/10.1080/03610929308831038. (page 70)

[55] Van Hasselt M, Bollinger CR, Bray JW. A Bayesian approach to account for misclassification in prevalence and trend estimation. Journal of Applied Econometrics. 2022;37(2):351-67. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2879. (page 70)

[56] Bayer DM, Fay MP, Graubard BI. Confidence intervals for prevalence estimates from complex surveys with imperfect assays. Statistics in Medicine. 2023;42(11):1822-67. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.9701. (page 71)

[57] L M, James PH. Logistic regression when the outcome is measured with uncertainty. American Journal of Epidemiology. 1997. (pages 74 and 80)

[58] Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. International Journal of Epidemiology. 2014 03;43(3):962-70. Available from: https://doi.org/10.1093/ije/dyu029. (page 75)

[59] Liu H, Zhang Z. Logistic regression with misclassification in binary outcome variables: a method and software. Behaviormetrika. 2017;44:447-76. Available from: https://doi.org/10.1007/s41237-017-0031-y. (pages 75, 79, 82, 83, 85, 102, and 103)

[60] D V, Joanna MTL, J M, P A, Ubydul H. Bias in logistic regression due to imperfect diagnostic test results and practical correction approaches. Malaria Journal. 2015. Available from: https://malariajournal.biomedcentral.com/track/pdf/10.1186/s12936-015-0966-y. (pages 75, 81, 82, 83, 92, and 102)

[61] L C, M S, P M. Methods for estimating prevalence ratios in cross-sectional studies. Revista de Saúde Pública. 2008. (page 79)

[62] Goette WF, Fatima H, Schaffert J, Carlew AR, Rossetti H, Lacritz LH, et al. 52 Bayesian Logistic Regression Bias Adjustment for Data Observed without a Gold Standard: A Simulation Study of Clinical Alzheimer's Disease. Journal of the International Neuropsychological Society. 2023;29(s1):259–260. (page 80)

[63] Fraser IL, M SV, P RT. Association between covariates and disease occurrence in the presence of diagnostic error. Epidemiology and Infection. 2011. (page 80)

[64] J Q, Han Z, Pengfei L, D A, Kai Y. Using covariate-specific disease prevalence information to increase the power of case-control studies. 2015. (page 80)

[65] Torsten s, J D, Martin RP, L E. Prevalence proportion ratios: estimation and hypothesis testing. International Journal of Epidemiology. 1998. (page 80)

[66] X T, J K, G J. Bayesian analysis of prevalence with covariates using simulation-based techniques: applications to HIV screening. Statistics in Medicine. 1999. (page 80)

[67] Vansteelandt S, Goetghebeur E, Verstraeten T. Regression Models for Disease Prevalence with Diagnostic Tests on Pools of Serum Samples. Biometrics. 2004 05;56(4):1126-33. Available from: https://doi.org/10.1111/j.0006-341X.2000.01126.x. (pages 81 and 102)

[68] Neuhaus J. Bias and efficiency loss due to misclassified responses in binary regression. Biometrika. 1999 12;86(4):843-55. Available from: https://doi.org/10.1093/biomet/86.4.843. (page 85)

[69] Hausman JA, Abrevaya J, Scott-Morton FM. Misclassification of the dependent variable in a discrete-response setting. Journal of Econometrics. 1998;87(2):239-69. Available from: https://www.sciencedirect.com/science/article/pii/S0304407698000153. (page 85)

[70] Newman KB, Villa C, King R. Logistic regression models: practical induced prior specification; 2025. Available from: https://arxiv.org/abs/2501.18106. (pages 88 and 104)

[71] 5. In: Eliciting Distributions – General. John Wiley Sons, Ltd; 2006. p. 97-119. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470033312.ch5. (page 89)

[72] 6. In: Eliciting and Fitting a Parametric Distribution. John Wiley Sons, Ltd; 2006. p. 121-51. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470033312.ch6. (page 89)

[73] Newman KB, Villa C, King R. Logistic regression models: practical induced prior specification; 2025. Available from: https://arxiv.org/abs/2501.18106. (page 89)

[74] 5. In: Choosing the Prior Distribution. John Wiley Sons, Ltd; 2012. p. 104-38. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119942412.ch5. (page 89)

[75] III JWS, Jr JWS, and JDS. Hidden Dangers of Specifying Noninformative Priors. The American Statistician. 2012;66(2):77-84. Available from: https://doi.org/10.1080/00031305.2012.695938. (page 90)

[76] Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Statistics in Medicine. 2019;38(11):2074-102. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8086. (page 93)

[77] Li J, Fine JP. Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis. Biostatistics. 2011 04;12(4):710-22. Available from: https://doi.org/10.1093/biostatistics/kxr008. (page 102)

[78] Villalobos-Arias M. Estimation of population infected by Covid-19 using regression Generalized logistics and optimization heuristics; 2020. Available from: https://arxiv.org/abs/2004.01207. (page 102)

[79] Gordóvil Merino A, Guàrdia Olmos J, Peró Cebollero M. Estimation of logistic regression models in small samples. A simulation study using a weakly informative default prior distribution. Psicológica: Revista de metodología y psicología experimental. 2012;33(2):345-61. (page 104)