



TRABAJO FIN DE MASTER EN BIOESTADÍSTICA

**META-ANÁLISIS DE
PROPORCIONES CON
MODELOS MIXTOS LINEALES
GENERALIZADOS PARA
ESTIMAR LA PREVALENCIAS**

SEPTIEMBRE 2025

Nombre: Alejandro Ballester Palomar

Tutor: David Lora Pablos

Declaración Responsable sobre Autoría y Uso Ético de Herramientas de Inteligencia Artificial (IA)

Yo, BALLESTER PALOMAR ALEJANDRO

Con DNI:74446540W

declaro de manera responsable que el/la presente:

- Trabajo de Fin de Grado (TFG)
- Trabajo de Fin de Máster (TFM)
- Tesis Doctoral

Titulado/a

META-ANÁLISIS DE PROPORCIONES CON MODELOS MIXTOS LINEALES GENERALIZADOS PARA ESTIMAR LA PREVALENCIAS

es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a 10/09/2025



FIRMA

ALEJANDRO BALLESTER PALOMAR

ÍNDICE DE CONTENIDOS

Índice de contenidos	4
Abreviaturas.....	6
Lista de Tablas.....	7
Lista de Figuras	8
Lista de Gráficos.....	9
Resumen	10
Abstract.....	11
1. Introducción.....	12
2. Objetivos.....	15
3. Metodología.....	16
Meta-análisis: metodología convencional	16
Modelo de efectos fijos.....	17
Modelo de efectos aleatorios	18
Meta-análisis de proporciones	21
Alternativa: Uso de los modelos mixtos lineales generalizados (GLMMs).....	22
Heterogeneidad.....	23
Comparación entre los modelos	25
Datos.....	28
Software estadístico.....	29
4. Resultados.....	31
Descripción de los datos	31
Meta-análisis de prevalencias: GLMMs vs. CONVENCIONAL.....	33
Enlace cauchit: Error de convergencia	34
Análisis de la heterogeneidad.....	35
Comparación entre metodología de dos pasos (convencional) o de un paso (GLMMs):	38
5. Conclusiones.....	55
6. Bibliografía.....	57

Anexo 1 – BASE DE DATOS.....	60
Anexo 2 – CÓDIGO DE R.....	61

ABREVIATURAS

TFM	Trabajo de Fin de Máster
GLMMs	Modelos Lineales Generalizados de Efectos Mixtos (<i>Generalised Linear Mixed Models</i>)
SJS	Síndrome de Stevens-Johnson (<i>Steven Johnson Syndrome</i>)
TEN	Necrólisis epidérmica tóxica (<i>Toxic Epidermic Necrosis</i>)
AIC	Criterio de Información de Akaike (<i>Akaike Information Criteria</i>)
IC	Intervalo de confianza
OR	Odds ratio
RR	Riesgo relativo
IDE	Entorno de desarrollo integrado (<i>Integrated Development Environment</i>)

LISTA DE TABLAS

Tabla 1. Distribución de frecuencias relativas y absolutas de las prevalencias estimadas de SJS y TEN en los 38 estudios incluidos en el artículo de Lee <i>et al.</i> (2023) (15).	32
Tabla 2. Análisis de prevalencias usando GLMMs y el modelo de efectos aleatorios convencional (de dos pasos).....	33
Tabla 3. Corrección problema de convergencia enlace cauchit. Se muestra la estimación puntual como su IC 95%.	35
Tabla 4 Análisis de la heterogeneidad.....	36
Tabla 5. Comparación de las estimaciones de AIC de los diferentes modelos.	41
Tabla 6. Análisis de prevalencias en función del tamaño de muestra.....	43
Tabla 7. Estimaciones del tamaño de efecto global para cada uno de los modelos en presencia de valores atípicos muy elevados.	44
Tabla 8. Estimaciones del tamaño de efecto global para cada uno de los modelos en presencia de valores atípicos muy bajos.....	44
Tabla 9. Datos simulados en el supuesto de baja heterogeneidad.....	46
Tabla 10. Estimaciones globales por los diferentes modelos en el supuesto de baja heterogeneidad.....	46
Tabla 11. Análisis de la heterogeneidad en el supuesto de baja heterogeneidad.	48
Tabla 12. Valores de AIC por tipo de enlace. Supuesto de baja heterogeneidad.....	49
Tabla 13. Datos simulados en el supuesto de alta heterogeneidad.....	50
Tabla 14. Estimaciones globales por los diferentes modelos en el supuesto de alta heterogeneidad.....	50
Tabla 15. Análisis de la heterogeneidad en el supuesto de alta heterogeneidad.	53
Tabla 16. Valores de AIC de cada uno de los modelos ajustados. Supuesto de alta heterogeneidad.....	53

LISTA DE FIGURAS

Figura 1. Representación del error de muestreo. Figura adaptada del libro de Borenstein <i>et al.</i> (2009) (5).....	17
Figura 2. Modelo de efectos aleatorios - tamaño del efecto real y el observado en un estudio. Imagen adaptada del libro de Borenstein <i>et al.</i> (2009) (5).....	19
Figura 3. Comparación entre las estimaciones puntuales de prevalencia de cada estudio y las predicciones realizadas por los diferentes modelos.	40

LISTA DE GRÁFICOS

Gráfico 1. Representación de las prevalencias de los estudios incluidos en el metaanálisis de Lee <i>et al.</i> (2023) (15).	31
Gráfico 2. Forestplot de los 38 estudios incluidos en el metaanálisis. Estimación realizada con GLMMs y el enlace probit.....	36
Gráfico 3. Comparación entre las diferentes metodologías a la hora de realizar las estimaciones del efecto agrupado.	39
Gráfico 4. Forestplot de los 38 estudios simulados y la estimación global del efecto (39) en el supuesto de baja heterogeneidad. Estimación realizada con GLMMs y el enlace probit.....	47
Gráfico 5. Forestplot de los 38 estudios simulados y la estimación global del efecto (39) en el supuesto de alta heterogeneidad. Estimación realizada con GLMMs y el enlace logit	51
Gráfico 6. Comparación entre las estimaciones globales del efecto realizadas por los diferentes modelos cuando hay una elevada heterogeneidad	52

RESUMEN

El metaanálisis es una técnica estadística crucial en la medicina basada en la evidencia, permitiendo sintetizar la información de múltiples estudios. Sin embargo, los modelos convencionales presentan desafíos metodológicos a la hora de analizar proporciones, debido a la naturaleza restringida entre 0 y 1. Una de las alternativas es el uso de los Modelos Lineales Generalizados de Efectos Mixtos (GLMMs), que, gracias a sus atributos, no tienen las limitaciones de los modelos convencionales, los cuales requieren de una transformación previa a su análisis, usando la metodología de dos pasos.

Este Trabajo de Fin de Máster tuvo como objetivo principal evaluar el uso de los GLMMs como alternativa a los métodos convencionales en el metaanálisis de proporciones.

Para lograrlo, se reanalizaron los datos de un metaanálisis publicado sobre la prevalencia del síndrome de Stevens-Johnson y la necrólisis epidérmica tóxica asociada a antibióticos. Se compararon diferentes GLMMs entre sí y con el modelo convencional utilizado por los autores, prestando especial atención a la estimación del efecto global y la heterogeneidad.

Los resultados mostraron que, tanto el modelo convencional como los GLMMs produjeron estimaciones similares. Sin embargo, el modelo convencional demostró infraestimar la variabilidad de los datos. La comparación mediante el Criterio de Información de Akaike confirmó que, los enlaces probit y logit son los más adecuados.

En conclusión, los GLMMs demostraron ser una alternativa para el metaanálisis de proporciones. Los modelos convencionales de dos pasos pueden requerir de una evaluación minuciosa de diferentes aspectos, haciendo que puedan ser más susceptibles a errores.

Palabras clave: Metaanálisis, Proporciones, prevalencia, Modelos Lineales Generalizados de Efectos Mixtos, epidemiología.

ABSTRACT

Meta-analysis is a crucial statistical technique in evidence-based medicine, allowing for the synthesis of information from multiple studies. However, conventional models present methodological challenges when analyzing proportions due to their constrained nature between 0 and 1. An alternative is the use of Generalized Linear Mixed Models (GLMMs), which, due to their properties, avoid the limitations of conventional models that necessitate a two-step transformation.

The primary objective of this Master's thesis was to assess the application of GLMMs as an alternative to conventional methods in meta-analysis of proportions.

To achieve this, data from a published meta-analysis on the prevalence of Stevens-Johnson Syndrome and toxic epidermal necrolysis associated with antibiotics were reanalyzed. Different GLMMs were compared with each other and with the conventional model used by the authors, paying special attention to the estimation of the overall effect and heterogeneity.

The results showed that both the conventional model and the GLMMs produced similar estimates. However, the conventional model was found to underestimate the variability of the data. The comparison using the Akaike Information Criterion confirmed that the probit and logit links are the most suitable.

In conclusion, GLMMs proved to be a viable alternative for the meta-analysis of proportions. Conventional two-step models may require a thorough evaluation of different aspects, making them more susceptible to errors.

Keywords: Meta-analysis, proportions, prevalence, Generalized Linear Mixed Models, epidemiology.

1. INTRODUCCIÓN

Actualmente, el rápido avance de la ciencia, especialmente en el ámbito de las ciencias de la salud, ha dado lugar a la creación de una gran cantidad de datos, que en ocasiones llegan a ser inmanejables. Esto se traduce en que existe un elevado número de estudios sobre un tema en concreto hoy en día, y en varias ocasiones estudios similares pueden dar evidencias distintas e incluso contradictorias. Un claro ejemplo de esto es el caso del consumo de huevos y la relación con la elevación del colesterol en sangre. Durante décadas se recomendó evitar el consumo de más de 3 huevos a la semana por su alto contenido de colesterol, basándose en estudios que relacionaban el colesterol dietético con el riesgo cardiovascular (1).

Sin embargo, existen investigaciones que a su vez no muestran un aumento significativo de estos niveles en las personas que consumen este tipo de alimentos de manera regular (2–4). Por lo tanto, es necesario obtener la manera de poder sintetizar y analizar toda esta información y así poder obtener conclusiones al respecto.

Las revisiones sistemáticas permiten agrupar toda esta evidencia de acuerdo a una metodología definida y bien establecida, permitiendo a los autores dar respuesta a una pregunta concreta. Sin embargo, dichas revisiones se basan en análisis cualitativos, sujetos a interpretación y posiblemente influenciados, ya sea directa o indirectamente, por la experiencia y creencias previas de los investigadores (5). En este caso, sería necesaria una metodología que permita resumir toda esta evidencia y poder dar respuesta a esa pregunta de manera objetiva. Aquí es donde el metaanálisis juega su papel principal.

Un metaanálisis es una técnica estadística utilizada para combinar y sintetizar cuantitativamente los resultados de múltiples estudios independientes que investigan una misma pregunta de investigación o un tema similar (6). Se puede decir que el metaanálisis es la estadística inferencial de la revisión sistemática (7).

A diferencia de las revisiones narrativas, los metaanálisis son capaces de sintetizar los tamaños de los efectos de los diferentes estudios individuales en una estimación global, con un intervalo de confianza (IC). Al sintetizar toda esta evidencia disponible, se puede obtener un resultado estadísticamente concluyente y que, al menos teóricamente, estaría muy cercano al valor del parámetro poblacional (7).

Por esta razón las revisiones sistemáticas con metaanálisis son consideradas unas de las principales bases en las que se sustenta la medicina basada en la evidencia (8), que utiliza la evidencia científica para el tratamiento de las enfermedades (9).

Actualmente, existen varias revistas y entidades especializadas en ello, siendo Cochrane uno de los principales exponentes de este tipo de estudios¹. Otras revistas como *the British Medical Journal (BMJ)* publican una serie de revistas tituladas *Evidence Based Medicine*, basándose principalmente en los resultados de las revisiones sistemáticas y los metaanálisis².

Los metaanálisis permiten tratar múltiples temas de interés para el ámbito de la medicina, como en la investigación clínica para conocer la eficacia de diferentes tratamientos, la adecuación de llevar a cabo pruebas diagnósticas específicas o bien para realizar estimaciones de la prevalencia de una enfermedad en el ámbito de la epidemiología. Siendo este último el ámbito en el que se va a centrar el presente Trabajo Fin de Máster (TFM).

El metaanálisis de prevalencias busca estimar la proporción de pacientes con una enfermedad en una población o área específica (también conocida como prevalencia). Sin embargo, este tipo de metaanálisis presenta requisitos adicionales en comparación con otros tipos de variables. Según el artículo de Lin y Chu (2020) (10), dado que las proporciones están restringidas entre 0 y 1, es necesario realizar transformaciones en los datos para aproximarlos a una distribución normal o estabilizar las varianzas. Estas transformaciones llevan a varias limitaciones que pueden afectar tanto la interpretabilidad de los resultados como la inferencia estadística subsiguiente.

Las nuevas técnicas de computación han permitido el desarrollo de los modelos lineales mixtos generalizados (GLMMs, por sus siglas en inglés), los cuales se postulan como una posible alternativa para realizar este tipo de análisis (11–14). La implementación de estos modelos en el metaanálisis de proporciones ha sido posible gracias a sus características intrínsecas que facilitan en gran medida el análisis y la interpretación de los resultados (10).

Sin embargo, parece que debido a la complejidad que puede acarrear el comprender cómo funcionan este tipo de modelos y sus propiedades, rara vez son aplicados en la

¹ <https://www.cochranelibrary.com/es/>

² <https://ebm.bmj.com/>

investigación. Por lo tanto, este TFM pretende profundizar en su uso, con el fin de desgranar sus ventajas y limitaciones y evaluar si realmente son tan buena herramienta para realizar este tipo de análisis.

Para cumplir con dicho propósito, se ha realizado un reanálisis de los datos de un metaanálisis publicado que pretendía evaluar la prevalencia a nivel global de dos efectos adversos graves, que son el síndrome de Steven Johnson (SJS, por sus siglas en inglés) y la necrólisis epidérmica tóxica (TEN, por sus siglas en inglés), en aquellos pacientes que han consumido antibióticos (15). En este TFM se utilizan las siglas en inglés, ya que son generalmente reconocidas en la literatura científica.

Tanto el SJS como TEN son consideradas enfermedades raras que se caracterizan por una extensión de la necrosis epidérmica y el desprendiendo de la piel (16). La diferencia entre estas dos enfermedades es la cantidad de superficie corporal que se ve afectada (SJS: <10%, solapamiento SJS/TEN: 10-30% y TEN: >30%) (16). Se asocian con una elevada mortalidad, siendo el diagnóstico temprano crucial para un correcto manejo de la enfermedad. Se considera que los fármacos son los principales desencadenantes, entre los cuales se encuentran los antibióticos (16).

No se conoce el mecanismo fisiopatológico por el cual se produce esta enfermedad, sin embargo, una de las teorías más sonadas es que los fármacos producen una reacción de hipersensibilidad tipo IV (16), también conocida como hipersensibilidad retardada. Esta respuesta inmunitaria es mediada por los linfocitos T tras varios días de exposición al agente desencadenante (17).

Conocer la prevalencia de estas enfermedades en los pacientes expuestos permite evaluar el riesgo, el rápido diagnóstico y, por lo tanto, un mejor tratamiento. En el caso de los antibióticos, en 2023 se estimó que alrededor de 34,4 billones de dosis diarias definidas (unidad de medida para estimar la exposición a un fármaco) y se prevé que este número pueda aumentar hasta un 52% en 2030 (18). Lo cual incrementará el número de personas que padezcan SJS y TEN. Por esta razón, la correcta estimación de su prevalencia se convierte en un aspecto crucial y el objetivo de estudio del presente TFM.

2. OBJETIVOS

Objetivo Principal

El objetivo principal de este TFM es evaluar la idoneidad y caracterizar la aplicación de los GLMMs como alternativa a los modelos convencionales para el metaanálisis de proporciones.

Objetivos Específicos

1. Explorar el marco teórico del metaanálisis de proporciones:
 - a. Analizar y comparar las principales metodologías de metaanálisis de proporciones descritas en la literatura científica.
 - b. Sintetizar las ventajas y limitaciones de cada metodología, prestando especial atención a su aplicabilidad.
2. Validar la aplicación de los GMMLs en un caso práctico:
 - a. Reanalizar un estudio publicado que estima la prevalencia del SJS y TEN asociada a antibióticos.
 - b. Evaluar la integridad de este tipo de modelos cuando se presentan datos atípicos.
 - c. Demostrar la implementación de los GMMLs en este contexto, comparando los resultados obtenidos con los del estudio original y evaluando la robustez de esta metodología.

3. METODOLOGÍA

En la presente sección se va a describir la metodología usada para llevar a cabo el presente TFM.

Metaanálisis: metodología convencional

Primero, se explica qué es un metaanálisis, sus características principales y los modelos más habituales para realizarlo. Después de esta introducción, se dedica atención al metaanálisis de proporciones con mayor detalle.

Como se ha comentado anteriormente, un metaanálisis es una técnica estadística utilizada para combinar y sintetizar cuantitativamente los resultados de múltiples estudios independientes que investigan una misma pregunta de investigación o un tema similar (6).

Tal cual se indica en la definición, los datos proceden de estudios independientes, de los cuales se obtiene la estimación del tamaño del efecto.

El tamaño del efecto (en inglés, *effect size*) en sí mismo se usa para dar respuesta a la pregunta de nuestro metaanálisis. El tamaño del efecto es la cuantificación de la asociación entre la exposición, bien una intervención o tratamiento, y el evento. El objetivo principal del metaanálisis es estudiar la consistencia de los efectos observados en estudios individuales y dar una estimación del efecto global. Este tamaño del efecto está unido a un IC, generalmente al 95% (5). Este da información acerca de la precisión con la que se ha realizado esta medida y, por lo tanto, de su variabilidad. Estudios con ICs más anchos implican una menor precisión y, por lo tanto, una mayor variabilidad.

La variabilidad es otro aspecto muy importante en los metaanálisis, ya que va a determinar el peso que se le da a cada estudio. Con el peso se hace referencia a cuánta influencia tiene ese estudio en la estimación global. La manera más común de determinar el peso es la inversa de la varianza (Ec. 1). De esta manera se prioriza a los estudios con menor variabilidad a la hora de obtener la estimación del efecto.

$$W_i = \frac{1}{V_{Y_i}} \quad (\text{Ec. 1})$$

Principalmente existen dos modelos en el ámbito del metaanálisis, los cuales se diferencian principalmente en el manejo de la variabilidad. Estos se conocen como el modelo de efectos fijos y el modelo de efectos aleatorios.

Modelo de efectos fijos

El modelo de efectos fijos es aquel que considera que existe un único tamaño del efecto real y que las diferencias entre las observaciones de los diferentes estudios se deben a errores de muestreo. En el libro de Borenstein *et al.* (2009) (5), se incluye la siguiente figura que ejemplifica el impacto del error de muestreo en la estimación de los efectos individuales de cada estudio:

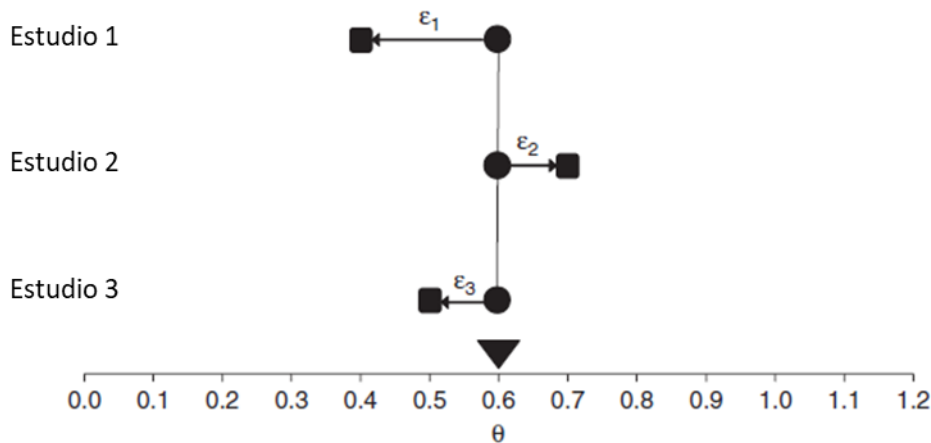


Figura 1. Representación del error de muestreo. Figura adaptada del libro de Borenstein *et al.* (2009) (5).

Como se puede ver en la Figura 1, la estimación del efecto observado (Y_i) en un estudio viene dado por la media poblacional (θ) más el error de muestreo de dicho estudio (ε_i). Es importante destacar que el modelo asume que los errores son aleatorios y siguen una distribución normal. De manera general, esto se puede describir matemáticamente mediante la siguiente expresión:

$$Y_i = \theta + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad (\text{Ec. 2})$$

Es decir, si el tamaño muestral fuera infinito (o el de toda la población), el error de muestreo sería 0 y por lo tanto todos los estudios reportarían el mismo tamaño de efecto real. En la práctica, el tamaño de la muestra es finito, por lo que el tamaño del efecto se ve influenciado por este error de muestreo, como se puede ver en la Figura 1.

Por lo tanto, para obtener una estimación del efecto global (M) más precisa, se calcula una media ponderada por el peso de cada estudio (W_i), penalizando aquellos estudios con mayor error o, lo que es lo mismo, con mayor variabilidad.

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i} \quad (\text{Ec. 3})$$

Como se muestra en la (Ec. 1), el peso viene dado por la inversa de la varianza de los estudios individuales. Por otro lado, la estimación global también tiene una varianza asociada, la cual se corresponde con la inversa de la suma de los pesos

$$V_M = \frac{1}{\sum_{i=1}^k W_i} \quad (\text{Ec. 4})$$

Dado que se asume que se sigue una distribución normal, se pueden calcular los ICs al 95% para la estimación global de la siguiente manera:

$$LL_M = M - 1,96 \times SE_M \quad (\text{Ec. 5})$$

$$UL_M = M + 1,96 \times SE_M \quad (\text{Ec. 6})$$

Donde SE_M se corresponde con $\sqrt{V_M}$.

Modelo de efectos aleatorios

A diferencia del modelo de efectos fijos, que asume un único tamaño del efecto verdadero y que las diferencias entre estudios son solo por errores de muestreo, el modelo de efectos aleatorios reconoce que el tamaño del efecto puede variar entre estudios. Es decir, este modelo acepta que existen diferencias causadas por diversos factores, algunos de los cuales podrían no haberse considerado en los estudios, e incluso desconocerse si influyen en la estimación del tamaño del efecto. Ejemplos de estos factores incluyen la edad de los pacientes, la actividad física, los cuidados del centro hospitalario, entre otros.

En este modelo se asume que los verdaderos tamaños de efecto observados en los diferentes estudios están normalmente distribuidos, por lo que se podrá estimar el rango en el que se encuentren la mayoría de estos efectos verdaderos. De nuevo, siguiendo el libro de Borenstein *et al.*, (2009) (5), esto se puede representar como en la siguiente figura:

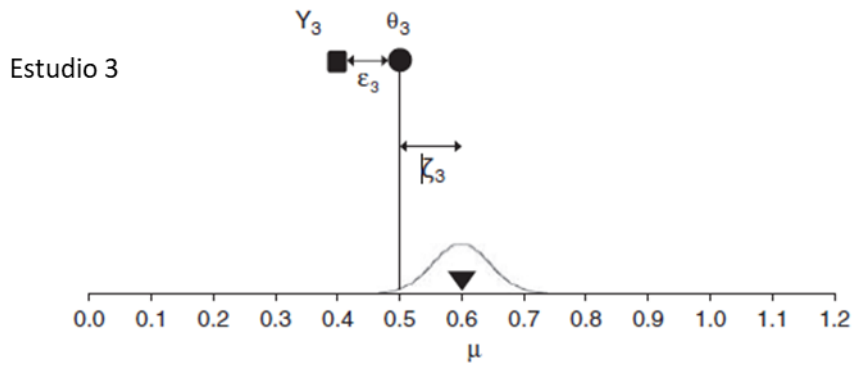


Figura 2. Modelo de efectos aleatorios - tamaño del efecto real y el observado en un estudio. Imagen adaptada del libro de Borenstein et al. (2009) (5)

En la Figura 2 se puede observar cómo el tamaño del efecto reportado en un estudio determinado (Y_3) contiene un error de muestreo (ε_3) al igual que en el modelo de efectos fijos. Sin embargo, aparece otro término que es la variabilidad inter-estudio (ζ_3), que permite explicar por qué los tamaños de los efectos reales no tienen por qué ser exactamente iguales entre los diferentes estudios. Esto se puede escribir de manera general:

$$Y_i = \theta + \zeta_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2); \quad \zeta_i \sim N(0, \tau^2) \quad (\text{Ec. 7})$$

De acuerdo con la expresión dada en la (Ec. 7), para predecir cuánto de alejado se encuentra el tamaño del efecto observado del tamaño del efecto real de la población tendremos que considerar la varianza real que existe entre los estudios, la cual se conoce como τ^2 (τ^2). La distancia entre lo observado y el real viene dada por su desviación estándar (τ), es decir, la distribución estándar de los tamaños de los efectos reales.

Un aspecto muy importante a destacar es la relación que hay entre ambos modelos, en el caso de que $\tau^2 = 0$, el modelo de efectos aleatorios sería equivalente al de efectos fijos. Sin embargo, si directamente se opta por la opción de los efectos fijos, no estaríamos contemplando la posibilidad de la variabilidad real entre los tamaños del efecto. Por lo tanto, seleccionar el modelo correcto para iniciar el análisis debe basarse en criterios científicos que permitan evaluar si se puede asumir un tipo de modelo u otro.

Respecto a la estimación global, esta sigue un razonamiento igual que en el modelo de efectos fijos, solo que ahora se incluye un término más que es la variabilidad entre-estudios. Por lo tanto, el peso que recibirá cada estudio viene dado por la siguiente expresión:

$$W_i^* = \frac{1}{V_{Y_i}^*} = \frac{1}{V_{Y_i} + T^2} \quad \text{Ec. 8}$$

A diferencia del modelo de efectos fijos, ahora se incluye T^2 que hace referencia a la estimación de la varianza muestral (que incluye la varianza intra-estudio, así como la varianza entre-estudio). La forma convencional de calcularlo es mediante el método de DerSimonian and Laird, en el que se estima que:

$$T^2 = \frac{Q - df}{C} \quad \text{Ec. 9}$$

donde Q es el estadístico de heterogeneidad de Cochran, k es el número de estudios, y C es un factor de ponderación.

$$Q = \sum_{i=1}^k W_i Y_i^2 - \frac{(\sum_{i=1}^k W_i Y_i)^2}{\sum_{i=1}^k W_i}, \quad \text{Ec. 10}$$

$$df = k - 1$$

$$C = \sum W_i - \frac{(\sum W_i)^2}{\sum W_i} \quad \text{Ec. 11}$$

Por todo lo demás, la estimación del tamaño del efecto global y su IC sigue el mismo razonamiento que el modelo de efectos fijos. En el cual se calcula la media ponderada con los pesos de los estudios, pero ahora considerando esta variabilidad entre estudios.

Para llevar a cabo correctamente el metaanálisis hay que tener en cuenta principalmente la naturaleza de las variables de resultado empleadas en los estudios primarios. Así, cuando los desenlaces son de naturaleza cuantitativa continua, como una puntuación en una escala de ansiedad o un nivel de glucosa en sangre, la medida del efecto comúnmente utilizada es la diferencia de medias o la diferencia de medias estandarizada (5). Por el contrario, si las variables son dicotómicas y representan la ocurrencia de un evento (p. ej., éxito o fracaso de un tratamiento), se emplean métricas como el Odds Ratio (OR) o el Riesgo Relativo (RR) (5,19). Otros tipos de datos, como las proporciones o los coeficientes de correlación, a menudo requieren una preparación específica. En estos casos, es habitual aplicar transformaciones para estabilizar sus varianzas y asegurar que se cumplan las asunciones del modelo, como la distribución normal de los estimadores del efecto (10).

Metaanálisis de proporciones

A diferencia del resto de metaanálisis en los que se estudia la comparación entre tratamientos a través de medidas de efecto relativas, como, por ejemplo, OR, RR, etc., el metaanálisis de proporciones trata de estimar la proporción general, independientemente de la transformación usada.

En la práctica, se usa la metodología de dos pasos, que implica una transformación de los datos previa a ser analizados. Una vez aplicado el modelo correspondiente, generalmente el modelo de efectos aleatorios, se lleva a cabo de nuevo una transformación inversa, que devuelve el resultado a la escala original (10).

En este TFM se ha referenciado a este tipo de metodología y al modelo resultante, como “modelo convencional de dos pasos”, “modelo convencional”, “metodología de dos pasos”, etc. Todos ellos hacen referencia al modelo descrito a continuación.

De manera general, cada uno de los estudios incluidos en un metaanálisis de proporciones contiene un número de casos (e_i) en una muestra definida (n_i), por lo que la prevalencia estimada es $\hat{p}_i = \frac{e_i}{n_i}$. Las transformaciones más comunes para este tipo de metaanálisis son: log, logit, arcoseno y la transformación de doble arcoseno de Freeman–Tukey (20–22). Esta transformación tiene como objetivo principal el de estabilizar la variabilidad, así como aproximar los datos a una distribución normal.

Estas transformaciones tienen algunas limitaciones, por ejemplo, las transformaciones log y logit no pueden manejar directamente los valores nulos, que suelen corregirse generalmente añadiendo 0,5 (10). En el caso de las transformaciones basadas en arcoseno, presentan otras limitaciones, como su difícil comprensión y en algunas ocasiones no siguen una distribución normal.

En el caso de la transformación de doble arcoseno de Freeman–Tukey presenta una transformación inversa muy compleja, la cual depende de un tamaño muestral que representa el resultado final. Este tamaño muestral no se define correctamente en la configuración del metaanálisis. Dependiendo de lo que se seleccione, puede dar lugar a resultados diferentes que produzcan malinterpretaciones de los resultados (20).

Recientemente, gracias al avance en la tecnología, se pueden usar metodologías que hace unos años no eran tan comunes. Por lo que el uso de los GLMMs supone una alternativa para este tipo de análisis que no presenta las limitaciones indicadas anteriormente.

Alternativa: Uso de los modelos mixtos lineales generalizados (GLMMs)

Varios investigadores y estadísticos han estado desarrollando y poniendo a punto esta metodología para su uso en los metaanálisis. El presente TFM ha sido realizado siguiendo las indicaciones proporcionadas por Lin y Xu (2020) (10) en su artículo titulado: *Meta-analysis of proportions using generalized linear mixed models*.

Desde el punto de vista de la estadística, se ha asumido que típicamente el conteo de eventos sigue una distribución binomial. No obstante, los modelos de efectos fijos y aleatorios no trabajan directamente con esta distribución, haciendo necesario el uso de las transformaciones para ajustarlos a una distribución normal.

Sin embargo, los GLMMs permiten resolver este problema en un único paso. Esta metodología modela directamente los conteos a través de una distribución binomial, y además tiene en cuenta el total de la variabilidad, ya sea inter- como intra- estudios (20). Estos modelos usan unas funciones de enlace que transforman cada una de las proporciones de los diferentes estudios a una escala lineal.

Las funciones de enlace más comunes son el enlace log, logit y probit. Sin embargo, otros como el cauchy o el cloglog también pueden usarse.

De manera general, los casos de una determinada enfermedad (e_i) siguen una distribución binomial ($e_i \sim bin(n_i, p_i)$), de forma que se especifica la función de enlace ($g(\cdot)$), se puede especificar el siguiente modelo:

$$\begin{aligned} g(p_i) &= \mu + \theta_i, & \text{Ec. 12} \\ \theta_i &\sim N(0, \tau^2) \end{aligned}$$

Donde μ es la proporción general estimada por el modelo en la misma escala que el enlace ($g(\cdot)$) y τ^2 es la variabilidad entre estudios.

A lo largo del trabajo, al modelo de efectos aleatorios especificado usando los GLMMs se le va a llamar directamente GLMMs para no confundirlo con el modelo de efectos aleatorios convencional.

Las ventajas del uso de estos modelos son que no presentan las limitaciones comentadas anteriormente, como puede ser la corrección de ciertos valores nulos, proporcionan resultados más interpretables, no hay problemas de correlación intrínseca debido a las transformaciones log o logit (20). Por lo tanto, es una herramienta muy adecuada para este tipo de análisis. Actualmente, ya existen paquetes conocidos de R para metaanálisis (función “metaprop” del paquete “meta” (23)) que ya los han incorporado.

Heterogeneidad

Como se ha ido comentando, uno de los factores más importantes a tener en cuenta a la hora de realizar un metaanálisis es la heterogeneidad. Esta se define como la variabilidad real que existe entre los diferentes estudios que han sido incluidos en el análisis (5).

Su principal impacto afecta a la interpretabilidad de los resultados. Este tipo de análisis son incluidos de manera rutinaria en los paquetes estadísticos destinados a los metaanálisis. Sin embargo, los GLMMs no cuentan con dichas estimaciones. Por lo tanto, en este TFM se ha llevado a cabo la obtención de los estadísticos que cuantifican la heterogeneidad a partir de los resultados de los GLMMs.

Para ello, a partir de la información de libros (5), artículos (24,25) e información publicada de los paquetes estadísticos destinados a los metaanálisis de proporciones, como la función “metaprop”, se ha desarrollado en el presente TFM el código necesario para obtener las métricas para cuantificar la heterogeneidad más comunes. El código en el que se incluye todo el desarrollo para el cálculo de estos índices se encuentra en el Anexo 2 – CÓDIGO DE R.

A continuación, se describen los diferentes estadísticos. En primer lugar, el estadístico Q se utiliza para determinar si los tamaños de efecto individuales de los diferentes estudios incluidos en un metaanálisis son homogéneos. Es importante indicar que, a partir de este, se pueden estimar el resto de parámetros que se describen a continuación.

Matemáticamente, Q (Ec. 10) es equivalente a una suma ponderada de las desviaciones al cuadrado de los tamaños de efecto (Y_i) de cada uno de los diferentes k estudios respecto al tamaño del efecto global (M), donde la ponderación es el peso que recibe el estudio en concreto (W_i).

$$Q = \sum_{i=1}^k W_i (Y_i - M)^2$$

Ec. 13

El valor de Q sigue una distribución de chi-cuadrado (χ^2) con $k-1$ grados de libertad, donde $k-1$ corresponde con el valor estandarizado de la variación esperada debida a un error de muestreo. Si el p-valor asociado es menor que un umbral de significación predefinido (generalmente $\alpha=0,05$), se rechaza la hipótesis nula y se concluye que existe una heterogeneidad estadísticamente significativa entre los estudios. Esto se conoce como prueba de homogeneidad. Cabe destacar que la presencia de heterogeneidad estadísticamente significativa no implica que directamente haya un alto grado de esta.

Es importante destacar que el estadístico Q tiene una limitación significativa: su poder estadístico es bajo en metaanálisis con un número pequeño de estudios, lo que puede llevar a no detectar la heterogeneidad incluso cuando está presente. Por esta razón, se complementa con otros estadísticos como el I^2 y el τ^2 .

Para calcular Q en los GLMMs, se realizan cálculos independientes del modelo, ya que no es posible obtener directamente el peso de cada estudio ni calcular los estimadores de heterogeneidad (excepto τ^2). Por ello, se procedió a estimar la variabilidad entre los estudios para cada enlace utilizando el método delta.

El método delta es un procedimiento basado en la aproximación de las series de Taylor, el cual permite aproximar la media y la varianza de una función a una variable aleatoria (25). En el caso de la varianza, aproxima mediante la siguiente expresión:

$$Var(g(\hat{\theta})) = \left(g'(\hat{\theta})\right)^2 Var(\hat{\theta}) \quad \text{Ec. 14}$$

En el caso del presente TFM, la distribución especificada en los GLMMs es una binomial, por lo que su varianza se corresponde con $\frac{p(1-p)}{n}$. Usando esta expresión se ha estimado cada una de las variabilidades para los diferentes enlaces. De esta manera se pudieron estimar los pesos y ya por lo tanto se pudo calcular el estadístico Q para cada tipo de enlace. El código implementado por el autor del TFM se encuentra en el Anexo 2 – CÓDIGO DE R.

El estadístico I^2 fue propuesto por Higgins *et al.* (2003) (26), es una medida de la heterogeneidad que cuantifica la proporción de la varianza total en un metaanálisis atribuible a la heterogeneidad real entre los estudios, en lugar de a un error muestral. La expresión de este estadístico es la siguiente:

$$I^2 = \frac{(Q - df)}{Q} \quad \text{Ec. 15}$$

Existen otras maneras de calcular este estadístico, como muestra Higgins *et al.* (2002) (24) en su artículo:

$$I^2 = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{H^2 - 1}{H^2} \quad \text{Ec. 16}$$
$$H^2 = \frac{Q}{k - 1}$$

En su artículo, Higgins *et al.* (2003) (26) proponen unos valores arbitrarios para clarificar la heterogeneidad en función de I^2 , indicando que los valores de 25%, 50% y 75% corresponden con los límites de baja, moderada y alta heterogeneidad, respectivamente.

A pesar de estas guías, es fundamental interpretar el valor de I^2 en el contexto del metaanálisis específico, ya que su interpretación puede variar. Como se discutirá más adelante, en metaanálisis de proporciones donde estas estimaciones arbitrarias pueden no ser completamente apropiadas.

Para mejorar su interpretación, se han calculado los ICs al 95% para este estadístico, de acuerdo a las ecuaciones especificadas en el libro de Borenstein *et al.* (2009)(5).

Por último, se ha obtenido la variabilidad entre estudios o variabilidad real conocida como τ^2 , utilizada principalmente en el modelo de efectos aleatorios. A diferencia de medidas relativas como el I^2 , que indica la proporción de la variabilidad total debida a la heterogeneidad, τ^2 estima la varianza real entre los efectos verdaderos de los estudios incluidos. Su principal ventaja es que se expresa en la misma escala que el tamaño del efecto, lo que permite una interpretación más intuitiva de la dispersión de los resultados. La expresión para su cálculo se muestra en la Ec. 9, sin embargo, este valor se puede extraer directamente de los GLMMs.

Comparación entre los modelos

Con el objetivo de identificar cuál de los modelos es el más adecuado para realizar un metaanálisis de proporciones, se compararon entre sí.

La primera es la comparación directa entre las estimaciones globales de los tamaños del efecto obtenidas con los diferentes modelos. Además, se ha realizado una estimación del cambio relativo de cada una de las diferentes estimaciones respecto al modelo considerado como referencia, y esto ha sido representado en un gráfico, como puede observarse en el apartado de resultados. Esta metodología ha sido empleada en estudios

publicados y por lo tanto han sido considerados para realizar este TFM (22). En el artículo, se toma como umbral de 1,2, siendo los valores por encima de este umbral considerados como muy elevados. En el TFM se ha seguido el mismo criterio.

Por otro lado, se ha realizado una comparación mediante el Criterio de Información de Akaike (AIC) (27). Es un método utilizado para comparar la calidad de diferentes modelos estadísticos. AIC busca un equilibrio entre la complejidad del modelo y su capacidad para ajustarse a los datos. Un valor de AIC más bajo indica un modelo superior.

La fórmula del AIC es la siguiente:

$$AIC = 2k - 2\ln(\hat{L}) \quad \text{Ec. 17}$$

Donde k representa el número de parámetros del modelo, y \hat{L} es el valor máximo de la función de verosimilitud para el modelo. Por lo tanto, definir modelos con los mismos efectos aleatorios, solo cambiando el tipo de enlace, las diferencias observadas pueden deberse directamente al enlace en sí.

Además, también se ha comparado la estimación cruda de la prevalencia de cada estudio respecto a lo que predice el modelo en los efectos aleatorios. El objetivo de esta comparativa es encontrar patrones o irregularidades que pudieran afectar a la cuantificación del efecto global o de la heterogeneidad.

En cuanto a la caracterización de la heterogeneidad, se ha realizado una comparación directa entre los estadísticos calculados en los diferentes modelos. Además, se han realizado forestplots, cuyo código aparece en el Anexo 2 – CÓDIGO DE R, con el fin de evaluar la heterogeneidad gráficamente.

Por otro lado, la robustez de un modelo es un pilar fundamental para garantizar su fiabilidad y aplicabilidad. Un modelo robusto es aquel que mantiene su integridad, su capacidad predictiva y la validez de sus inferencias, incluso en situaciones que se desvían de las condiciones ideales o en presencia de datos atípicos.

Para llevar a cabo una evaluación exhaustiva acerca de la robustez de los GLMMs para el metaanálisis de proporciones, se han planteado una serie de situaciones que representan desafíos comunes en los conjuntos de datos reales relacionados con las proporciones. En este caso se va a tratar principalmente el efecto de los diferentes tamaños muestrales, la presencia de valores atípicos y la heterogeneidad.

Análisis de Subgrupos por Tamaño Muestral

Se sabe que tamaños de muestras muy pequeños aumentan la variabilidad de las estimaciones de los estudios independientes, por lo tanto, afectan al peso que se les da, así como la heterogeneidad.

Para investigar si el tamaño muestral de los estudios influía en la prevalencia estimada, se realizó un análisis de subgrupos. Primero, los estudios se estratificaron en tres grupos de igual tamaño (terciles) según su número de participantes (N). Los puntos de corte para definir los grupos fueron 51 y 82 participantes, resultando en las categorías de tamaño bajo, medio y alto.

Posteriormente, estas categorías se introdujeron como variables en los diferentes modelos ajustados, tanto en los GLMM como en el modelo convencional. El objetivo era estimar una prevalencia específica para cada subgrupo y determinar si existían diferencias estadísticamente significativas entre ellos. Finalmente, se comparó el ajuste de los cinco modelos con funciones de enlace y el modelo convencional mediante el criterio AIC para este escenario específico.

Simulación con Baja Heterogeneidad

Para evaluar el comportamiento de los modelos en un escenario con poca variabilidad, se simuló un conjunto de datos. Para ello se generaron 38 estudios con tamaños muestrales aleatorios entre 100 y 500. Las prevalencias de cada estudio se extrajeron de una distribución normal con una media de 0,26 y una desviación estándar muy pequeña (0,03). Esto aseguró que la variabilidad real entre los estudios fuera mínima.

Sobre este conjunto de datos simulado, se ajustaron los cinco modelos GLMM y el modelo convencional de dos pasos. El análisis se centró en verificar si los modelos eran capaces de estimar correctamente la prevalencia media (0,26) y en cuantificar la heterogeneidad (I^2 y τ^2). La comparación del AIC permitió identificar qué modelo ofrecía un mejor ajuste en estas condiciones de baja variabilidad.

Simulación con Alta Heterogeneidad

Además, para conocer cómo se comportan los modelos cuando hay una alta heterogeneidad, se simuló otra muestra distinta. El procedimiento fue análogo al de baja heterogeneidad, generando 38 estudios con la misma prevalencia media (0,26). Sin

embargo, la desviación estándar se incrementó significativamente (0,35). Este cambio introdujo una dispersión considerable en las prevalencias de los estudios simulados.

De nuevo, se aplicó la serie completa de modelos (cinco GLMM y el convencional) a estos datos. El objetivo principal era observar cómo la alta heterogeneidad afectaba a las estimaciones de la prevalencia global y a la precisión de los ICs. Se calcularon los estimadores de heterogeneidad para confirmar el alto grado de variabilidad y se utilizó el AIC para comparar el rendimiento del modelo en este escenario.

Análisis de Sensibilidad con Prevalencias Altas

Finalmente, se realizó un análisis de sensibilidad para comprobar la estabilidad de los modelos al incluir estudios con prevalencias extremas. Para ello, por un lado se incluyeron 3 estudios con una prevalencia muy baja (1×10^{-5} , 2×10^{-5} y 3×10^{-5}) y por otro lado se han incluido 3 estudios con prevalencias muy altas (0,97, 0,98 y 0,99). De esta manera se podrá evaluar si todos los modelos son capaces de trabajar con estos valores o bien hay alguno que genere algún tipo de problema.

El propósito era evaluar la estabilidad y la validez de las estimaciones de cada modelo, ya que las proporciones cercanas a 0 o 1 pueden ser problemáticas para ciertas funciones de enlace. Se ajustaron los cinco GLMM y el modelo convencional a este nuevo conjunto de datos aumentado y se compararon las prevalencias agrupadas resultantes para detectar posibles sesgos o fallos en la convergencia del modelo.

Datos

Los datos seleccionados para llevar a cabo el presente TFM proceden de un estudio publicado que se titula: *Worldwide Prevalence of Antibiotic-Associated Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis A Systematic Review and Meta-analysis*. Este estudio fue realizado por Lee *et al.* (15), en el año 2023, en el cual trataron de evaluar la prevalencia de SJS/TEN en pacientes que habían consumido antibióticos a nivel mundial. Para ello llevaron a cabo una revisión sistemática con metaanálisis en la que se incluyeron estudios que describían adecuadamente el origen de SJS/TEN y especificaban los antibióticos asociados a estas enfermedades. Un total de 64 estudios fueron seleccionados que cumplían con los criterios de selección propuestos por los autores para ser incluidos en la revisión sistemática (Registro en PROSPERO: CRD42022303846). Sin embargo,

de estos 64 estudios solo 38 cumplieron los requisitos para ser incluidos en el metaanálisis.

Se trata de 38 estudios observacionales retrospectivos en los que se incluyeron 2917 pacientes de más de 20 países. Más específicamente, 16 estudios se llevaron a cabo en Asia, 14 en América del Norte, 4 en Europa, 3 en África y 1 en Australia. La mayoría de los estudios proporcionaban datos de un único país; solo 2 estudios reportaron datos de más de 2 países.

Por lo tanto, en lo que respecta a este TFM, se extrajeron los datos del artículo publicado, así como de su material suplementario con el fin de obtener una base de datos inicial. Los datos se caracterizan, principalmente, por el autor del artículo, el país de origen, la muestra estudiada, los casos de TEN y SJS, población (adulta, niños, no especificado o ambos), edad, número de mujeres, casos de muerte y el número de casos asociado a cada causa (tipos de antibióticos, otros fármacos u otras causas). En el Anexo 1 – BASE DE DATOS se muestra la tabla en la que se incluyen los datos que han sido usados para el presente TFM.

Como se puede observar en el Anexo 2 – CÓDIGO DE R, en la primera parte de limpieza y depuración de los datos se observó que los datos presentados en el material suplementario no terminaban de coincidir con los datos reportados por los autores en el artículo final. Por lo que se tuvo que llevar a cabo un proceso de corrección, revisando la fuente primaria de aquellos estudios en los que no coincidía el número de casos, es decir, el artículo original.

Software estadístico

Para llevar a cabo la depuración, el análisis y la visualización de los datos, se ha utilizado el lenguaje de programación R. Este lenguaje de programación es comúnmente usado en el análisis estadístico y en la ciencia de datos. R se distingue por su naturaleza de código abierto, lo que garantiza la transparencia y reproducibilidad de los resultados.

Sin embargo, como bien se sabe, trabajar directamente con R es muy costoso; se ha empleado RStudio, un entorno de desarrollo integrado (IDE, por sus siglas en inglés) que optimiza el flujo de trabajo. RStudio proporciona herramientas esenciales para la gestión de proyectos, la depuración de código y la creación de informes, permitiendo una experiencia más intuitiva y eficiente (28).

Respecto a los paquetes usados, cabría destacar los paquetes “tidyverse”, “lme4”, “meta” y “metafor”.

En primer lugar, el paquete “tidyverse” es el que se ha usado para gestionar la depuración de los datos, manejo de la base de datos y la representación gráfica. En realidad, más que un paquete se trata de un conjunto de paquetes que se usan para las diferentes tareas más comunes dentro del análisis de datos de una forma más sencilla e intuitiva. Entre estas tareas pueden destacarse la importación de datos, organización, visualización, manipulación y la programación. Por lo tanto, “tidyverse” ofrece un lenguaje sencillo que permite resolver los retos más comunes de la ciencia de datos mediante el uso de R (29).

El paquete “lme4” permite el análisis de modelos de efectos mixtos lineales y no lineales (30). Su principal propósito es permitir el modelado de datos con estructuras jerárquicas o anidadas, donde las observaciones no son independientes. Esto es común en estudios longitudinales o en datos de encuestas con estructuras de clústeres. Dentro de este paquete encontramos la función “glmer”, la cual es la que permite ajustar los GLMMs, en la cual los datos no siguen una distribución normal, como es el caso de los conteos y proporciones (31). El enlace actúa como un puente, transformando el valor esperado de la variable de respuesta para que su relación con los predictores se vuelva lineal. Esto permite aplicar la estructura de un modelo lineal a datos con distribuciones no normales, como el logit para proporciones o el log para conteos (30).

Por último, destacar los paquetes “meta” (23) y “metafor” (32). Estos paquetes han servido como referencia a la hora de elaborar este TFM. El paquete “meta” contiene la función “metaprop”, la cual incluye los GLMMs para el metaanálisis de proporciones (23). Por otro lado, el paquete “metafor” ofrece la función “rma”, la cual es muy completa y flexible (32). Con esta función ha sido con la que se ha conseguido replicar el análisis llevado a cabo por los autores en su artículo.

4. RESULTADOS

A continuación, se exponen los resultados obtenidos en el presente TFM.

Descripción de los datos

Antes de iniciar el análisis de las prevalencias, se ha realizado un análisis descriptivo de los datos con los que se ha trabajado.

Los datos proceden de 38 estudios seleccionados por los autores en los que se especifican casos de SJS o TEN y en los que se describieran las etiologías y cómo se procedió al diagnóstico de la enfermedad. Los artículos debían estar publicados en inglés y podían proceder de cualquier parte del mundo.

Por otro lado, se han representado las prevalencias de cada uno de los estudios mediante un histograma (Gráfico 1) con el fin de representar su distribución.

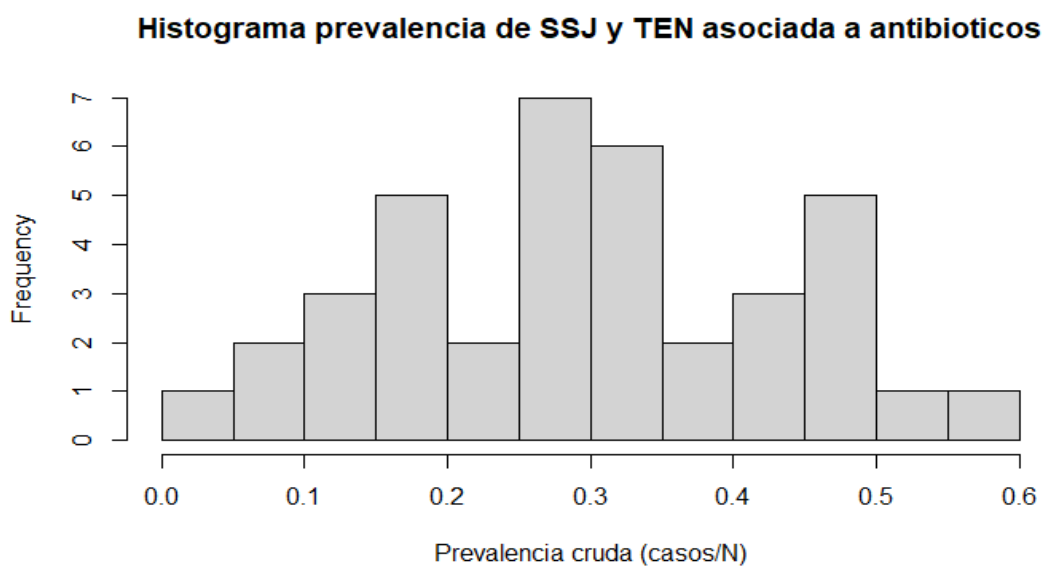


Gráfico 1. Representación de las prevalencias de los estudios incluidos en el metaanálisis de Lee *et al.* (2023) (15).

El Gráfico 1 representa la distribución muestral de las prevalencias de los casos SSJ y TEN asociados a antibióticos de los estudios que fueron incluidos en el metaanálisis del estudio de Lee *et al.* (2023) (15). Como se puede observar, no parece que haya una distribución definida o a priori conocida. Para continuar caracterizándola se ha realizado un test de normalidad.

El test de normalidad de Shapiro-Wilk no permite rechazar la hipótesis nula de normalidad ($p = 0,8546$). Sin embargo, la poca muestra ($N = 38$ estudios) y la forma irregular del histograma no permiten sacar conclusiones definitivas respecto a la normalidad de la distribución. Afortunadamente, esto no sugiere un problema, ya que una de las principales ventajas de los GLMMs es que no tienen por qué seguir una distribución normal. Como se ha comentado en la sección de Metodología (Alternativa: Uso de los modelos mixtos lineales generalizados (GLMMs)), como la variable es un conteo, se ha utilizado una distribución binomial para modelizarlos.

En el caso de los modelos convencionales que sí asumen una distribución normal, puede requerirse el uso de transformaciones, tal y como se ha descrito en la sección de Metodología (Metaanálisis de proporciones).

Por otro lado, se han descrito de forma cuantitativa la distribución de las prevalencias a través la siguiente tabla resumen:

Prevalencia	N estudios	Frecuencia relativa	Frecuencia acumulada
0-0,1	3	0,0789	0,0789
0,1 – 0,2	8	0,2105	0,2894
0,2-0,3	9	0,2368	0,5263
0,3-0,4	8	0,2105	0,7368
0,4-0,5	8	0,2105	0,9473
0,5-0,6	2	0,0526	1,0000

Tabla 1. Distribución de frecuencias relativas y absolutas de las prevalencias estimadas de SJS y TEN en los 38 estudios incluidos en el artículo de Lee *et al.* (2023) (15).

Tal y como se puede observar en la Tabla 1, las prevalencias se encuentran en el rango de 0 a 0,6, con más de la mitad por debajo de 0,3, ya que la frecuencia acumulada del intervalo 0,2-0,3 es 0,5263 (52,63%). La moda se encuentra en el intervalo de 0,2 a 0,3, ya que es el que contiene la mayor cantidad de estudios (9 estudios).

Por otro lado, tanto en el Gráfico 1 como en la Tabla 1 se muestra que los valores extremos se encuentran muy poco representados (0,079 y 0,053 para los intervalos de 0-0,1 y 0,5-0,6, respectivamente). Esto hace que la mayoría de los valores de prevalencia estimados se encuentren en torno a 0,1 y 0,5 (33 de los 38 estudios, es decir, aproximadamente el 87% de las prevalencias estimadas, se encuentran dentro de este rango).

Metaanálisis de prevalencias: GLMMs vs. CONVENCIONAL.

El objetivo principal de esta sección es comparar la robustez de los GLMMs frente al modelo convencional. Para ello, en primer lugar, se han descrito las estimaciones globales del tamaño del efecto entre los diferentes tipos de enlace (log, logit, probit, etc.) y frente al modelo convencional de efectos aleatorios. Además, se ha estudiado cómo cada uno de los diferentes modelos gestiona la variabilidad de los datos.

Como se ha explicado en la sección de Metodología (Heterogeneidad), para realizar el estudio de la heterogeneidad se han calculado los estadísticos (Q , I^2 , τ^2 y τ) para cada uno de los modelos estudiados. La información detallada de sus expresiones y de sus principales características se encuentra recogida en la sección citada al inicio del párrafo. La información detallada sobre el código de R usado se encuentra en el Anexo 2 – CÓDIGO DE R.

En la

Tabla 2 se muestran los resultados obtenidos con cada uno de los modelos acerca de las estimaciones de las prevalencias, así como los intervalos de confianza (IC) al 95%.

Enlace	Prevalencia	IC 95%
Log	0,2654	0,2233-0,3152
Logit	0,2735	0,2276-0,3246
Probit	0,2762	0,2299-0,3267
Cauchit	0,2613	0,2599-0,2626
clog-log	0,2695	0,2254-0,3202
Convencional*	0,2808	0,2353-0,3287
Lee <i>et al.</i> , (2023)(15)	0,28	0,23-0,33

*Nota: la transformación arcoseno es la usada en el modelo de efectos aleatorios que imita al realizado por los autores en su trabajo.

Tabla 2. Análisis de prevalencias usando GLMMs y el modelo de efectos aleatorios convencional (de dos pasos).

Tal y como se muestra en la

Tabla 2, se ha podido recrear el análisis llevado a cabo por los autores usando el modelo de efectos aleatorios convencional y la transformación arcoseno. Las estimaciones puntuales de la prevalencia obtenidas con los modelos GLMM varían entre 0,2613 (enlace cauchit) y 0,2762 (enlace probit), lo que indica una alta consistencia en la estimación del efecto global por este tipo de metodología, independientemente del enlace utilizado. Además, estas estimaciones son muy cercanas a la del modelo de efectos aleatorios

convencional, ya sea el replicado en el presente TFM (0,2808) o el del estudio original de Lee *et al.* (2023)(15) (0,28), lo que refuerza la robustez de los resultados.

Por otro lado, en cuanto a las estimaciones de los ICs 95% se puede observar que son prácticamente iguales en todos los casos, excepto para el enlace cauchit. Este hecho se comentará más adelante.

Esta semejanza en cuanto a los ICs soportaría la robustez de la metodología utilizada, ya que su estimación es prácticamente independiente del tipo de enlace. Sin embargo, algunos modelos pueden presentar algunas ventajas en cuanto al análisis de diferentes tipos de proporciones, como se detallará más adelante.

Enlace cauchit: Error de convergencia

La función de enlace cauchit es atractiva por su robustez ante valores atípicos, ya que se basa en la distribución de Cauchy con colas más pesadas. Sin embargo, esta misma característica puede generar importantes problemas de convergencia.

Las principales razones por las que ocurren este tipo de problemas de convergencia son que [1] el enlace cauchit puede requerir hacer ajustes gigantescos para pequeñas variaciones en las probabilidades cercanas a 0 o 1. Esto dificulta la convergencia del algoritmo de optimización (33), [2] la función que el algoritmo intenta minimizar puede tener una forma más irregular, impidiendo que se encuentre una solución óptima y estable(34), [3] cuando alguno de las estimaciones tienen una proporción de 1 o 0, el modelo intenta estimar parámetros infinitos, lo que causa fallos en la convergencia (35), y [4] es un enlace muy sensible a los valores atípicos, por lo que la falta de observaciones suficientes en algunas categorías o un fuerte desequilibrio entre los resultados puede agravar la inestabilidad provocando una falta de convergencia (36).

En este caso, el problema de convergencia del enlace cauchit se debe a que la proporción estimada del estudio 7 es 0, por lo que generaría parámetros infinitos que generan dichos problemas. Para solucionar el problema, se han probado dos formas, la primera ha sido eliminar el estudio 7 y la otra darle el valor más pequeño al número de casos que tolere este tipo de enlace, es decir, corregir el valor nulo indicando el valor mínimo de casos tolerado por el enlace. En este caso, el valor mínimo tolerado es de 2. Los resultados se resumen en la Tabla 3.

Método/Enlace	Datos sin corregir	Eliminar estudio 7	Incluir valor mínimo de casos tolerado
Log	0,2654 0,2233-0,3152	0,2796 0,2403-0,3252	0,2683 0,2276-0,3164
Logit	0,2735 0,2276-0,3246	0,2867 0,2445 0,3331	0,2763 0,2321-0,3253
Probit	0,2762 0,2299-0,3267	0,2891 0,2466 0,3347	0,2791 0,2346-0,3272
Cauchit	0,2613 0,2599-0,2626	0,2750 0,2365-0,3236	0,2634 0,2240-0,3142
clog-log	0,2695 0,2254-0,3202	0,2832 0,2424 0,3293	0,2723 0,2298-0,3210
Dos pasos (arcoseno)	0,2808 0,2353-0,3287	0,2921 0,2495 0,3365	0,2829 0,2389-0,3293

Tabla 3. Corrección problema de convergencia enlace cauchit. Se muestra la estimación puntual como su IC 95%.

Tal y como se muestra en la Tabla 3, ambos enfoques permiten corregir la falta de convergencia con el enlace cauchit. Sin embargo, parece que el incluir un valor muy pequeño que permita estimar la prevalencia más baja posible en el estudio que generaba la problemática parece proporcionar estimaciones similares a las que se obtienen sin hacer ninguna modificación de los datos, es decir, se seguirá teniendo en cuenta esta baja prevalencia dada por este estudio. Por lo tanto, este resultado nos indica que el uso del enlace cauchit no sería adecuado cuando tenemos proporciones muy próximas a 0.

Para evitar más problemas de convergencia, se ha continuado trabajando con los datos incluyendo el menor número de casos para que el enlace cauchit converja.

Análisis de la heterogeneidad

Tal y como se ha comentado previamente en la sección de Metodología (Heterogeneidad), existen varias formas de evaluar la heterogeneidad en un metaanálisis. Principalmente, están los métodos gráficos y los analíticos.

Respecto a los métodos gráficos, el más común es el forestplot. En este gráfico podemos observar a simple vista cuán alejadas están las estimaciones de los tamaños del efecto individuales de los diferentes estudios entre sí y respecto a la estimación global del efecto. El forestplot de los estudios evaluados en el presente trabajo se muestra en el Gráfico 2.

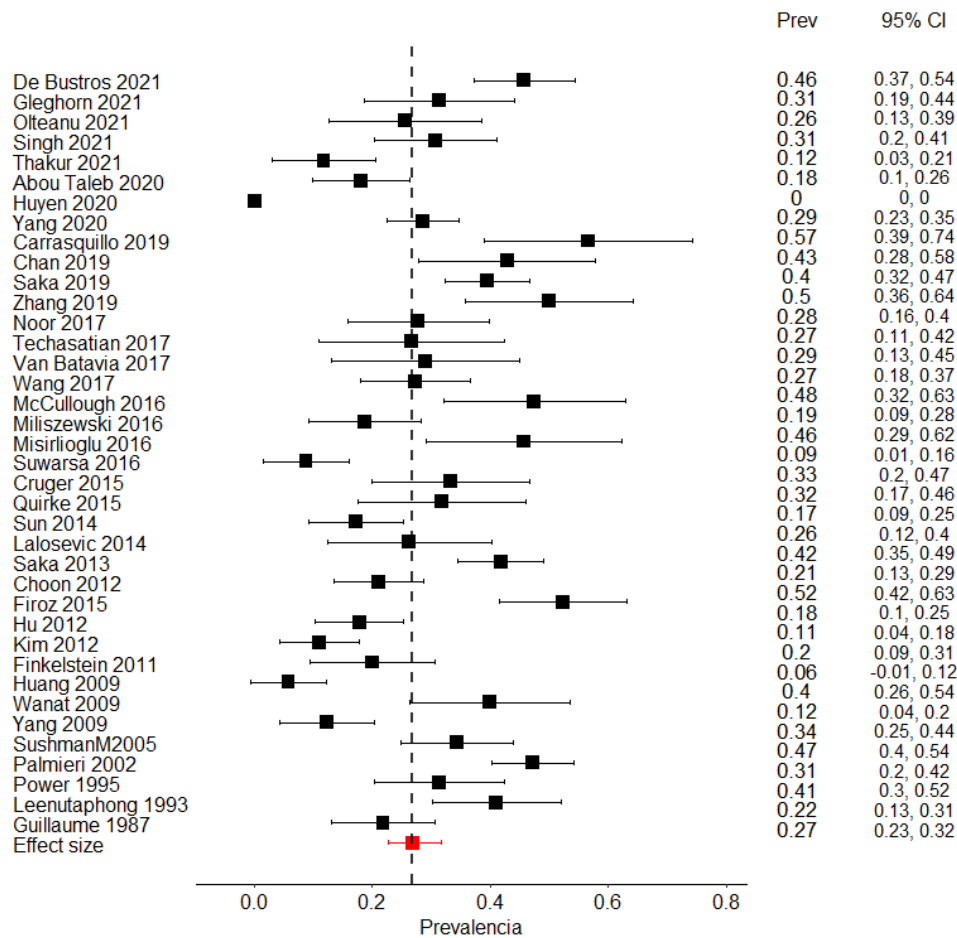


Gráfico 2. Forestplot de los 38 estudios incluidos en el metaanálisis. Estimación realizada con GLMMs y el enlace probit

Como se puede observar, los estudios parecen estar bastante repartidos a lo largo del gráfico, sugiriendo que la heterogeneidad entre estudios es real. Además, cabe destacar que hay estudios que se alejan mucho de la estimación global realizada por el modelo. En algunos casos, el IC 95% de los estudios individuales no recoge el valor de la estimación global del modelo, sugiriendo una heterogeneidad considerable.

En cuanto a la estimación con métodos analíticos, los valores obtenidos para los diferentes estadísticos se muestran en la Tabla 4

Enlace	Q	Q test	I ² (%)	I ² (IC 95%)	τ ²	τ
Log	231,90	4,59×10 ⁻²⁷	84,05	78,98-87,89	0,2156	0,4643
Logit	208,89	7,45×10 ⁻²⁶	82,29	76,48-86,66	0,4448	0,6669
Probit	225,58	6,70×10 ⁻²⁹	83,60	78,35-87,58	0,1564	0,3954
Cauchit	1798,34	0	97,94	97,62-92,23	0,5025	0,7089
clog-log	216,36	3,72×10 ⁻²⁷	82,90	77,35-87,09	0,3147	0,5609
Convencional*	272,09	< .0001	86,40	83,63-90,23	0,0209	0,1446
Lee <i>et al.</i> , (2023)(5)	NR	< .001	87	NR	0,022	NR

*Nota: la transformación arcoseno es la usada en el modelo de efectos aleatorios que imita al realizado por los autores en su trabajo.

Tabla 4 Análisis de la heterogeneidad

En primer lugar, se puede apreciar cómo en todos los casos el valor de p para Q es significativo, lo que nos indicaría que podríamos rechazar la hipótesis nula de que todos los estudios comparten un tamaño del efecto común, y aceptar la alternativa, que el efecto no es el mismo en todos los estudios.

Respecto al estadístico I^2 , podríamos observar que en todos los casos I^2 se encontraría encima del 75% que Higgins *et al.* (2003) (26) propusieron arbitrariamente como “alta heterogeneidad”. Es más, los ICs 95% estimados para I^2 están todos por encima del 75%, por lo que podríamos concluir con un 95% de confianza que la heterogeneidad real debida a las diferencias entre los estudios es real y muy elevada.

No obstante, de acuerdo a lo especificado por Migliavaca *et al.* (2021) (37), puede que este, puede que varios factores distorsionen la estimación de I^2 en el metaanálisis de prevalencias. En su artículo, los autores indican algunos factores que pueden contribuir a que se muestren valores elevados de este estadístico en este tipo de estudios en específico. En primer lugar, comentan que verdaderamente pueden existir factores como el momento, la localización o subgrupo de pacientes, que muestren una verdadera heterogeneidad. Por otro lado, por el elevado tamaño muestral de los estudios observacionales, típicos para la estimación de prevalencias de enfermedades, se estiman unas varianzas muy pequeñas, lo que promueve la heterogeneidad entre estudios. Por último, es que la naturaleza no comparativa de la variable aumenta la heterogeneidad. Los autores comentan que hay metaanálisis con otras variables comparativas, como el RR o el OR, que muestran estimaciones más dispersas, pero que, sin embargo, no se observan valores tan elevados de I^2 .

Por otro lado, las estimaciones de τ^2 (τ^2) y τ (τ) muestran la varianza de los tamaños de efecto reales y su desviación estándar. Las estimaciones de estos valores entre los diferentes modelos pueden ser diferentes, lo cual sería esperable ya que cada enlace opera sobre una escala distinta. No obstante, todos los modelos GLMMs muestran altos valores de τ (de 0,4185 a 0,7334), lo cual refleja que estos modelos son capaces de capturar la variabilidad intrínseca de los datos, sugiriendo en todos ellos que la variabilidad puede ser elevada.

Por otro lado, el hallazgo más relevante reside en la comparación con el modelo de efectos aleatorios obtenido con la metodología convencional, el cual muestra un τ significativamente menor (0,1507). Esta diferencia sugiere que la transformación

arcoseno empleada en el método convencional infraestima la varianza entre estudios, lo cual es un problema documentado con este tipo de transformaciones (20). Esta hipótesis es consistente con lo observado en la Figura 3.

Por lo tanto, de acuerdo con varios estudios, los GLMMs ofrecen unas ventajas sustanciales respecto a los metaanálisis convencionales, ya que ofrecen una estimación más real de la variabilidad (20,21).

Comparación entre metodología de dos pasos (convencional) o de un paso (GLMMs):

Por otro lado, esta sección tiene como principal objetivo comparar las diferentes metodologías (modelo convencional de efectos aleatorios de dos pasos y el método con un único paso usando GLMMs) y los diferentes tipos de enlace en los GLMMs.

En primer lugar, se ha comparado la estimación del tamaño del efecto obtenida con cada modelo. Por otro lado, se ha hecho una comparativa de las estimaciones de los efectos aleatorios realizadas por los diferentes modelos. Y por último, se ha generado una serie de simulaciones de nuevos datos suponiendo diferentes casuísticas con el fin de poner a prueba los diferentes modelos.

Estimación puntual del tamaño del efecto

Con el fin de distinguir las diferencias entre estimaciones globales de cada uno de los diferentes modelos y si esta diferencia es significativa, se ha calculado la ratio entre las estimaciones realizadas por cada modelo respecto a la estimación del método tomado como referencia (modelo de efectos aleatorios convencional con transformación arcoseno). Como se ha comentado en la Sección de Metodología (Comparación entre los modelos), artículos publicados (22) realizaron una comparativa similar.

En el siguiente gráfico se muestran los resultados obtenidos de la comparación de las estimaciones globales de los GLMMs frente al modelo de referencia:

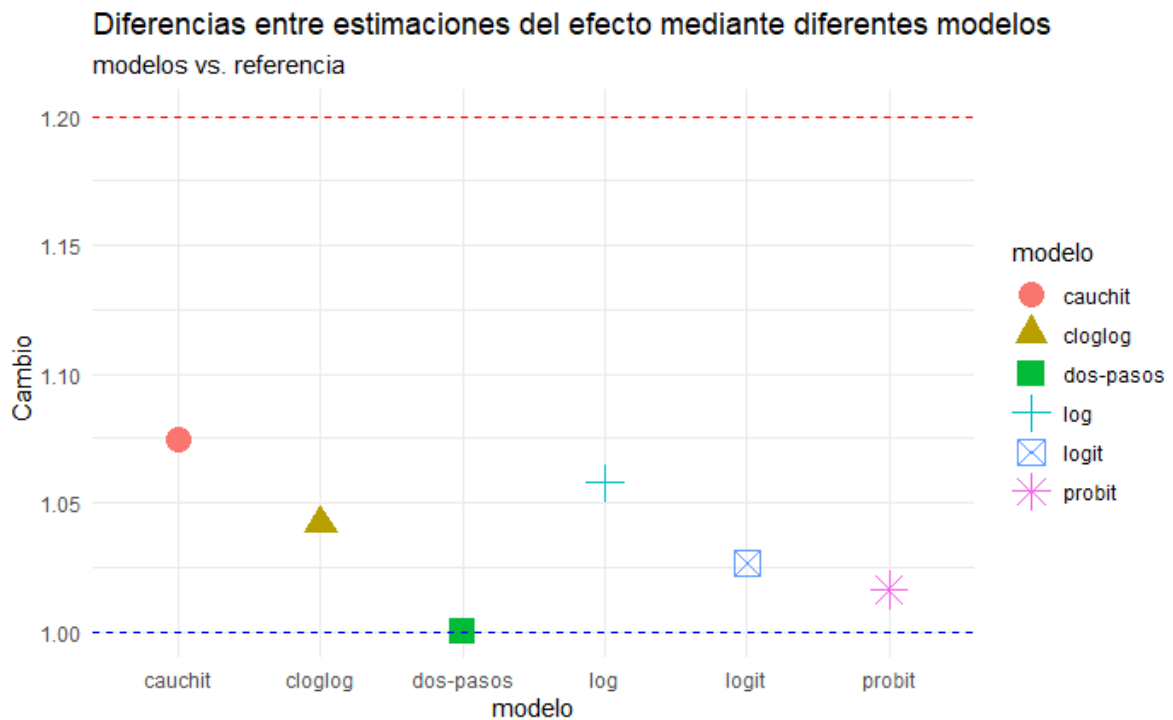
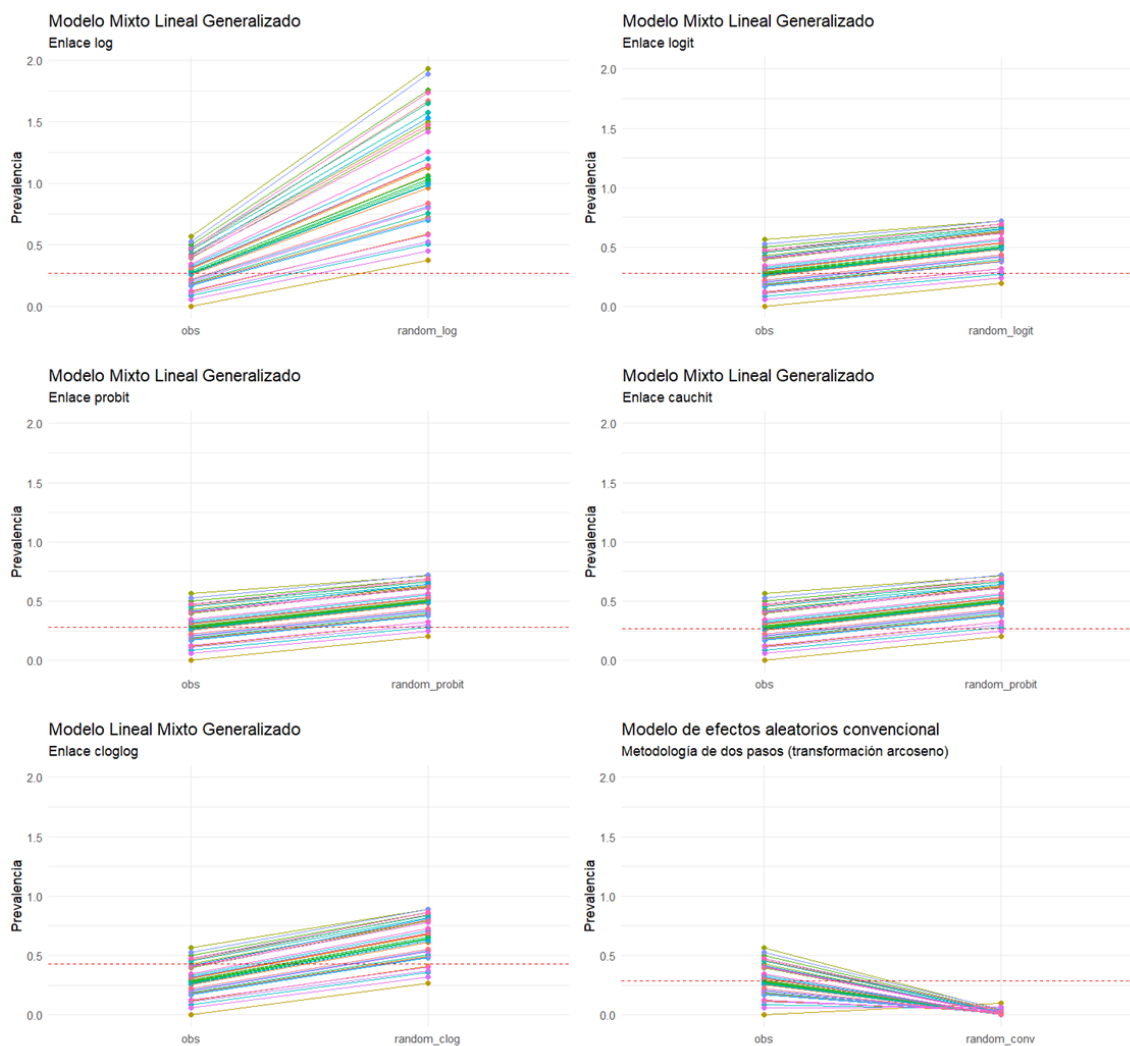


Gráfico 3. Comparación entre las diferentes metodologías a la hora de realizar las estimaciones del efecto agrupado.

Como se puede observar en el Gráfico 3, ninguna de las estimaciones se encuentra por encima del límite de 1,2, el cual se considera como una diferencia considerable (22). El enlace probit sería el que más se asemeja a la estimación obtenida mediante el método de referencia (1,0164), mientras que el enlace cauchit es el que más se diferencia (1,0746). Se puede concluir que la elección del modelo GLMM no altera significativamente la estimación de la prevalencia media de SJS/TEN.

Estimación de los efectos aleatorios

A continuación, se ha tratado de comparar las predicciones de los efectos aleatorios de los diferentes modelos respecto a las estimaciones puntuales de cada estudio. Esta comparación permite evaluar la capacidad de cada modelo para capturar la variabilidad a nivel de grupo y comprender cómo las distintas especificaciones del modelo influyen en la estimación de estas desviaciones individuales.



Nota: Las prevalencias expresadas en tanto por uno, deberían encontrarse entre 0 y 1. Valores por encima de 1 son incorrectos, pese a ello se ha considerado mostrar el eje de ordenadas más amplio para señalar la repercusión de usar el modelo con enlace log

Figura 3. Comparación entre las estimaciones puntuales de prevalencia de cada estudio y las predicciones realizadas por los diferentes modelos.

Como se muestra en la Figura 3, existen diferencias en las predicciones de los diferentes modelos sobre las estimaciones puntuales para cada uno de los estudios. La línea roja discontinua representa la estimación del tamaño del efecto global realizada por cada modelo.

Por lo general, los GLMMs mantienen las distancias entre los diferentes estudios, lo que indicaría que el modelo maneja la variabilidad real observada por los estudios. Simplemente aplican la transformación de la prevalencia, pero la varianza entre estudios se mantiene.

Sin embargo, en el caso del modelo convencional de dos pasos, toda esa variabilidad ha sido reducida, produciendo predicciones cercanas a 0. Esto demuestra visualmente que el método de dos pasos infraestima la variabilidad real porque fuerza las estimaciones de los estudios a parecerse más entre sí. Esto estaría en línea con el valor tan bajo de τ observado con este modelo mostrado en la Tabla 4

Por otro lado, al evaluar las predicciones de los efectos aleatorios realizadas con el modelo usando el enlace log, se puede observar que algunas de ellas superan el valor de 1,0, un resultado matemáticamente imposible para una proporción. Por lo tanto, a pesar de dar una estimación adecuada de la prevalencia global, el enlace log podría no ser adecuado para realizar un metaanálisis de proporciones.

Con el fin de comparar los GLMMs entre sí, se ha obtenido el valor de AIC. Brevemente, este estadístico permite comparar los diferentes modelos entre sí, ofreciendo una métrica de manera que valores más bajos indican mejores modelos. Más información sobre sus principales características y aplicaciones se encuentra descrita en la Sección de Metodología (Comparación entre los modelos).

A continuación, se muestran los valores de AIC para los diferentes modelos.

Enlace	AIC
Probit	279,32
Logit	280,06
Cloglog	281,73
Log	284,01
Cauchit	287,03
Modelo convencional	-30,98

Tabla 5. Comparación de las estimaciones de AIC de los diferentes modelos.

Como se muestra en la Tabla 5, el modelo de efectos aleatorios convencional es el que ha dado un valor de AIC más bajo, incluso negativo (-30,98). Esta gran diferencia respecto al resto de modelos estaría en línea con la infraestimación de la variabilidad comentada previamente, sugiriendo un sobreajuste de los datos. Por lo tanto, este valor no debería ser comparable con el del resto de modelos.

En el caso de los GLMMs, el modelo con enlace probit fue el que dio el valor de AIC más bajo (279,3212), indicando el mejor equilibrio entre bondad de ajuste y complejidad.

Los modelos logit y cloglog también mostraron un ajuste excelente. Por el contrario, los enlaces log y cauchit exhibieron los AIC más altos.

De acuerdo con lo especificado por Burnham y Anderson (2004) (38), una diferencia de AIC (ΔAIC) mayor de 5 entre modelos respecto al que presentó el AIC más bajo puede ser considerada sustancial. Al aplicar este criterio, se corrobora que los enlaces cauchit ($\Delta AIC=7,71$) y log ($\Delta AIC=4,69$) tienen un ajuste significativamente inferior en comparación con el modelo con el enlace probit (AIC=279,32). Sin embargo, dado que la diferencia entre los enlaces probit y logit (AIC=280,06) es mínima ($\Delta AIC=0,74$), ambos modelos se consideran igualmente plausibles.

En resumen, todos los modelos arrojaron estimaciones consistentes para la prevalencia de SJS/TEN. Sin embargo, el análisis detallado de la variabilidad y el ajuste reveló diferencias cruciales entre modelos. En primer lugar, se observó que el modelo de efectos aleatorios ajustado mediante método convencional con transformación arco seno infraestima significativamente la variabilidad entre estudios (τ^2). Por otro lado, las estimaciones de los efectos aleatorios realizadas por el enlace log no fueron correctas, ya que mostró predicciones de prevalencia superiores a 1. Por último, el enlace probit y logit parece que fueron capaces de capturar la variabilidad real mostrada por los estudios, además, mostraron el mejor ajuste según el criterio AIC.

Evaluación robustez de las metodologías

Tal y como se ha descrito en la sección de Metodología (Comparación entre los modelos), se muestran los resultados de los diferentes supuestos utilizados para poner a prueba los diferentes modelos:

Diferentes tamaños de muestra

De acuerdo a lo que nos comenta Borenstein *et al.* (2009) (5) en su libro, el tamaño de la muestra de los estudios individuales es un factor crítico en la estimación del efecto global. La necesidad de otorgar un mayor peso a los estudios con un mayor número de sujetos se debe a que sus estimaciones son más precisas (es decir, tienen menor variabilidad). Si bien el uso de modelos de efectos aleatorios modera ligeramente esta asignación, los estudios más grandes conservan una influencia considerable en el efecto resumen.

Al evaluar las estimaciones de prevalencia en función del tamaño muestral, se observaron diferencias notables entre los subgrupos. La siguiente tabla presenta los resultados para los diferentes enlaces de función:

Tamaño muestral	Modelos					
	Log	Logit	Probit	Cauchit	Cloglog	Convencional
Bajo	0.3364	0.3462	0.3479	0.3342	0.3418	0.3497
Medio	0.1951	0.1968	0.1990	0.1962	0.1957	0.2030
Alto	0.2855	0.2926	0.2948	0.2789	0.2891	0.2977

Tabla 6. Análisis de prevalencias en función del tamaño de muestra

La prevalencia estimada en los estudios con un tamaño muestral medio es considerablemente menor que en los grupos de tamaño bajo y alto. La diferencia entre los estudios medianos y grandes roza la significancia estadística con valores de p muy cercanos a 0,05. Sin embargo, no se llega a alcanzar, mientras que la diferencia entre los estudios pequeños y grandes está totalmente clara que no lo es ($p \approx 0,33$).

La observación de que la prevalencia estimada no sigue una tendencia lineal con el tamaño muestral resalta la complejidad de la heterogeneidad de los datos. Este hallazgo puede deberse a diferencias sistemáticas no detectadas entre los estudios, como el tipo de población, la metodología empleada o el periodo de tiempo en que fueron realizados.

Presencia de valores atípicos

Un valor atípico, en el contexto de un metaanálisis, es un estudio con un tamaño de efecto que difiere sustancialmente de los demás. La presencia de un estudio atípico puede comprometer la validez de la estimación global. Estos estudios pueden tener una influencia desproporcionada en el resultado final (5).

La identificación de valores atípicos es crucial para el análisis de sensibilidad, donde se evalúa si las conclusiones del metaanálisis son robustas y no dependen de la presencia de uno o varios estudios con resultados aberrantes.

El problema de convergencia observado con el enlace cauchit (Tabla 3), subraya la importancia de evaluar la influencia de los estudios individuales en el metaanálisis. Por lo tanto, se ha realizado un análisis de sensibilidad más exhaustivo para evaluar el impacto de otros posibles valores atípicos.

Los resultados obtenidos con los estudios de alta prevalencia son los siguientes:

Enlace	Prevalencia	IC 95%
Log	0,2930809	0,2416133 0,3555118
Logit	0,3375818	0,2493744 0,4387523
Probit	0,3405638	0,2558124 0,4342596
Cauchit	0,3068325	0,3056505 0,3080208
clog-log	0,3126674	0,2451151 0,3934321
Convencional	0,3436778	0,2617188 0,4306189

Tabla 7. Estimaciones del tamaño de efecto global para cada uno de los modelos en presencia de valores atípicos muy elevados.

Como era esperable, la estimación global de la prevalencia ha aumentado. Aquí lo interesante es que el enlace cauchit vuelve a dar problemas de convergencia cuando se incluyen valores muy cercanos a 1. El resto de enlaces parece que dan estimaciones similares y no muestran problemas a la hora de ejecutar el modelo, siendo los mejores los modelos con los enlaces probit y logit.

En cuanto a los valores de prevalencias muy pequeños, estos son los resultados:

Enlace	Prevalencia	IC 95%
Log	0,2100033	0,1586115 0,2780465
Logit	0,2252016	0,1691804 0,2932262
Probit	0,2343517	0,1789072 0,2981798
Cauchit	0,1901745	0,1896117 0,1907403
clog-log	0,2176503	0,1637882 0,2859597
Convencional	0,2503112	0,1906412 0,3151663

Tabla 8. Estimaciones del tamaño de efecto global para cada uno de los modelos en presencia de valores atípicos muy bajos.

Al igual que cuando incluimos la alta prevalencia, parece que solo el enlace cauchit tiene problemas de convergencia. El resto de modelos dan resultados válidos.

En ninguno de los dos supuestos la adición de valores tan extremos produce resultados muy dispares respecto a las estimaciones de los datos iniciales (0,26 - 0,28). Por lo tanto, se podría concluir que las metodologías usadas para el metaanálisis, tanto el modelo convencional como los GLMMs, son herramientas robustas que permiten llevar a cabo una estimación precisa a pesar de la presencia de valores muy extremos.

Heterogeneidad

Uno de los principales problemas de los metaanálisis es ser capaces de poder determinar la alta variabilidad en los efectos reales, o heterogeneidad, que hay entre los estudios. Este problema radica en la capacidad de interpretar los resultados, ya que esta puede influir en las conclusiones obtenidas en el metaanálisis (24).

En el caso de los metaanálisis de proporciones la evaluación de la heterogeneidad es incluso algo más compleja. Como se ha comentado anteriormente, el estadístico I^2

(ampliamente reconocido por la comunidad científica para cuantificar la heterogeneidad) se puede ver afectado por varias características implícitas en el estudio de las prevalencias (estudios observacionales con elevados tamaños muestrales y la propia naturaleza no comparativa de la variable) (37).

Como se ha comentado en la Sección de Metodología (Comparación entre los modelos), para evaluar el impacto de la heterogeneidad en las estimaciones de los GLMMs se han simulado dos conjuntos de datos con el mismo número de estudios y el mismo valor medio de prevalencia, pero con variabilidades muy diferentes. En cada uno de ellos se ha realizado la comparación en cuanto a la estimación global del tamaño del efecto y la heterogeneidad como se ha hecho anteriormente.

En primer lugar, en el caso de baja heterogeneidad, estos son los datos simulados:

Estudio	Tamaño de la Muestra (n)	Número de Casos	Prevalencia Estimada
1	148	44	0,2973
2	420	133	0,3167
3	252	62	0,246
4	173	44	0,2543
5	327	68	0,208
6	245	67	0,2735
7	221	53	0,2398
8	148	41	0,277
9	227	64	0,2819
10	402	117	0,291
11	123	30	0,2439
12	426	117	0,2746
13	455	95	0,2088
14	188	44	0,234
15	264	62	0,2348
16	209	39	0,1866
17	119	31	0,2605
18	469	125	0,2665
19	466	116	0,2489
20	486	137	0,2819
21	396	94	0,2374
22	188	41	0,2181
23	382	104	0,2723
24	208	49	0,2356
25	104	32	0,3077
26	311	77	0,2476

Estudio	Tamaño de la Muestra (n)	Número de Casos	Prevalencia Estimada
27	447	125	0,2796
28	459	124	0,2702
29	358	85	0,2374
30	413	127	0,3075
31	397	111	0,2796
32	123	32	0,2602
33	257	69	0,2685
34	398	112	0,2814
35	498	131	0,2631
36	413	70	0,1695
37	235	63	0,2681
38	391	97	0,2481

Tabla 9. Datos simulados en el supuesto de baja heterogeneidad

Las estimaciones de los tamaños de los efectos se muestran en la Tabla 10, así como el forestplot resultante en el Gráfico 4.

Enlace	Prevalencia	IC 95%
Log	0,2567	0,2462-0,2677
Logit	0,2568	0,2462-0,2678
Probit	0,2569	0,2462-0,2678
Cauchit	0,2565	0,2462-0,2676
clog-log	0,2568	0,2462-0,2677
Convencional	0,2567	0,2459-0,2677

Tabla 10. Estimaciones globales por los diferentes modelos en el supuesto de baja heterogeneidad

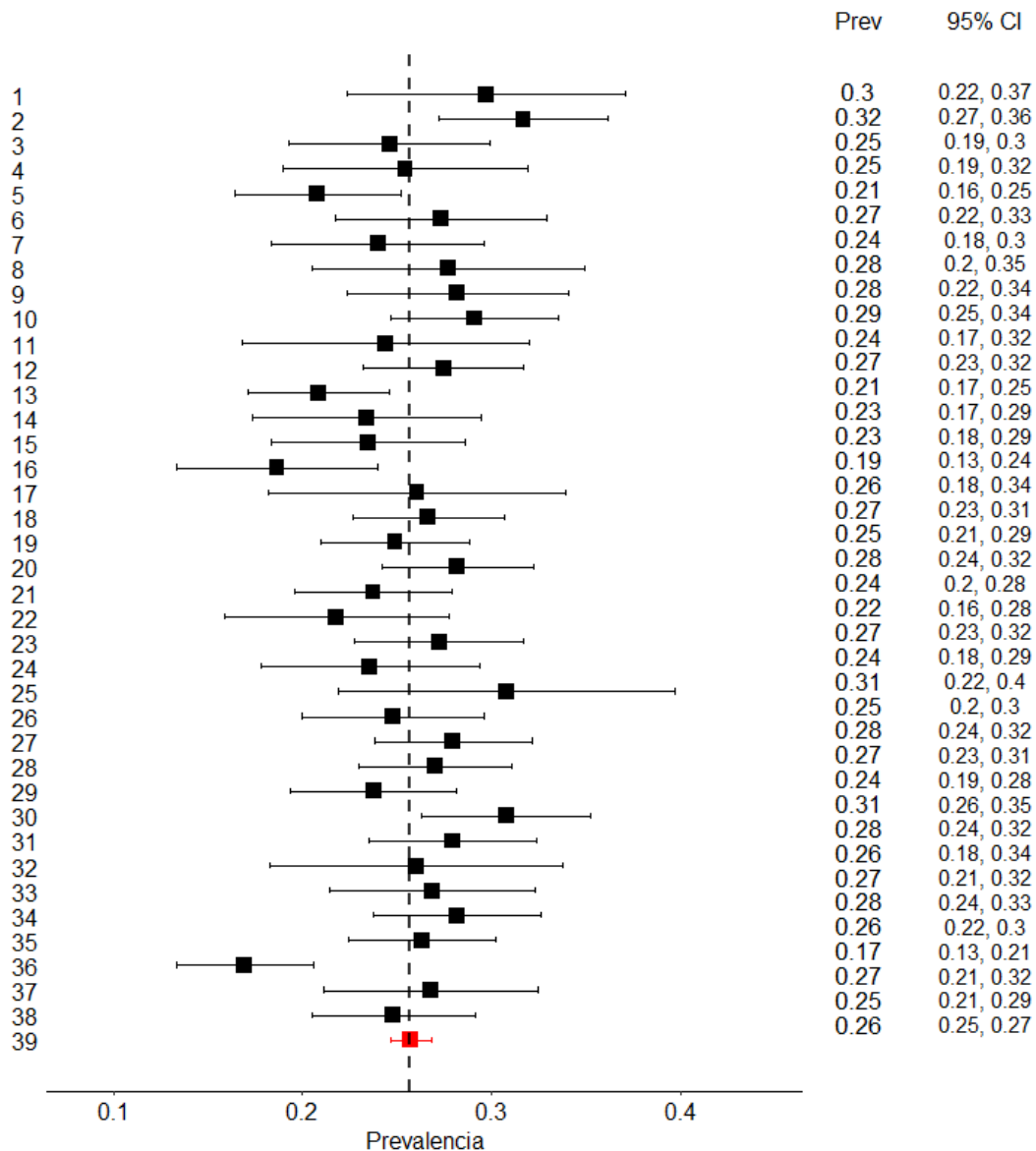


Gráfico 4. Forestplot de los 38 estudios simulados y la estimación global del efecto (39) en el supuesto de baja heterogeneidad. Estimación realizada con GLMMs y el enlace probit

Las estimaciones de prevalencia para todos los modelos (Tabla 10), tanto GLMMs como el método convencional, se concentran alrededor de 0,257, que corresponde a la media de la distribución que se ha usado para poder hacer la simulación. Además, los ICs son muy estrechos y están superpuestos unos a otros. Este resultado valida la robustez de todos los modelos cuando la variabilidad entre estudios es baja.

El forestplot (Gráfico 4) estaría mostrando que en 34 de los 38 estudios simulados (89,5%) se muestra una prevalencia entre 0,2 y 0,3. Además, excepto 4 estudios, todos los ICs 95% estimados para cada uno de los estudios individuales contienen la estimación global obtenida con el modelo. Este primer análisis gráfico estaría en línea con el supuesto inicial de baja heterogeneidad.

A continuación, se presentan los estadísticos usados para evaluar la heterogeneidad de manera analítica:

Enlace	Q	Q test	I^2 (%)	I^2 (IC 95%)	τ^2	τ
Log	61,7723	0,0064	40,10	12,24-59,12	0,0070	0,0841
Logit	64,3056	0,0035	42,46	15,91-60,63	0,0131	0,1146
Probit	66,1674	0,0022	44,08	18,44-61,66	0,0046	0,0682
Cauchit	430,866	$1,35 \times 10^{-68}$	91,41	89,19-93,18	0,0161	0,1269
clog-log	63,0227	0,0048	41,29	14,09-59,88	0,0097	0,0986
Convencional*	67,2914	0,0017	45,02	19,90-62,26	0,0007	0,0258

*Modelo de efectos aleatorios convencional, datos transformados con la función arcoseno.

Tabla 11. Análisis de la heterogeneidad en el supuesto de baja heterogeneidad.

En primer lugar, se puede observar cómo el enlace cauchit proporciona valores de Q como de I^2 muy por encima de lo estimado para el resto de modelos y de enlaces. Esto puede deberse a que este tipo de escala produce unas estimaciones de la variabilidad mucho más elevadas de las observadas realmente en los datos. Por lo que podría no ser adecuado considerar este tipo de enlace para casos con poca heterogeneidad.

Respecto al resto de enlaces y metodologías, se puede apreciar cómo el estadístico Q reporta valores alrededor de 60, por lo que al ser mayor que los grados de libertad ($df = 37$), da valores de p muy bajos. En este caso se puede rechazar la hipótesis nula de homogeneidad, lo que significa que a pesar de la baja heterogeneidad existiría una diferencia significativa entre los estudios.

Por otro lado, en cuanto a I^2 , las estimaciones están en torno al 40%, lo que se clasificaría comúnmente dentro del rango de heterogeneidad moderada. Sin embargo, es crucial destacar que la estimación de este estadístico no es precisa, tal y como se puede apreciar en los amplios ICs 95% calculados. La interpretación de estos intervalos sugiere que la verdadera magnitud de la heterogeneidad podría variar desde un nivel bajo hasta un nivel moderado-alto, lo que indica que el I^2 no es un estimador preciso en este caso.

En cuanto a la estimación directa de la variabilidad (τ y τ^2) se puede observar que por lo general los GLMMs hacen una estimación similar de la variabilidad (en torno a 0,1). Curiosamente, como ocurría con los datos reales, la transformación arcoseno da una estimación mucho menor de la variabilidad que la observada con los modelos. Se podría concluir que los GLMMs siguen siendo mejores para capturar la variabilidad real entre estudios, incluso cuando esta es baja.

Por último, para comparar los diferentes modelos entre sí, se ha vuelto a recurrir al AIC. Los valores reportados se muestran a continuación:

Enlace	AIC
Probit	281,86
Logit	281,98
Cloglog	282,10
Log	282,25
Cauchit	282,68
Modelo convencional	-136,15

Tabla 12. Valores de AIC por tipo de enlace. Supuesto de baja heterogeneidad

Como se puede observar en la Tabla 12, prácticamente todos los modelos presentan un ajuste similar, a excepción del modelo de efectos aleatorios convencional, cuyo valor de AIC tan bajo evidencia un sobreajuste de los datos con este tipo de transformación.

Una vez estudiado el caso de una baja heterogeneidad, se va a realizar el mismo análisis, pero ahora en una muestra simulada con una elevada variabilidad entre estudios, es decir, una alta heterogeneidad.

Como en el caso anterior, primero se muestran los datos simulados. Cabe destacar que en este caso se ha tratado de usar una variabilidad de valores diferentes entre sí pero que entren dentro de valores posibles para proporciones, sin que sean valores muy extremos.

Estudio	Tamaño de la Muestra (n)	Número de Casos	Prevalencia Estimada
1	148	101	0,6824
2	420	388	0,9238
3	252	28	0,1111
4	173	29	0,1676
5	327	117	0,3578
6	245	103	0,4204
7	221	8	0,0362
8	148	62	0,4189
9	227	115	0,5066
10	402	250	0,6219
11	123	6	0,0488
12	426	186	0,4366
13	455	155	0,3407
14	188	3	0,0160
15	264	10	0,0379
16	209	122	0,5837
17	119	32	0,2689
18	469	156	0,3326
19	466	62	0,1330
20	486	255	0,5247
21	396	2	0,0051
22	188	41	0,2181

Estudio	Tamaño de la Muestra (n)	Número de Casos	Prevalencia Estimada
23	382	157	0,4110
24	208	5	0,0240
25	104	80	0,7692
26	311	34	0,1093
27	447	219	0,4899
28	459	171	0,3725
29	358	5	0,0140
30	413	335	0,8111
31	397	193	0,4861
32	123	36	0,2927
33	257	92	0,3580
34	398	198	0,4975
35	498	145	0,2912
36	413	325	0,7869
37	235	85	0,3617
38	391	51	0,1304

Tabla 13. Datos simulados en el supuesto de alta heterogeneidad

Las estimaciones de los tamaños de los efectos se muestran en la

Enlace	Prevalencia	IC 95%
Log	0,2264	0,1549-0,3308
Logit	0,27720	0,1846-0,3937
Probit	0,29677	0,2054-0,4031
Cauchit	0,15736	0,1004-0,3168
clog-log	0,25021	0,1690-0,3610
Convencional	0,32387	0,2371-0,4171

Tabla 14. Estimaciones globales por los diferentes modelos en el supuesto de alta heterogeneidad

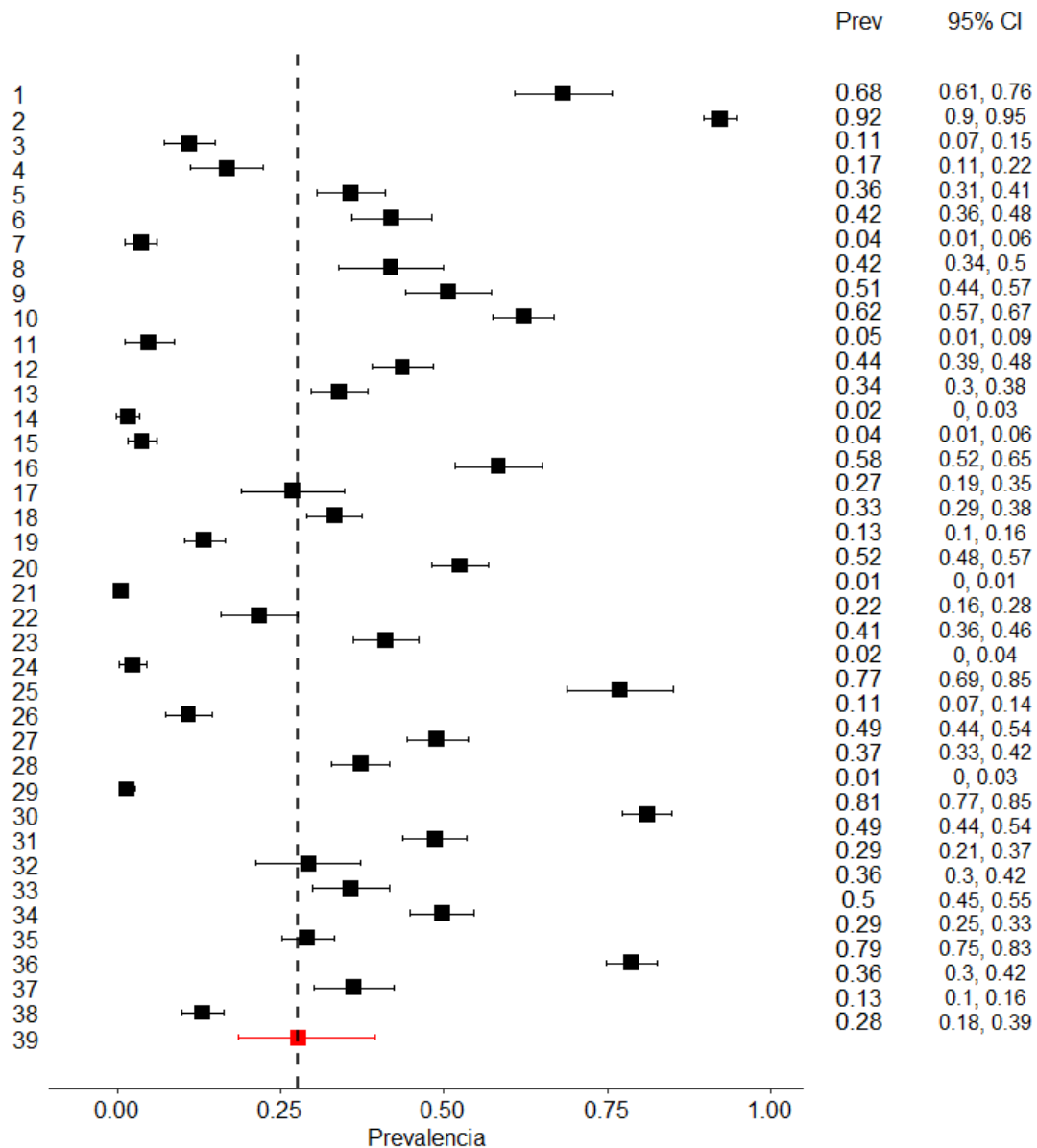


Gráfico 5. Forestplot de los 38 estudios simulados y la estimación global del efecto (39) en el supuesto de alta heterogeneidad. Estimación realizada con GLMMs y el enlace logit

Curiosamente, en este caso sí es posible observar diferencias entre los diferentes enlaces y metodologías. Como anteriormente, la media de la distribución fue fijada en 0,26. En el caso de alta heterogeneidad solo los enlaces logit, probit y cloglog mostraron los resultados más cercanos a este valor (0,27720; 0,29677 y 0,2502, respectivamente). El resto se mantuvo muy por debajo o por encima de esta media.

Para poder mostrar esta diferencia de manera gráfica, se ha empleado el mismo procedimiento que para el Gráfico 3. A continuación, se muestra el resultado:

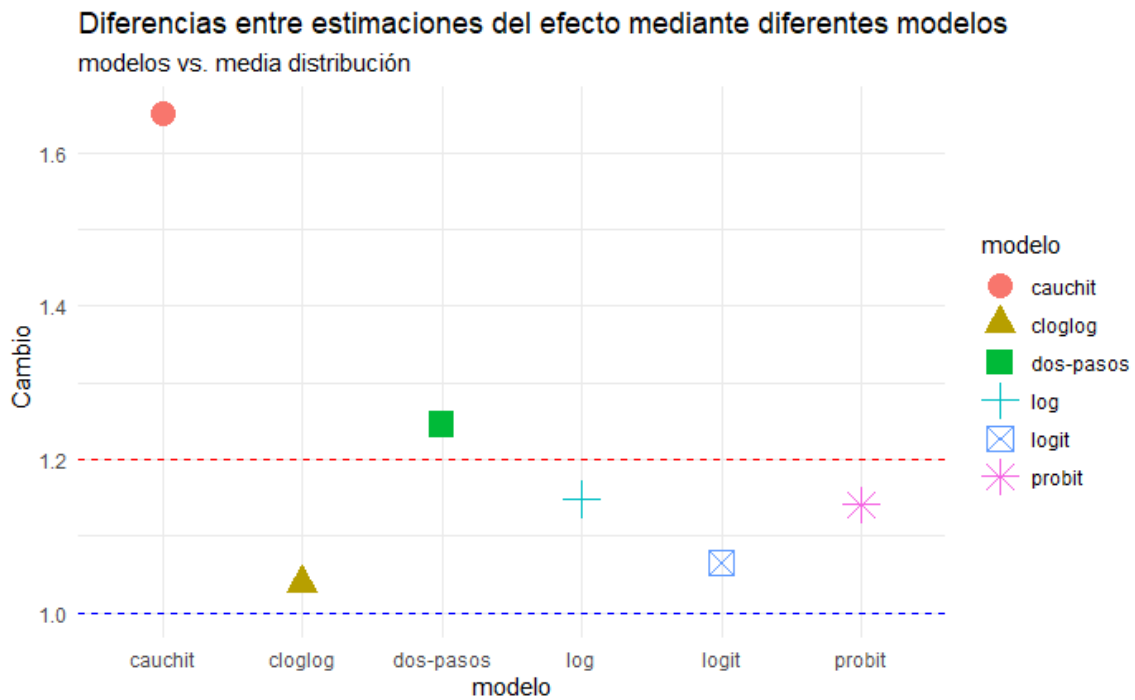


Gráfico 6. Comparación entre las estimaciones globales del efecto realizadas por los diferentes modelos cuando hay una elevada heterogeneidad

Como se puede observar en el Gráfico 6, los enlaces logit y cloglog fueron los que más se parecían a la media usada para poder hacer la simulación. Luego encontraríamos los enlaces probit y log dentro del rango de 1,0 a 1,2, lo que indicaría que las diferencias no serían significativas. Sin embargo, en los casos del enlace cauchit y el método convencional, sus estimaciones de cambio superan el límite de 1,2 (1,652 y 1,245, respectivamente), lo cual indicaría que las diferencias observadas con estos modelos sí son significativas.

Respecto a los ICs al 95%, se puede observar que estos son más amplios que en el supuesto de baja heterogeneidad, lo cual era esperable. Sin embargo, muchos de ellos presentan bastantes regiones en común.

A diferencia del caso de baja heterogeneidad, el forestplot (Gráfico 5) estaría mostrando que solo 4 de los 38 estudios simulados (10,5%) contienen la estimación global obtenida con el modelo en su IC al 95%. Este primer análisis gráfico estaría en línea con el supuesto inicial de alta heterogeneidad.

A continuación, se presentan los estadísticos usados para evaluar la heterogeneidad de manera analítica:

Enlace	Q	Q test	I^2 (%)	I^2 (IC 95%)	τ^2	τ
Log	6712,992	0	99,45	99,39-99,50	1,386	1,177
Logit	1858,921	0	98,01	97,70-98,28	2,694	1,641
Probit	2339,844	0	98,42	98,19-98,62	0,810	0,900
Cauchit	51770,86	0	99,93	99,93-99,93	13,087	3,617
clog-log	2515,957	0	98,53	98,32-98,71	1,887	1,3738
Convencional	4208,554	<0,0001	99,12	99,02-99,21	0,0916	0,3026f

Tabla 15. Análisis de la heterogeneidad en el supuesto de alta heterogeneidad.

En primer lugar, se puede observar cómo de nuevo el enlace cauchit proporciona valores de Q y variabilidad (τ^2 y τ) muy por encima de lo estimado para el resto de modelos y de enlaces.

Respecto al resto de enlaces y metodologías, se puede apreciar cómo el estadístico Q es bastante dispar, pero todos son muy elevados que permiten rechazar la hipótesis nula de homogeneidad.

Por otro lado, en cuanto a I^2 , las estimaciones están en torno al 98-99%, lo que se clasificaría comúnmente dentro del rango de heterogeneidad alta. Es crucial destacar que los ICs 95% calculados son muy estrechos y todos ellos por encima de 97%, lo cual permite concluir con un 95% de seguridad que la heterogeneidad real de los datos es muy elevada.

En cuanto a la estimación directa de la variabilidad (τ y τ^2) se puede observar que se ha venido comentando hasta el momento, la transformación arcoseno da una estimación mucho menor de la variabilidad que la observada con los modelos.

Por último, para comparar los diferentes modelos entre sí, se ha vuelto a recurrir al AIC. Los valores reportados se muestran a continuación:

Enlace	AIC
Probit	421,55
Logit	423,35
Cloglog	429,58
Log	441,44
Cauchit	469,11
Modelo convencional	20,59

Tabla 16. Valores de AIC de cada uno de los modelos ajustados. Supuesto de alta heterogeneidad

Los resultados que se muestran en la Tabla 16 muestran que al igual que en los casos anteriores, el modelo convencional de efectos aleatorios presenta el valor de AIC más

bajo. Sin embargo, este puede deberse al sobreajuste debido a la transformación de los datos.

Respecto a los GLMMs, los enlaces probit y logit son los que mejor relación entre ajuste y bondad a los datos presentan respecto al resto de modelos. Es decir, estos enlaces parecen ser los más versátiles a la hora de modelar los datos de proporciones. Por otro lado, como se ha venido anticipando, los enlaces log y cauchit son los que peor funcionan.

5. CONCLUSIONES

El presente TFM ha tratado de evaluar la idoneidad y la robustez de los GLMMs como una metodología para el metaanálisis de proporciones, utilizando como caso de estudio la prevalencia de SJS y TEN asociada al consumo de antibióticos. Los resultados obtenidos respaldan la idea de que los GLMMs ofrecen ventajas sustanciales sobre el método convencional de dos pasos.

En primer lugar, a pesar de la heterogeneidad inherente a los datos, los GLMMs, independientemente del enlace de función utilizado, proporcionaron estimaciones del efecto agregado que fueron muy similares entre sí y altamente consistentes con el resultado del análisis convencional y el del estudio original de Lee *et al.* (2023) (15). Esto demuestra que la elección del enlace de función en el metaanálisis de prevalencias tiene un impacto mínimo en la estimación del efecto global.

Sin embargo, el análisis detallado de la heterogeneidad reveló diferencias cruciales. Se concluyó que el método de dos pasos con la transformación arcoseno infraestima significativamente la varianza real entre estudios (τ^2), lo que genera una representación irreal de la variabilidad inherente a los datos. Por el contrario, los GLMMs demostraron ser superiores al capturar de forma más precisa y consistente la verdadera heterogeneidad, proporcionando una estimación más fiable de la variabilidad.

La comparación entre los diferentes modelos basada en el AIC permitió identificar los modelos con el mejor equilibrio entre bondad de ajuste y complejidad. En línea con lo observado en la evaluación de la heterogeneidad, parece que la transformación arcoseno sobreajusta los datos, haciendo que los valores de AIC obtenidos no sean compatibles con los del resto de modelos. Los enlaces probit y logit mostraron el mejor ajuste tanto en el escenario de los datos reales como en las simulaciones con alta heterogeneidad, confirmando su versatilidad y fiabilidad para modelar datos de proporciones. Por otro lado, los enlaces log y cauchit se mostraron menos adecuados. El enlace log generó predicciones de prevalencia por encima de 1, mientras que el enlace cauchit presentó problemas de convergencia y una menor capacidad de ajuste, especialmente con valores atípicos.

En definitiva, este TFM demuestra que los GLMMs son una alternativa válida y muy potente para el metaanálisis de proporciones. Ofreciendo una metodología de un solo paso que proporciona estimaciones robustas del efecto global y una representación más precisa

de la heterogeneidad. En particular, se pueden destacar los enlaces probit y logit, los que mejores resultados han demostrado, en cuanto a estimación y adaptación a las diferentes situaciones que se les han dado en este TFM.

No obstante, los hallazgos abren la puerta a varias líneas de investigación futuras que podrían expandir y profundizar en la metodología aquí presentada.

1. Desarrollo de herramientas accesibles para la investigación

A pesar de sus ventajas metodológicas, la complejidad percibida de los GLMMs limita su uso en la práctica clínica y epidemiológica. Una vía de investigación prometedora sería el desarrollo de interfaces gráficas o funciones simplificadas en R que hagan más accesible la implementación de los GLMMs para metaanálisis de proporciones, eliminando barreras técnicas y promoviendo su adopción entre investigadores con menos experiencia en programación.

2. Comparación con otros métodos recientes

Investigaciones recientes han propuesto nuevas transformaciones y enfoques bayesianos para el metaanálisis de proporciones (20). Un futuro estudio podría comparar los GLMMs con estos métodos emergentes para determinar cuál ofrece la mejor combinación de precisión, robustez e interpretabilidad en distintos escenarios de datos.

3. Metarregresión y heterogeneidad por covariables

El metaanálisis de prevalencias a menudo presenta una heterogeneidad considerable debido a factores no contemplados. Un área de investigación futura podría ser la aplicación de la metarregresión en el contexto de los GLMMs para identificar y cuantificar la contribución de covariables como la edad, la ubicación geográfica o el tipo de población a la heterogeneidad observada. Esto no solo mejoraría la interpretación de los resultados, sino que también permitiría desarrollar modelos más precisos.

6. BIBLIOGRAFÍA

1. McNamara DJ. The Fifty Year Rehabilitation of the Egg. *Nutrients*. 21 de octubre de 2015;7(10):8716-22.
2. Drouin-Chartier JP, Chen S, Li Y, Schwab AL, Stampfer MJ, Sacks FM, et al. Egg consumption and risk of cardiovascular disease: three large prospective US cohort studies, systematic review, and updated meta-analysis. *BMJ*. 4 de marzo de 2020;368:m513.
3. Carson JAS, Lichtenstein AH, Anderson CAM, Appel LJ, Kris-Etherton PM, Meyer KA, et al. Dietary Cholesterol and Cardiovascular Risk: A Science Advisory From the American Heart Association. *Circulation*. 21 de enero de 2020;141(3):e39-53.
4. Mah E, Chen CYO, Liska DJ. The effect of egg consumption on cardiometabolic health outcomes: an umbrella review. *Public Health Nutr*. 2020;23(5):935-55.
5. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to meta-analysis*. John Wiley & Sons; 2009.
6. Vera Carrasco O. El metaanálisis en la investigación científica médica. *Rev Médica Paz*. 2023;29(2):80-5.
7. Sos VJE, Abella JAL, Villach LG, Oliver MB. Metaanálisis: una forma básica de entender e interpretar su evidencia. *Rev Senol Patol Mamar*. 2021;34(1):44-51.
8. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. julio de 2011;128(1):305-10.
9. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 13 de enero de 1996;312(7023):71-2.
10. Lin L, Chu H. Meta-analysis of Proportions Using Generalized Linear Mixed Models. *Epidemiol Camb Mass*. septiembre de 2020;31(5):713-7.
11. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med*. 30 de diciembre de 2000;19(24):3417-32.
12. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 20 de diciembre de 2010;29(29):3046-67.
13. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *J Clin Epidemiol*. enero de 2008;61(1):41-51.
14. Bakbergenuly I, Kulinskaya E. Meta-analysis of binary outcomes via generalized linear mixed models: a simulation study. *BMC Med Res Methodol*. 4 de julio de 2018;18(1):70.

15. Lee EY, Knox C, Phillips EJ. Worldwide Prevalence of Antibiotic-Associated Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis: A Systematic Review and Meta-analysis. *JAMA Dermatol.* 1 de abril de 2023;159(4):384-92.
16. Frantz R, Huang S, Are A, Motaparathi K. Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis: A Review of Diagnosis and Management. *Med Kaunas Lith.* 28 de agosto de 2021;57(9).
17. Yun J, Cai F, Lee FJ, Pichler WJ. T-cell-mediated drug hypersensitivity: immune mechanisms and their clinical relevance. *Asia Pac Allergy.* abril de 2016;6(2):77-89.
18. Klein EY, Impalli I, Poleon S, Denoel P, Cipriano M, Van Boeckel TP, et al. Global trends in antibiotic consumption during 2016-2023 and future projections through 2030. *Proc Natl Acad Sci U S A.* 3 de diciembre de 2024;121(49):e2411919121.
19. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Pitfalls of using the risk ratio in meta-analysis. *Res Synth Methods.* septiembre de 2019;10(3):398-419.
20. Lin L, Xu C. Arcsine-based transformations for meta-analysis of proportions: Pros, cons, and alternatives. *Health Sci Rep.* septiembre de 2020;3(3):e178.
21. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods.* septiembre de 2019;10(3):476-83.
22. Lin L, Xu C, Chu H. Empirical Comparisons of 12 Meta-analysis Methods for Synthesizing Proportions of Binary Outcomes. *J Gen Intern Med.* febrero de 2022;37(2):308-17.
23. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health.* 2019;(22):153-60.
24. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 15 de junio de 2002;21(11):1539-58.
25. Casella G, Berger R. *Statistical inference.* Chapman and Hall/CRC; 2002.
26. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 6 de septiembre de 2003;327(7414):557-60.
27. Akaike H. A New Look at the Statistical Model Identification. En: Parzen E, Tanabe K, Kitagawa G, editores. *Selected Papers of Hirotugu Akaike* [Internet]. New York, NY: Springer New York; 1998. p. 215-22. Disponible en: https://doi.org/10.1007/978-1-4612-1694-0_16
28. Posit. *RStudio Integrated Development Environment - User Guide*, [Internet]. 2025. Disponible en: <https://docs.posit.co/ide/user/>
29. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4(43):1686.

30. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw.* 7 de octubre de 2015;67(1):1-48.
31. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol.* marzo de 2009;24(3):127-35.
32. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw.* 5 de agosto de 2010;36(3):1-48.
33. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* [Internet]. Cambridge: Cambridge University Press; 2006. (Analytical Methods for Social Research). Disponible en: <https://www.cambridge.org/core/product/32A29531C7FD730C3A68951A17C9D983>
34. McCullagh P. *Generalized linear models.* Routledge; 2019.
35. Agresti A, Kateri M. Categorical data analysis. En: *International encyclopedia of statistical science.* Springer; 2025. p. 408-11.
36. Koenker R, Bassett G. Regression Quantiles. *Econometrica.* 1978;46(1):33-50.
37. Migliavaca CB, Stein C, Colpani V, Barker TH, Ziegelmann PK, Munn Z, et al. Meta-analysis of prevalence: I(2) statistic and how to deal with heterogeneity. *Res Synth Methods.* mayo de 2022;13(3):363-7.
38. Burnham KP, Anderson DR. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol Methods Res.* 1 de noviembre de 2004;33(2):261-304.

ANEXO 1 – BASE DE DATOS

ID	study	country	period	population	N	SJS	SJS_TEN	TEN	Setting	Age	Female	Death	beta	fq	sulfa	macro	tetra	vanco	other_abx	total	prev	continente	n	
1	De Bustros 2021	US	2000 to 2019	Adults	131	44	63	24	Inpatient	44.9+/-23.0	78	0	15	10	25	0	0	0	0	10	60	0.45801527	North_America	60
2	Gleghorn 2021	US	1990 to 2015	Children	51	0	3	48	Inpatient	8 (0.25-17)	16	7	1	0	14	0	1	0	0	16	0.31372549	North_America	16	
3	Olteanu 2021	Canada	1999 to 2015	Adults	43	0	7	36	Inpatient	54+/-19	26	9	7	0	3	0	0	0	1	11	0.25581395	North_America	11	
4	Singh 2021	India	Jan 2015 to Dec 2019	Both	75	48	18	9	Inpatient	NR	42	0	10	2	7	1	1	0	2	23	0.30666667	Asia_south	23	
5	Thakur 2021	India	Jan 2014 to Dec 2018	Unclear	51	7	16	28	Inpatient	38.2+/-17.6	24	11	6	0	0	0	0	0	0	6	0.11764706	Asia_south	6	
6	Abou Taleb 2020	Egypt	Jan 2012-Dec 2017	Adults	83	56	22	5	Inpatient	37.9+/-14.5	39	18	10	2	0	3	0	0	0	15	0.18072289	Africa	15	
7	Huyen 2020	Vietnam	Jul 2018 to Jul 2019	Adults	60	30	0	30	Inpatient	52+/-16	28	0	0	0	0	0	0	0	0	2	0.00000000	Asia_southeast	2	
8	Yang 2020	China	Jan 2008- Dec 2018	Adults	213	141	19	53	Inpatient	48.2+/-18.4	98	8	28	6	14	4	1	0	8	61	0.28638498	Asia_east	61	
9	Carrasquillo 2019	Puerto Rico	2006-2017	Both	30	3	10	17	Inpatient	35.9+/-21.7	17	4	6	3	7	1	0	0	0	17	0.56666667	North_America	17	
10	Chan 2019	Australia	Jan 2006 to Dec 2016	Adults	42	10	6	26	Inpatient	56	20	8	11	0	5	1	0	1	0	18	0.42857143	Australia	18	
11	Saka 2019	Togo	Jan 1995 to Dec 2017	Both	177	138	29	10	Inpatient	31.7+/-13.0	113	17	7	0	63	0	0	0	0	70	0.39548023	Africa	70	
12	Zhang 2019	US	Jan 2007 to Feb 2017	Both	48	11	9	29	Inpatient	49.2+/-22.6	28	6	8	1	12	0	0	0	3	24	0.50000000	North_America	24	
13	Noor 2017	Pakistan	2013-2015	Both	54	NR	NR	NR	Inpatient	25.9+/-17.4	27	9	11	2	2	0	0	0	0	15	0.27777778	Asia_south	15	
14	Techasatjan 2017	Thailand	1992 to 2012	Children	30	24	0	6	Inpatient	6.9+/-4.4	12	2	3	0	2	3	0	0	0	8	0.26666667	Asia_southeast	8	
15	Van Batavia 2017	US	Jan 2008 to Jun 2015	Children	31	NR	NR	NR	Inpatient	10.9 [IQR 7.4- 14.1]	15	0	2	0	4	3	0	0	0	9	0.29032258	North_America	9	
16	Wang 2017	China	Jan 2006 to Dec 2015	Both	88	48	34	6	Inpatient	45+/-18	48	6	11	5	0	3	0	1	4	24	0.27272727	Asia_east	24	
17	McCullough 2016	US	2000 to 2015	Unclear	40	NR	NR	NR	Inpatient	45.9	22	4	5	3	6	3	1	1	0	19	0.47500000	North_America	19	
18	Milliszewski 2016	Canada	Jan 2001 to Dec 2011	Adults	64	28	19	17	Inpatient	55.4+/-20	36	14	4	1	4	1	1	1	0	12	0.18750000	North_America	12	
19	Misiriloglu 2016	Turkey	Jan 2011 to May 2016	Both	35	28	3	4	Inpatient	9.13 [IQR 7.05- 12.85]	17	1	13	0	0	2	0	0	1	16	0.45714286	Europe	16	
20	Suwarsa 2016	Indonesia	Jan 2009 to Dec 2013	Both	57	39	7	11	Inpatient	NR	27	7	3	0	0	0	0	0	2	5	0.08771930	Asia_southeast	5	
21	Cruger 2015	Denmark	Jan 2001 to Dec 2010	Both	48	38	0	10	Inpatient	52.4 (3-81, SJS), 62.9 (44-77, TEN)	24	0	8	1	3	0	0	0	4	16	0.33333333	Europe	16	
22	Quirke 2015	US	Oct 1997 to Aug 2012	Children	41	8	10	23	Inpatient	11.2+/-4.5	21	0	5	0	5	3	0	0	0	13	0.31707317	North_America	13	
23	Sun 2014	China	Jan 2007 to Dec 2013	Both	87	39	0	48	Inpatient	NR	51	2	11	1	0	0	0	0	3	15	0.17241379	Asia_east	15	
24	Lalosevic 2014	Serbia	1993 to 2012	Both	38	13	8	17	Inpatient	42.1 (1-94)	21	3	4	0	4	0	1	0	1	10	0.26315789	Europe	10	
25	Saka 2013	Africa	2000 to 2010	Both	177	129	11	37	Inpatient Outpatient	32.3+/-15.4	111	22	6	0	68	0	0	0	0	74	0.41807910	Africa	74	
26	Choon 2012	Malaysia	Jan 2001 to Dec 2010	Both	109	88	0	21	Inpatient	NR	NA	8	7	0	14	0	0	0	2	23	0.21100917	Asia_southeast	23	
27	Firoz 2015	US	5 years	Unclear	82	0	0	82	Inpatient	45.1+/-20.4	38	24	8	3	30	1	0	0	1	43	0.52439024	North_America	43	
28	Hu 2012	Taiwan	Jan 1992 to Dec 2009	Unclear	101	0	0	101	Inpatient	56.1+/-23.1	57	53	12	0	2	2	0	2	0	18	0.17821782	Asia_east	18	
29	Kim 2012	South Korea	Jun 2001 to Jun 2011	Adults	82	71	0	11	Inpatient	53.8 (18-94)	45	8	6	1	1	0	1	0	0	9	0.10975610	Asia_east	9	
30	Finkelstein 2011	US, Canada	Jan 2000 to Dec 2007	Children	55	47	3	5	Inpatient	9.6+/-4.8	20	1	2	0	7	2	0	0	0	11	0.20000000	North_America	11	
31	Huang 2009	Taiwan	Jan 2002 to Aug 2007	Adults	52	42	0	10	Inpatient	52+/-18	26	2	1	0	1	0	0	0	1	3	0.05769231	Asia_east	3	
32	Wanat 2009	US	Jan 1995 to Dec 2007	Adults	50	17	2	31	Inpatient	49.5+/-17.1	30	16	5	1	14	0	0	0	0	20	0.40000000	North_America	20	
33	Yang 2009	China	Jan 1993 to Oct 2007	Both	65	18	0	47	ICU	43.7+/-23.1	32	13	4	2	2	0	0	0	0	8	0.12307692	Asia_east	8	
34	SushmanM2005	India	Jan 1994 to Dec 2002	Unclear	96	75	0	21	Inpatient	NR	NA	0	11	19	1	1	1	0	0	33	0.34375000	Asia_south	33	
35	Palmieri 2002	US	Jan 1995 to Jan 2000	Both	199	0	0	199	Burn unit	46+/-1.7	NA	64	28	0	44	0	0	0	22	94	0.47236181	North_America	94	
36	Power 1995	US	1960 to 1994	Both	67	37	0	30	Inpatient	32 (SJS), 35 (TEN)	41	9	5	2	13	0	0	0	1	21	0.31343284	North_America	21	
37	Leenutaphong 1993	Thailand	Jan 1981 to Jan 1990	Both	78	58	0	20	Inpatient	32+/-19.5 (SJS) 36.2+/-20.7 (TEN)	47	11	19	0	9	1	3	0	0	32	0.41025641	Asia_southeast	32	
38	Guillaume 1987	France	Nov 1972 to Jul 1985	Both	87	10	32	45	Inpatient	45.8+/-22.7	55	22	0	0	18	1	0	0	0	19	0.21839080	Europe	19	

ANEXO 2 – CÓDIGO DE R

Código de R - Análisis de prevalencias con GLMMs

Alejandro Ballester Palomar

```
rm(list=ls())

#Librerías

library(readxl)
library(tidyverse)

— Attaching core tidyverse packages ————— tidyverse 2.0.0 —
✓ dplyr      1.1.3      ✓ readr      2.1.4
✓ forcats    1.0.0      ✓ stringr    1.5.0
✓ ggplot2    3.4.3      ✓ tibble     3.2.1
✓ lubridate  1.9.3      ✓ tidyr      1.3.0
✓ purrr      1.0.2

— Conflicts ————— tidyverse_conflicts() —
X dplyr::filter() masks stats::filter()
X dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(lme4)

Loading required package: Matrix

Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

  expand, pack, unpack

library(parallel)
library(DHARMa)

Warning: package 'DHARMa' was built under R version 4.3.3

This is DHARMa 0.4.7. For overview type '?DHARMa'. For recent changes,
type news(package = 'DHARMa')

library(lattice)
library(broom.mixed)

Warning: package 'broom.mixed' was built under R version 4.3.3

library(metafor)

Warning: package 'metafor' was built under R version 4.3.3
```

```
Loading required package: metadat
Warning: package 'metadat' was built under R version 4.3.3
Loading required package: numDeriv
Loading the 'metafor' package (version 4.8-0). For an
introduction to the package please type: help(metafor)
library(meta)
Warning: package 'meta' was built under R version 4.3.3
Loading 'meta' package (version 8.0-2).
Type 'help(meta)' for a brief overview.
library(glue)
```

FUNCIONES

```
conversion <- function(x, link) {
  model_summary <- summary(x)
  mu <- model_summary$coefficients[1]
  mu.se <- model_summary$coefficients[2]
  alpha <- 0.05
  z_crit <- qnorm(1 - alpha / 2)
  mu.ci <- c(mu - z_crit * mu.se, mu + z_crit * mu.se)
  tau2 <- as.numeric(model_summary$varcor)
  tau <- sqrt(tau2)

  if (link == "logit") {
    mu_transformado <- plogis(mu)
    ci_transformado <- plogis(mu.ci)
    return(list(mu = mu_transformado, "CI 95%" = ci_transformado, "tau2"
    "=tau2, "tau" = tau))
  } else if (link == "probit") {
    mu_transformado <- pnorm(mu)
    ci_transformado <- pnorm(mu.ci)
    return(list(mu = mu_transformado, "CI 95%" = ci_transformado, "tau2"
    "=tau2, "tau" = tau))
  } else if (link == "cauchit") {
    mu_transformado <- pcauchy(mu)
    ci_transformado <- pcauchy(mu.ci)
    return(list(mu = mu_transformado, "CI 95%" = ci_transformado, "tau2"
    "=tau2, "tau" = tau))
  } else if (link == "cloglog") {
    back.trans <- function(x2) 1 - exp(-exp(x2))
    mu_transformado <- back.trans(mu)
    ci_transformado <- back.trans(mu.ci)
    return(list(mu = mu_transformado, "CI 95%" = ci_transformado, "tau2"
    "=tau2, "tau" = tau))
  } else if (link == "log") {
    mu_transformado <- exp(mu)
  }
}
```

```
ci_transformado <- exp(mu.ci)
return(list(mu = mu_transformado, "CI 95%" = ci_transformado,"tau2
"=tau2, "tau" = tau))
} else {
  stop("El enlace especificado no es válido. Las opciones son: 'log'
, 'logit', 'probit', 'cauchit' o 'cloglog'.")
}
}
```

#Datos

```
data <- read_excel(path = "./01_Data.base/Datos_TFM_AB_V01.xlsx")
```

New names:

- `` -> `...23`
- `` -> `...24`

Depuración de los datos

```
data_t <- data |>
  select(-c(...23,...24))

data_t <- data_t|>
  rename("study"=Study ,
        "country" = Country,
        "period" = `Study period` ,
        "population" = `Study population`,
        "total_patients"=`Total patients`,
        "SJS" = SJS,
        "SJS_TEN" = `SJS / TEN`,
        "TEN" = TEN,
        "setting" =Setting,
        "age"=`Age*`,
        "female"=`Female, n`,
        "death" = `Death, n`,
        "pcn" = `PCN, n`,
        "cephalo" = Cephalo,
        "fq"=FQ,
        "sulfa" =Sulfa,
        "macro"=Macro,
        "tetra"=Tetra,
        "vanco"=Vanco,
        "other_abx"=Other_abx,
        "other_drugs"=Other_drugs,
        "other_causes" = `Non-drug_causes`) |>
  mutate(pcn = as.numeric(pcn),
        cephalo = as.numeric(cephalo),
        fq = as.numeric(fq),
        sulfa = as.numeric(sulfa),
        macro = as.numeric(macro),
        tetra = as.numeric(tetra),
        vanco = as.numeric(vanco),
        other_abx = as.numeric(other_abx),
        other_drugs = as.numeric(other_drugs),
```

```
      other_causes = as.numeric(other_causes)) |>
    mutate(ID = seq(from = 1, to = length(study)), .before = study)

Warning: There were 6 warnings in `mutate()`.
The first warning was:
i In argument: `pcn = as.numeric(pcn)`.
Caused by warning:
! NAs introducidos por coerción
i Run `dplyr::last_dplyr_warnings()` to see the 5 remaining warnings.

data_tt <- data_t |>
  pivot_longer(cols = pcn:other_causes, names_to = "causes", values_to = "cases")

cases2 <- data_tt |>
  pull(cases)

data_tt <- data_tt |>
  mutate(cases2 = is.na(cases2)) |>
  filter(cases2 == FALSE)

abx <- data_tt |>
  filter(causes != "other_drugs") |>
  filter(causes != "other_causes")
```

Prueba datos correctos

Vemos que el número de casos no coincide con el que especifican los autores en su artículo. Por lo que vamos a hacer una revisión de los casos con los materiales suplementarios.

Primero vemos cuáles son los que difieren del artículo:

```
study <- c(
  "De Bustros 2021", "Gleghorn 2021", "Olteanu 2021", "Singh 2021", "T
hakur 2021",
  "Abou-Taleb 2020", "Huyen 2020", "Yang 2020", "Carrasquillo 2019", "
Chan 2019",
  "Saka 2019", "Zhang 2019", "Noor 2017", "Techasatian 2017", "Van Bat
avia 2017",
  "Wang 2017", "McCullough 2016", "Miliszewski 2016", "Misirlioglu 201
6", "Suwarsa 2016",
  "Cruger 2015", "Quirke 2015", "Sun 2014", "Lalosevic 2014", "Saka 20
13",
  "Choon 2012", "Firoz 2012", "Hu 2012", "Kim 2012", "Finkelstein 2011
",
  "Huang 2009", "Wanat 2009", "Yang 2009", "Sushman 2005", "Palmieri 2
002",
  "Power 1995", "LEENUTAPHONG 1993", "Guillaume 1987"
)
events <- c(
  60, 16, 11, 23, 6, 15, 0, 61, 17, 18, 70, 24, 15, 8, 9,
  24, 19, 12, 16, 5, 16, 13, 15, 10, 74, 23, 43, 18, 9, 11,
  3, 20, 8, 33, 94, 21, 32, 19
```

```
)  
total <- c(  
  131, 51, 35, 75, 50, 77, 40, 175, 29, 34, 144, 41, 51, 30, 14,  
  49, 34, 47, 29, 23, 44, 37, 87, 33, 156, 109, 82, 62, 37, 28,  
  41, 41, 64, 96, 179, 43, 60, 67  
)  
id<- seq(length(events))
```

Creamos una nueva base de datos

```
study_data <- data.frame(study, id, events, total)
```

```
abx |>  
  summarise(sum(cases), .by = c(ID,study)) |>  
  full_join(study_data, by = c("ID" = "id")) |>  
  rename(inicial = "sum(cases)") |>  
  mutate(test = if_else(inicial==events, TRUE, FALSE)) |>  
  filter(test == FALSE)
```

A tibble: 6 × 7

	ID	study.x	inicial	study.y	events	tota
1	test					
	<int>	<chr>	<dbl>	<chr>	<dbl>	<dbl>
>	<lg1>					
1	1	"De Bustros 2021"	45	De Bustros 2021	60	13
1	FALSE					
2	8	"Yang 202040"	50	Yang 2020	61	17
5	FALSE					
3	9	"Carrasquillo\r\n201944"	11	Carrasquillo 2019	17	2
9	FALSE					
4	23	"Sun 201466"	14	Sun 2014	15	8
7	FALSE					
5	27	"Firoz 201575"	35	Firoz 2012	43	8
2	FALSE					
6	37	"Leenutaphong 199391"	34	LEENUTAPHONG 1993	32	6
0	FALSE					

En el material suplementario se especifica lo siguiente:

Four studies reported the numbers of beta-lactams only (n=25 in DeBustros 2021, n=20 in Kridin 2021, n=6 in Carasquillo 2020, and n=8 in Firoz 2012), so the numbers of penicillins and cephalosporins were denoted as “-“.

Hay una serie de errores en el material suplementario. Una vez revisados con los artículos originales vemos que los números coinciden con los reportados en el artículo. Por lo que cargamos de nuevo la base de datos con los datos correctos.

Datos finales

```
data <- read_excel(path = "./01_Data.base/Datos_TFM_AB_V02.xlsx") |>  
  mutate(beta = as.numeric(beta),  
         fq = as.numeric(fq),
```

```
    sulfa = as.numeric(sulfa),
    macro = as.numeric(macro),
    tetra = as.numeric(tetra),
    vanco = as.numeric(vanco),
    other_abx = as.numeric(other_abx),
    other_drugs = as.numeric(other_drugs),
    other_causes = as.numeric(other_causes)) |>
mutate(ID = seq(from = 1, to = length(study)), .before = study)

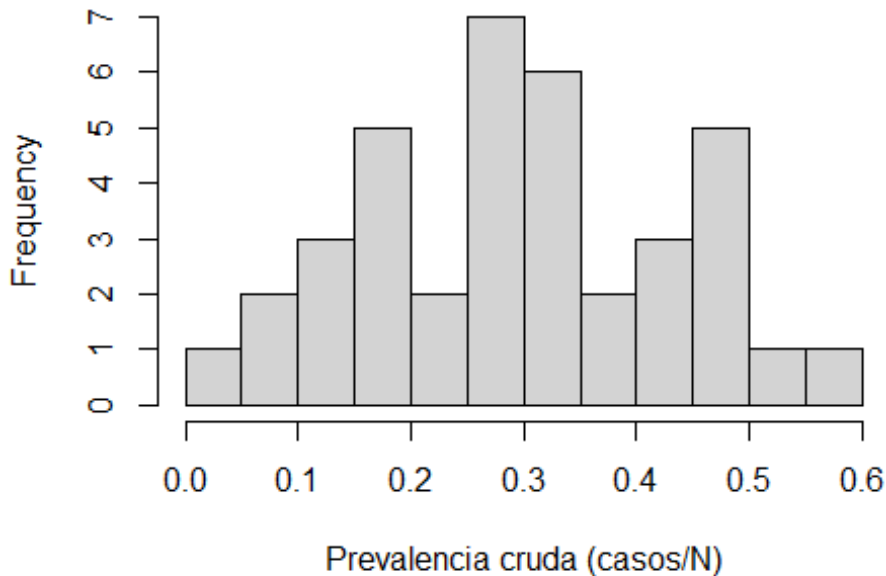
abx <- data |>
  select(-c(other_drugs, other_causes)) |>
  mutate(total = sum(c(beta, fq, sulfa, macro, tetra, vanco, other_abx)), .by = ID) |>
  mutate(prev = total/N, .by = ID) |>
  mutate(continente = factor(case_when(
    country == "Australia" ~ "Australia",
    country %in% c("US", "Canada", "Puerto Rico", "US, Canada") ~ "North_America",
    country %in% c("Egypt", "Togo", "Sub-Saharan Africa", "Africa") ~ "Africa",
    country %in% c("Turkey", "Denmark", "Serbia", "France") ~ "Europe",
    ,
    country %in% c("India", "Pakistan") ~ "Asia_south",
    country %in% c("China", "Taiwan", "South Korea") ~ "Asia_east",
    country %in% c("Vietnam", "Thailand", "Indonesia", "Malaysia") ~ "Asia_southeast",
    TRUE ~ "No clasificado"
  ))) |>
  mutate(continente = relevel(continente, "Australia"))
```

#Descripción de los datos

##Test normalidad

```
hist(abx$prev, xlab = "Prevalencia cruda (casos/N)", main = "Histograma de prevalencia de SSJ y TEN asociada a antibi\u00f3ticos", breaks = 12 )
```

Histograma prevalencia de SSJ y TEN asociada a antib



Frecuencias marginales y acumuladas

```
abx |>
  distinct(ID, .keep_all = TRUE) |>
  mutate(interval = case_when(prev <= 0.1 ~ "0-0.1",
                              prev > 0.1 & prev <= 0.2 ~ "0.1-0.2",
                              prev > 0.2 & prev <= 0.3 ~ "0.2-0.3",
                              prev > 0.3 & prev <= 0.4 ~ "0.3-0.4",
                              prev > 0.4 & prev <= 0.5 ~ "0.4-0.5",
                              prev > 0.5 & prev <= 0.6 ~ "0.5-0.6",
                              TRUE ~ "Funciona")) |>
  mutate(interval = factor(interval)) |>
  pull(interval) |>
  fct_count() |>
  mutate(freq = n/sum(n)) |>
  mutate(cum = cumsum(freq))
```

```
# A tibble: 6 × 4
  f      n  freq  cum
<fct> <int> <dbl> <dbl>
1 0-0.1     3 0.0789 0.0789
2 0.1-0.2    8 0.211  0.289
3 0.2-0.3    9 0.237  0.526
4 0.3-0.4    8 0.211  0.737
5 0.4-0.5    8 0.211  0.947
6 0.5-0.6    2 0.0526 1
```

```
#Meta-análisis de prevalencias
```

CONVERGENCIA ENLACE CAUCHIT

##Eliminar estudio 7

```
abx_cau <- abx |>
  mutate(n = total) |>
  filter(ID != 7)

mod_logc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, family =
binomial(link = "log"))
mod_logitc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, family
= binomial(link = "logit"))
mod_probitc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, famil
y = binomial(link = "probit"))
mod_cauchitc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, fami
ly = binomial(link = "cauchit"))
mod_cloglogc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, fami
ly = binomial(link = "cloglog"))

conversion(mod_logc, "log")

$mu
[1] 0.2796186

$`CI 95%`
[1] 0.2403626 0.3252860

$tau2
[1] 0.1725126

$tau
[1] 0.4153463

conversion(mod_logitc, "logit")

$mu
[1] 0.2867897

$`CI 95%`
[1] 0.2445384 0.3331221

$tau2
[1] 0.3649931

$tau
[1] 0.6041466

conversion(mod_probitc, "probit")

$mu
[1] 0.2891122

$`CI 95%`
[1] 0.2466391 0.3347421
```

```
$tau2
[1] 0.129858

$tau
[1] 0.3603581

conversion(mod_cauchitc, "cauchit")

$mu
[1] 0.2750865

$`CI 95%`
[1] 0.2365127 0.3236446

$tau2
[1] 0.3978478

$tau
[1] 0.6307518

conversion(mod_cloglogc, "cloglog")

$mu
[1] 0.2832665

$`CI 95%`
[1] 0.2424037 0.3293931

$tau2
[1] 0.2552074

$tau
[1] 0.5051806

test1 <- escalc("PAS", xi = n, ni = N, data = abx_cau, add = 0.5)

rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, )

sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)

(sin(mu.est))^2
[1] 0.2921384

(sin(mu.ci))^2
[1] 0.2495938 0.3365892

##Incluir valor mínimo de casos tolerado

abx_cau <- abx |>
  mutate(n = if_else(total==0, 2, total))
```

```
mod_logc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, family =  
binomial(link = "log"))  
mod_logitc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, family  
= binomial(link = "logit"))  
mod_probitc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, fami  
ly = binomial(link = "probit"))  
mod_cauchitc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, fami  
ly = binomial(link = "cauchit"))  
mod_cloglogc <- glmer(cbind(n, N - n)~ 1 +(1|ID), data = abx_cau, fami  
ly = binomial(link = "cloglog"))  
  
conversion(mod_logc, "log")  
  
$mu  
[1] 0.268396  
  
$`CI 95%`  
[1] 0.2276101 0.3164903  
  
$tau2  
[1] 0.215615  
  
$tau  
[1] 0.4643436  
  
conversion(mod_logitc, "logit")  
  
$mu  
[1] 0.2763156  
  
$`CI 95%`  
[1] 0.2321626 0.3253081  
  
$tau2  
[1] 0.4448279  
  
$tau  
[1] 0.6669542  
  
conversion(mod_probitc, "probit")  
  
$mu  
[1] 0.279117  
  
$`CI 95%`  
[1] 0.2346941 0.3272721  
  
$tau2  
[1] 0.1564142  
  
$tau  
[1] 0.3954924  
  
conversion(mod_cauchitc, "cauchit")
```

```
$mu
[1] 0.2634358

$`CI 95%`
[1] 0.2240972 0.3142377

$tau2
[1] 0.5025426

$tau
[1] 0.7089024

conversion(mod_cloglogc, "cloglog")

$mu
[1] 0.2723888

$`CI 95%`
[1] 0.229808 0.321078

$tau2
[1] 0.3147164

$tau
[1] 0.5609959

test1 <- escalc("PAS", xi = n, ni = N, data = abx_cau, add = 0.5)

rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, )

sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)

(sin(mu.est))^2
[1] 0.2829845

(sin(mu.ci))^2
[1] 0.238886 0.329273

abx <- abx_cau |>
  mutate(total = n)
```

Convencional: Dos pasos

Selección del tipo de transformación usada por los autores: Raw:

```
test1 <- escalc("PR", xi = total, ni = N, data = abx, add = 0.5)

rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt
```

Random-Effects Model (k = 38; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.0200 (SE = 0.0061)
tau (square root of estimated tau^2 value): 0.1415
I^2 (total heterogeneity / total variability): 90.30%
H^2 (total variability / sampling variability): 10.30

Test for Heterogeneity:
Q(df = 37) = 381.2561, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.2893	0.0247	11.7326	<.0001	0.2409	0.3376	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log:

```
test1 <- escalc("PLN", xi = total, ni = N, data = abx, add = 0.5)
rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt
```

Random-Effects Model (k = 38; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.1143 (SE = 0.0421)
tau (square root of estimated tau^2 value): 0.3380
I^2 (total heterogeneity / total variability): 80.83%
H^2 (total variability / sampling variability): 5.22

Test for Heterogeneity:
Q(df = 37) = 193.0374, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
-1.2289	0.0653	-18.8259	<.0001	-1.3568	-1.1010	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)
```

```
(exp(mu.est))
```

```
[1] 0.2926165
```

```
(exp(mu.ci))
```

```
[1] 0.2574751 0.3325541
```

Probit:

```
test1 <- escalc("PRZ", xi = total, ni = N, data = abx, add = 0.5)
rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt
```

Random-Effects Model (k = 38; tau² estimator: DL)

```
tau^2 (estimated amount of total heterogeneity): 0.1230 (SE = 0.0395)
tau (square root of estimated tau^2 value):      0.3507
I^2 (total heterogeneity / total variability):   83.48%
H^2 (total variability / sampling variability):  6.05
```

Test for Heterogeneity:

```
Q(df = 37) = 223.9994, p-val < .0001
```

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
-0.5716	0.0639	-8.9438	<.0001	-0.6968	-0.4463	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)
```

```
(pnorm(mu.est))
```

```
[1] 0.2838019
```

```
(pnorm(mu.ci))
```

```
[1] 0.2429509 0.3276808
```

Logit

```
test1 <- escalc("PLO", xi = total, ni = N, data = abx, add = 0.5)
rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt
```

Random-Effects Model (k = 38; tau² estimator: DL)

```
tau^2 (estimated amount of total heterogeneity): 0.3042 (SE = 0.1018)
tau (square root of estimated tau^2 value):      0.5515
I^2 (total heterogeneity / total variability):   81.78%
H^2 (total variability / sampling variability):  5.49
```

Test for Heterogeneity:

Q(df = 37) = 203.0341, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
-0.9115	0.1026	-8.8834	<.0001	-1.1126	-0.7104	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)
```

```
(plogis(mu.est))
```

```
[1] 0.2867007
```

```
(plogis(mu.ci))
```

```
[1] 0.2473939 0.3295184
```

Arcoseno:

```
test1 <- escalc("PAS", xi = total, ni = N, data = abx, add = 0.5)
```

```
rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt
```

Random-Effects Model (k = 38; tau² estimator: DL)

tau² (estimated amount of total heterogeneity): 0.0209 (SE = 0.0063)
tau (square root of estimated tau² value): 0.1446
I² (total heterogeneity / total variability): 86.40%
H² (total variability / sampling variability): 7.35

Test for Heterogeneity:

Q(df = 37) = 272.0995, p-val < .0001

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.5609	0.0256	21.8784	<.0001	0.5107	0.6112	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)
```

```
(sin(mu.est))^2
```

```
[1] 0.2829845
(sin(mu.ci))^2
[1] 0.238886 0.329273
B <- 0.5*((log(292.5256)-log(37))/(sqrt(2*292.5256)-sqrt(2*(37-1))))
L <- exp(0.5*log(292.5256/37)-1.96*B)
U <- exp(0.5*log(292.5256/37)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.8363, 0.9023
```

De acuerdo a los resultados obtenidos, es el link arcoseno el que han usado los autores

Modelos lineales mixtos generalizados (GLMMs)

Modelo 1: enlace log

```
mod_log <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx, famil
y = binomial(link = "log"))
sum <- conversion(mod_log, "log");sum
$mu
[1] 0.268396
$`CI 95%`
[1] 0.2276101 0.3164903
$tau2
[1] 0.215615
$tau
[1] 0.4643436
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = abx, family = binomial(
link = "log"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]

test <- abx |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(p)) |>
```

```
mutate(vc = (1-p)/(N*p)) |>
mutate(w = 1/vc) |>
mutate(yf = ((yc-(ccau))^2)) |>
mutate(wyf = w*yf)

Q <- sum(test$wyf);Q
[1] 231.903

df <- length(abx$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
p val = 4.59561661336037e-30

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.8405

glue("tau2 = {sum$tau2}")
tau2 = 0.215614992995372

glue("tau = {sum$tau}")
tau = 0.464343615219776

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.7898, 0.8789
```

Forestplot

```
data_forest <- abx |>
mutate (
  LL = prev - (1.96 * (sqrt(prev * (1 - prev) / N))),
  UL = prev + (1.96 * (sqrt(prev * (1 - prev) / N))) |>
add_row(prev = sum$mu,
  ID = 39,
  study = "Effect size",
  UL = sum$'CI 95%' [2],
  LL = sum$'CI 95%' [1]
```

```
)  
plot_data <- data_forest |>  
  mutate(CI_texto = as.character(glue("{round(LL,2)}, {round(UL,2)}"))  
)  
  
library(gridExtra)  
  
Warning: package 'gridExtra' was built under R version 4.3.3  
  
Attaching package: 'gridExtra'  
  
The following object is masked from 'package:dplyr':  
  
  combine  
  
plot1 <- ggplot(plot_data, aes(y = ID, x = prev)) +  
  geom_point(aes(color = ifelse(study == "Effect size", "red", "black"  
)),  
             shape = 15,  
             size = 3.5) +  
  geom_errorbarh(aes(xmin = LL,  
                    xmax = UL,  
                    color = ifelse(study == "Effect size", "red", "black"  
ack))),  
                height = 0.25) +  
  geom_vline(xintercept = sum$mu,  
            linetype = "dashed",  
            color = "black",  
            linewidth = 0.8,  
            alpha = 0.8) +  
  scale_y_continuous(name = "",  
                    breaks = plot_data$ID,  
                    labels = plot_data$study,  
                    trans = "reverse") +  
  xlab("Prevalencia") +  
  ylab(" ") +  
  labs(title = "hi")+  
  theme_bw() +  
  coord_cartesian(xlim = c(min(plot_data$LL) - 0.05, max(plot_data$UL)  
+ 0.05)) +  
  scale_color_identity() +  
  theme(plot.title = element_text(hjust = 0.5, size=12, colour = "white"),  
        legend.position = "none",  
        panel.border = element_blank(),  
        panel.background = element_blank(),  
        panel.grid.major = element_blank(),  
        panel.grid.minor = element_blank(),  
        axis.line.x = element_line(colour = "black"),  
        axis.line.y = element_blank(),  
        axis.ticks.y = element_blank(),  
        axis.text.y = element_text(size = 12, colour = "black", hjust  
= 0),  
        axis.text.x = element_text(size = 12, colour = "black"),
```

```
axis.title.x = element_text(size = 12, colour = "black"))

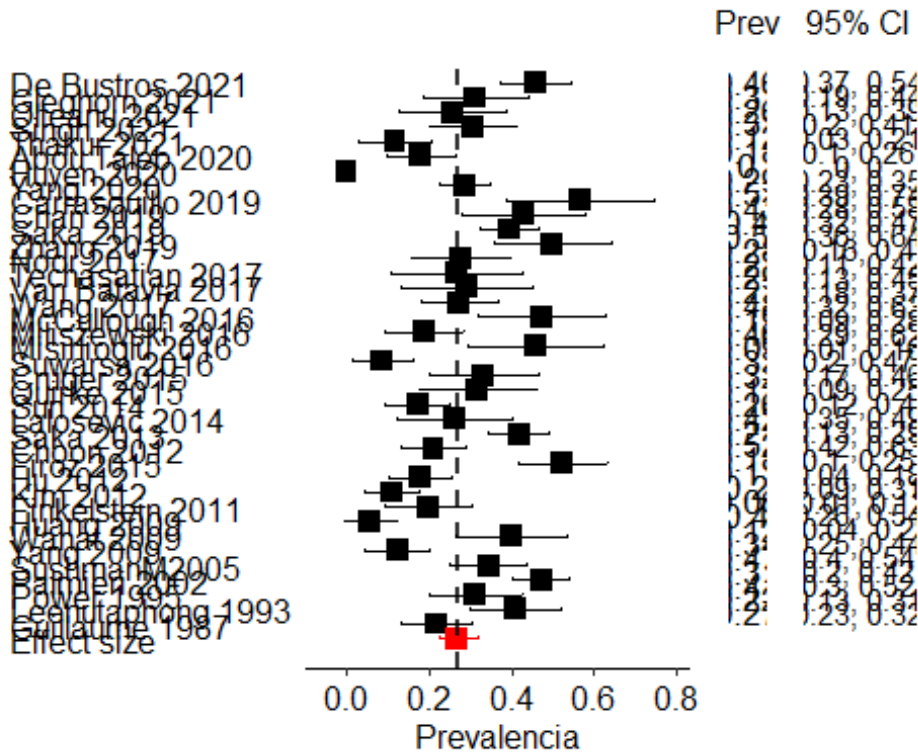
table_base <- ggplot(plot_data, aes(y=study)) +
  ylab(NULL) + xlab(" ") +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        axis.text.x = element_text(color="white", hjust = -3, size = 2
5),
        axis.line = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y = element_blank(),
        legend.position = "none",
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.background = element_blank())

tab1 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = as.character(round(prev,2)
), size = 4)) +
  ggtitle("Prev")

tab2 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = CI_texto), size = 4) +
  ggtitle("95% CI")

lay <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3), nrow = 1)

grid.arrange(plot1, tab1, tab2, layout_matrix = lay)
```



Modelo 2: enlace logit

```
mod_logit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx, family = binomial(link = "logit"))

sum <- conversion(mod_logit, "logit"); sum

$mu
[1] 0.2763156

$`CI 95%`
[1] 0.2321626 0.3253081

$tau2
[1] 0.4448279

$tau
[1] 0.6669542
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = abx, family = binomial(link = "logit"))
scau <-summary(cau)
ccau <- scau$coefficients[1]

test <- abx |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(n/(N-n))) |>
  mutate(vc = 1/(N*p*(1-p))) |>
```

```
mutate(w = 1/vc) |>
mutate(yf = ((yc-(ccau))^2)) |>
mutate(wyf = w*yf)

Q <- sum(test$wyf);Q
[1] 208.8979

df <- length(abx$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
p val = 7.45070480524152e-26

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.8229

glue("tau2 = {sum$tau2}")
tau2 = 0.444827859911675

glue("tau = {sum$tau}")
tau = 0.666954166275071

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.7648, 0.8666
```

Forestplot

```
data_forest <- abx |>
mutate (
  LL = prev - (1.96 * (sqrt(prev * (1 - prev) / N))),
  UL = prev + (1.96 * (sqrt(prev * (1 - prev) / N))) |>
add_row(prev = sum$mu,
  ID = 39,
  study = "Effect size",
  UL = sum$'CI 95%' [2],
  LL = sum$'CI 95%' [1]
)
```

```
plot_data <- data_forest |>
  mutate(CI_texto = as.character(glue("{round(LL,2)}, {round(UL,2)}"))
)

library(gridExtra)

plot1 <- ggplot(plot_data, aes(y = ID, x = prev)) +
  geom_point(aes(color = ifelse(study == "Effect size", "red", "black"
)),
            shape = 15,
            size = 3.5) +
  geom_errorbarh(aes(xmin = LL,
                    xmax = UL,
                    color = ifelse(study == "Effect size", "red", "black"
)),
                height = 0.25) +
  geom_vline(xintercept = sum$mu,
            linetype = "dashed",
            color = "black",
            linewidth = 0.8,
            alpha = 0.8) +
  scale_y_continuous(name = "",
                    breaks = plot_data$ID,
                    labels = plot_data$study,
                    trans = "reverse") +
  xlab("Prevalencia") +
  ylab(" ") +
  theme_bw() +
  coord_cartesian(xlim = c(min(plot_data$LL) - 0.05, max(plot_data$UL)
+ 0.05)) +
  scale_color_identity() +
  theme(legend.position = "none",
        panel.border = element_blank(),
        panel.background = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.line.x = element_line(colour = "black"),
        axis.line.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.text.y = element_text(size = 12, colour = "black", hjust
= 0),
        axis.text.x = element_text(size = 12, colour = "black"),
        axis.title.x = element_text(size = 12, colour = "black"))

table_base <- ggplot(plot_data, aes(y=study)) +
  ylab(NULL) + xlab(" ") +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        axis.text.x = element_text(color="white", hjust = -3, size = 2
5),
        axis.line = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y = element_blank(),
```

```

legend.position = "none",
panel.background = element_blank(),
panel.border = element_blank(),
panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
plot.background = element_blank())

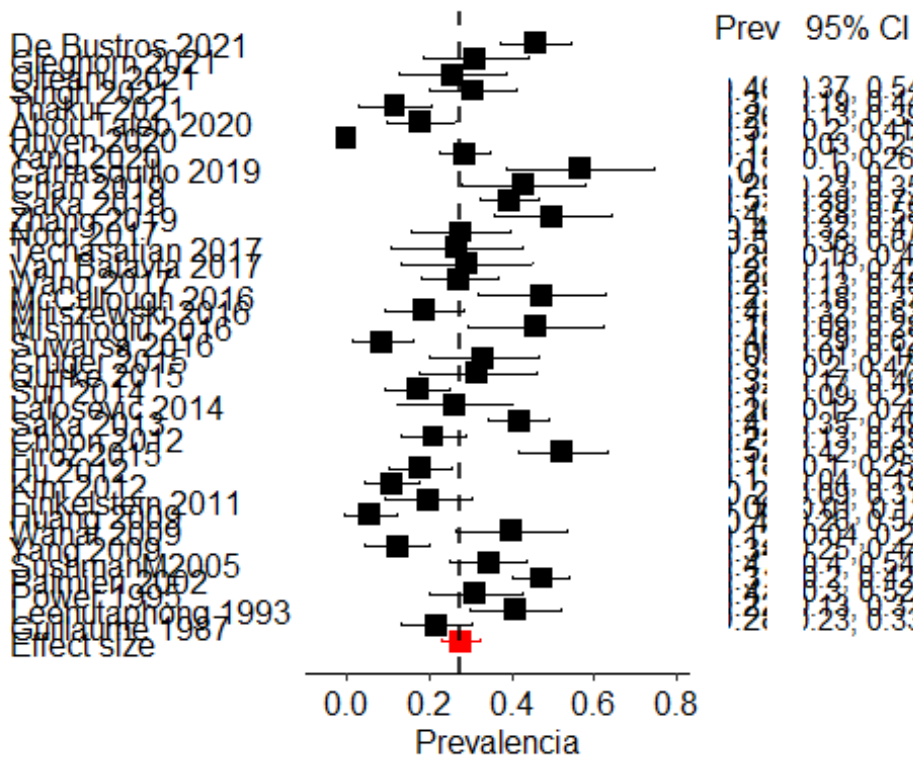
tab1 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = as.character(round(prev,2)
), size = 4)) +
  ggtitle("Prev")

tab2 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = CI_texto), size = 4) +
  ggtitle("95% CI")

lay <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3), nrow = 1)

grid.arrange(plot1, tab1, tab2, layout_matrix = lay)

```



###Modelo 3: enlace probit

```

mod_probit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx, fa
mily = binomial(link = "probit"))
sum <- conversion(mod_probit, "probit"); sum

$mu
[1] 0.279117

$`CI 95%`
[1] 0.2346941 0.3272721

```

```
$tau2  
[1] 0.1564142
```

```
$tau  
[1] 0.3954924
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = abx, family = binomial(  
link = "probit"))  
scau <-summary(cau)  
ccau <- scau$coefficients[1]
```

```
test <- abx |>  
  mutate(n = if_else(total==0, 0.5, total)) |>  
  mutate(p = n/N) |>  
  mutate(yc = qnorm(p)) |>  
  mutate(derivada_cuadrada = 2 * pi * exp(yc^2)) |>  
  mutate(var_p = p * (1 - p) / N) |>  
  mutate(vc = derivada_cuadrada*var_p)|>  
  mutate(w = 1/vc) |>  
  mutate(yf = ((yc-(ccau))^2)) |>  
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
```

```
[1] 225.5864
```

```
df <- length(abx$ID)-1
```

```
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
```

```
p val = 6.70192696407781e-29
```

```
H2 <- Q/(length(test$ID)-1)
```

```
H <-sqrt(H2)
```

```
I2 <- (H2-1)/H2
```

```
glue("I2 = {round(I2,4)}")
```

```
I2 = 0.836
```

```
glue("tau2 = {sum$tau2}")
```

```
tau2 = 0.15641424359939
```

```
glue("tau = {sum$tau}")
```

```
tau = 0.395492406500289
```

```
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
```

```
L <- exp(0.5*log(Q/df)-1.96*B)
```

```
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.7835, 0.8758
```

Forestplot

```
data_forest <- abx |>
  mutate (
    LL = prev - (1.96 * (sqrt(prev * (1 - prev) / N))),
    UL = prev + (1.96 * (sqrt(prev * (1 - prev) / N))) |>
  add_row(prev = sum$mu,
    ID = 39,
    study = "Effect size",
    UL = sum$'CI 95%'[2],
    LL = sum$'CI 95%'[1]
  )
plot_data <- data_forest |>
  mutate(CI_texto = as.character(glue("{round(LL,2)}, {round(UL,2)}"))
)

library(gridExtra)

plot1 <- ggplot(plot_data, aes(y = ID, x = prev)) +
  geom_point(aes(color = ifelse(study == "Effect size", "red", "black"
)),
    shape = 15,
    size = 3.5) +
  geom_errorbarh(aes(xmin = LL,
    xmax = UL,
    color = ifelse(study == "Effect size", "red", "black"
)),
    height = 0.25) +
  geom_vline(xintercept = sum$mu,
    linetype = "dashed",
    color = "black",
    linewidth = 0.8,
    alpha = 0.8) +
  scale_y_continuous(name = "",
    breaks = plot_data$ID,
    labels = plot_data$study,
    trans = "reverse") +
  xlab("Prevalencia") +
  ylab(" ") +
  theme_bw() +
  coord_cartesian(xlim = c(min(plot_data$LL) - 0.05, max(plot_data$UL)
+ 0.05)) +
  scale_color_identity() +
  theme(legend.position = "none",
    panel.border = element_blank(),
    panel.background = element_blank(),
```

```
panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
axis.line.x = element_line(colour = "black"),
axis.line.y = element_blank(),
axis.ticks.y = element_blank(),
axis.text.y = element_text(size = 12, colour = "black", hjust
= 0),
axis.text.x = element_text(size = 12, colour = "black"),
axis.title.x = element_text(size = 12, colour = "black"))

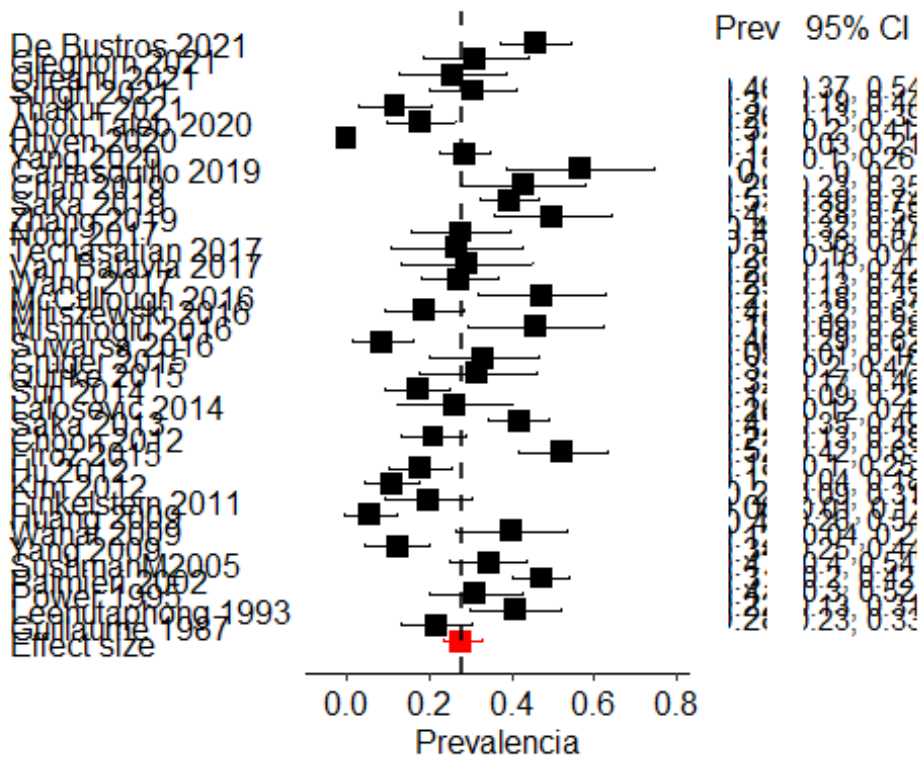
table_base <- ggplot(plot_data, aes(y=study)) +
  ylab(NULL) + xlab(" ") +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        axis.text.x = element_text(color="white", hjust = -3, size = 2
5),
        axis.line = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y = element_blank(),
        legend.position = "none",
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.background = element_blank())

tab1 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = as.character(round(prev,2)
), size = 4)) +
  ggtitle("Prev")

tab2 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = CI_texto), size = 4) +
  ggtitle("95% CI")

lay <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3), nrow = 1)

grid.arrange(plot1, tab1, tab2, layout_matrix = lay)
```



###Modelo 4: enlace cauchit

```
mod_cauchit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx, family = binomial(link = "cauchit"))
sum <- conversion(mod_cauchit, "cauchit"); sum
```

```
$mu
[1] 0.2634358

$`CI 95%`
[1] 0.2240972 0.3142377

$tau2
[1] 0.5025426

$tau
[1] 0.7089024
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = abx, family = binomial(link = "cauchit"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]
test <- abx |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = tan(pi*(p-0.5))) |>
  mutate(vc = (pi*p*(1-p))/(N*cos(pi*(p-0.5))*cos(pi*(p-0.5)))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccaу))^2)) |>
```

```
mutate(wyf = w*yf)

Q <- sum(test$wyf);Q
[1] 1798.855

df <- length(abx$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
p val = 0

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.9794

glue("tau2 = {sum$tau2}")
tau2 = 0.502542614670598

glue("tau = {sum$tau}")
tau = 0.708902401371725

df <- length(abx$ID)-1

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.9762, 0.9823

###Modelo 5: enlace cloglog

mod_cloglog <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx, family = binomial(link = "cloglog"))
sum <- conversion(mod_cloglog, "cloglog"); sum

$mu
[1] 0.2723888

$`CI 95%`
[1] 0.229808 0.321078
```

```
$tau2  
[1] 0.3147164
```

```
$tau  
[1] 0.5609959
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = abx, family = binomial(  
link = "cloglog"))  
scau <-summary(cau)  
ccau <- scau$coefficients[1]
```

```
test <- abx |>  
  mutate(n = if_else(total==0, 0.5, total)) |>  
  mutate(p = n/N) |>  
  mutate(yc = log(-log(1-p))) |>  
  mutate(vc = p/(N*(1-p)*(log(1-p)^2))) |>  
  mutate(w = 1/vc) |>  
  mutate(yf = ((yc-(ccau))^2)) |>  
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
```

```
[1] 216.3678
```

```
df <- length(abx$ID)-1
```

```
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
```

```
p val = 3.26752269383772e-27
```

```
H2 <- Q/(length(test$ID)-1)
```

```
H <-sqrt(H2)
```

```
I2 <- (H2-1)/H2
```

```
glue("I2 = {round(I2,4)}")
```

```
I2 = 0.829
```

```
glue("tau2 = {sum$tau2}")
```

```
tau2 = 0.314716392554994
```

```
glue("tau = {sum$tau}")
```

```
tau = 0.560995893527746
```

```
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
```

```
L <- exp(0.5*log(Q/df)-1.96*B)
```

```
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

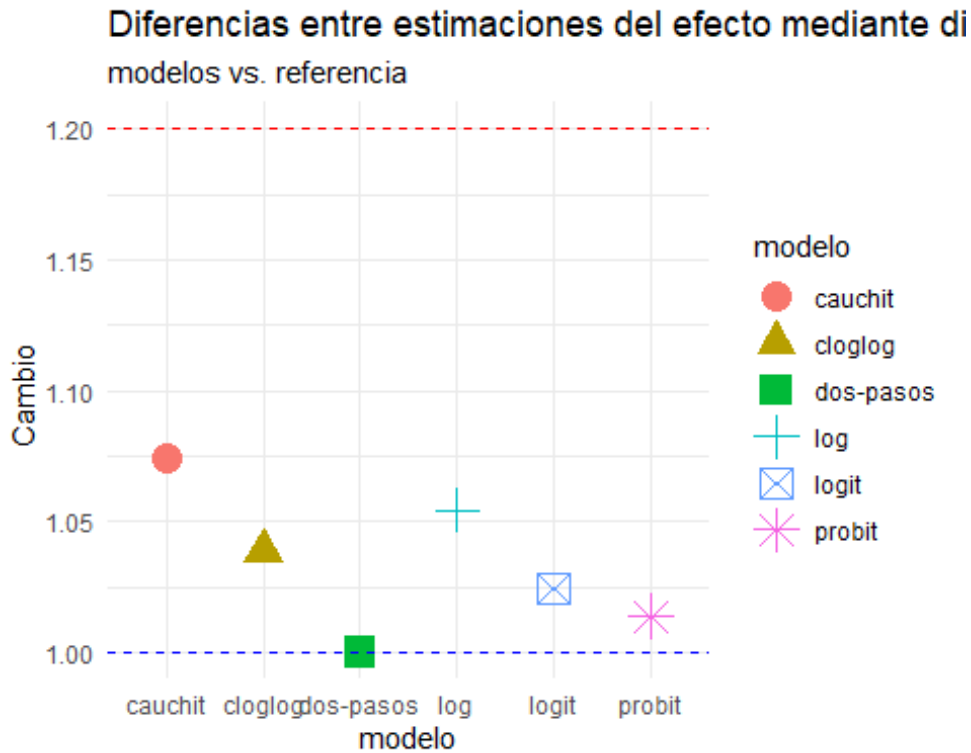
IC95% de I2 = 0.7735, 0.8709
```

COMPARACIÓN ENTRE MODELOS

Estimaciones puntuales

```
comp <- tibble(modelo = c("log",
                          "logit",
                          "probit",
                          "cauchit",
                          "cloglog",
                          "dos-pasos"),
               estimacion = c(conversion(mod_log, "log")$mu,
                              conversion(mod_logit, "logit")$mu,
                              conversion(mod_probit, "probit")$mu,
                              conversion(mod_cauchit, "cauchit")$mu,
                              conversion(mod_cloglog, "cloglog")$mu,
                              (((sin(mu.est))^2)))) |>
  mutate(ratio = ((sin(mu.est))^2)/estimacion)

ggplot(comp, aes(modelo, ratio))+
  geom_point(aes(color = modelo, shape = modelo), size = 5)+
  geom_hline(yintercept = 1, colour = "blue", linetype = "dashed")+
  geom_hline(yintercept = 1.2, colour = "red", linetype = "dashed")+
  theme_minimal() +
  labs(
    title = "Diferencias entre estimaciones del efecto mediante difere
ntes modelos",
    subtitle = "modelos vs. referencia ",
    x = "modelo",
    y = "Cambio "
  )
```



AIC

```
AIC(mod_logc)
```

```
[1] 284.0112
```

```
AIC(mod_logitc)
```

```
[1] 280.0582
```

```
AIC(mod_probitc)
```

```
[1] 279.3212
```

```
AIC(mod_cauchitc)
```

```
[1] 287.0299
```

```
AIC(mod_cloglogc)
```

```
[1] 281.7269
```

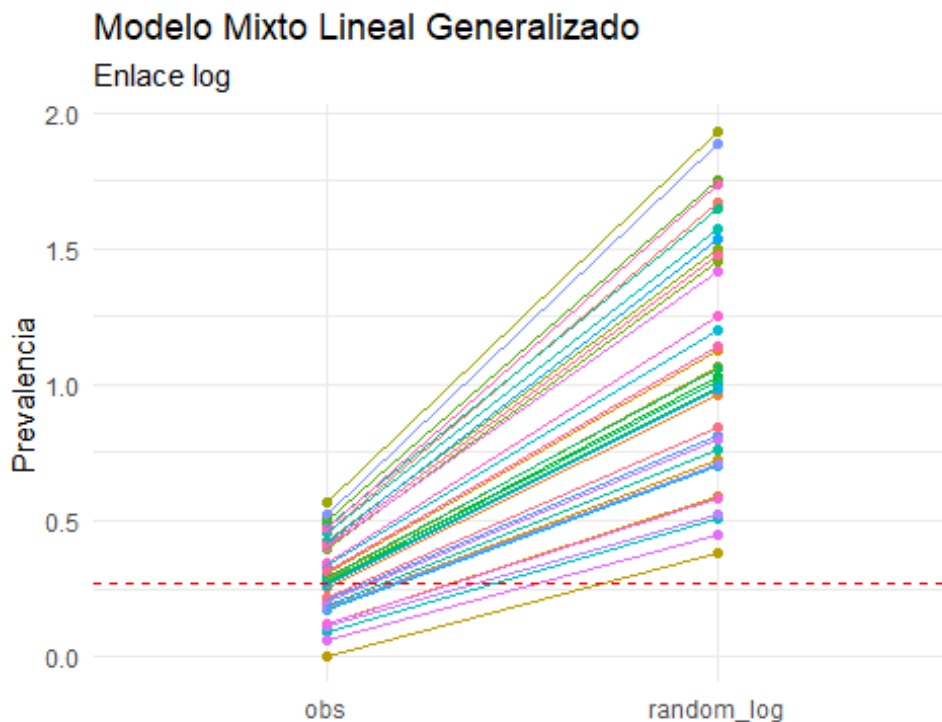
Efectos aleatorios

```
random_log <- exp(as.data.frame(ranef(mod_logc))$condval)
random_logit <- plogis(as.data.frame(ranef(mod_logitc))$condval)
random_probit <- pnorm(as.data.frame(ranef(mod_probitc))$condval)
random_cau <- pcauchy(as.data.frame(ranef(mod_cauchitc))$condval)
random_clog <- (1 - exp(-exp((as.data.frame(ranef(mod_cloglogc))$condval))))
random_conv <- (sin((ranef(rslt)$pred)))^2
```

```
obs <- abx$prev  
ID <- abx_cau$ID
```

enlace log

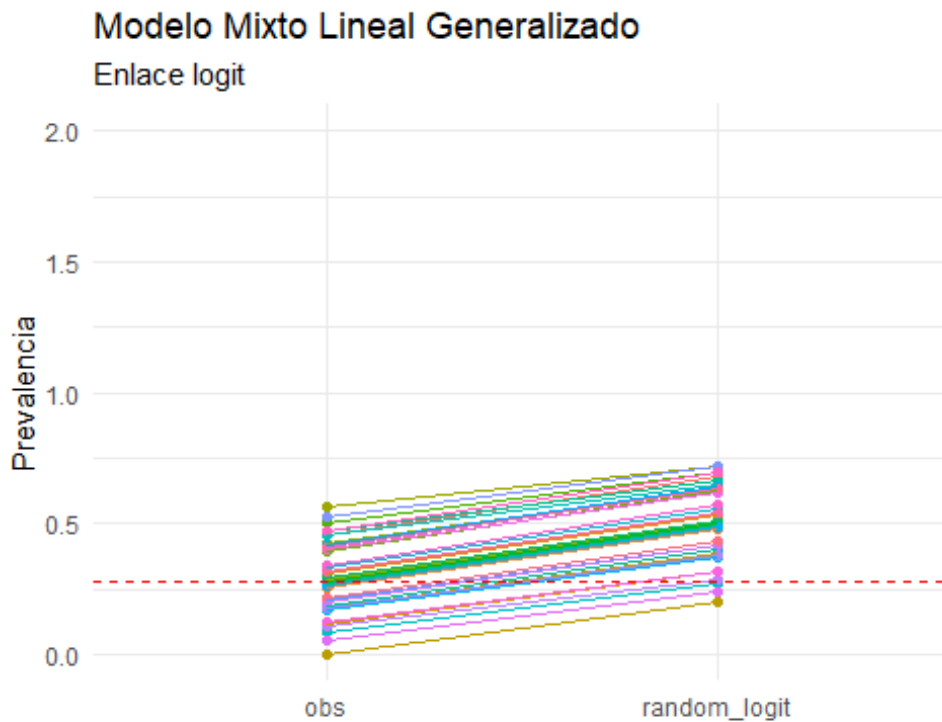
```
comp.random_1 <- tibble(ID,random_log, obs) |>  
  pivot_longer(cols = random_log:obs, names_to = "link", values_to = "  
Predictions")  
  
ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact  
or(ID)))+  
  geom_point()+  
  geom_line()+  
  geom_hline(yintercept = exp(summary(mod_logc)$coefficients[1]), line  
type = "dashed", color = "red")+  
  labs(y = "Prevalencia",  
       x = " ",  
       title = "Modelo Mixto Lineal Generalizado",  
       subtitle = "Enlace log")+  
  theme_minimal()+  
  theme(legend.position = "none")
```



Enlace logit

```
comp.random_1 <- tibble(ID,random_logit, obs) |>  
  pivot_longer(cols = random_logit:obs, names_to = "link", values_to = "  
Predictions")  
  
ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact  
or(ID)))+  
  geom_point()+
```

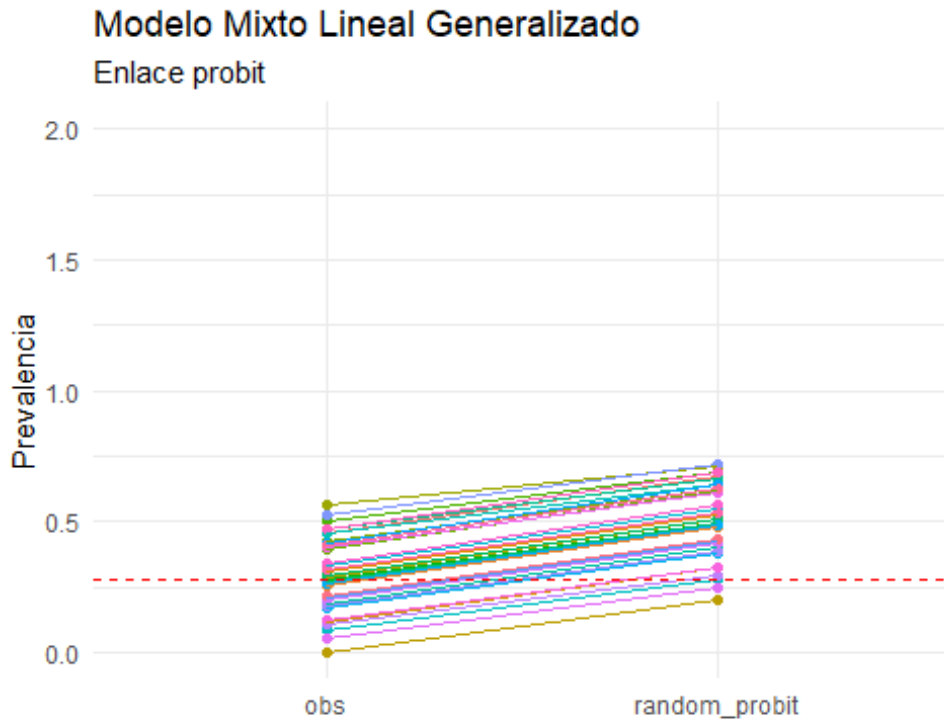
```
geom_line()+
ylim(c(0,2))+
geom_hline(yintercept = plogis(summary(mod_logitc)$coefficients[1]),
linetype = "dashed", color = "red")+
labs(y = "Prevalencia",
x = " ",
title = "Modelo Mixto Lineal Generalizado",
subtitle = "Enlace logit")+
theme_minimal()+
theme(legend.position = "none")
```



Enlace probit

```
comp.random_1 <- tibble(ID,random_probit, obs) |>
pivot_longer(cols = random_probit:obs, names_to = "link", values_to
= "Predictions")

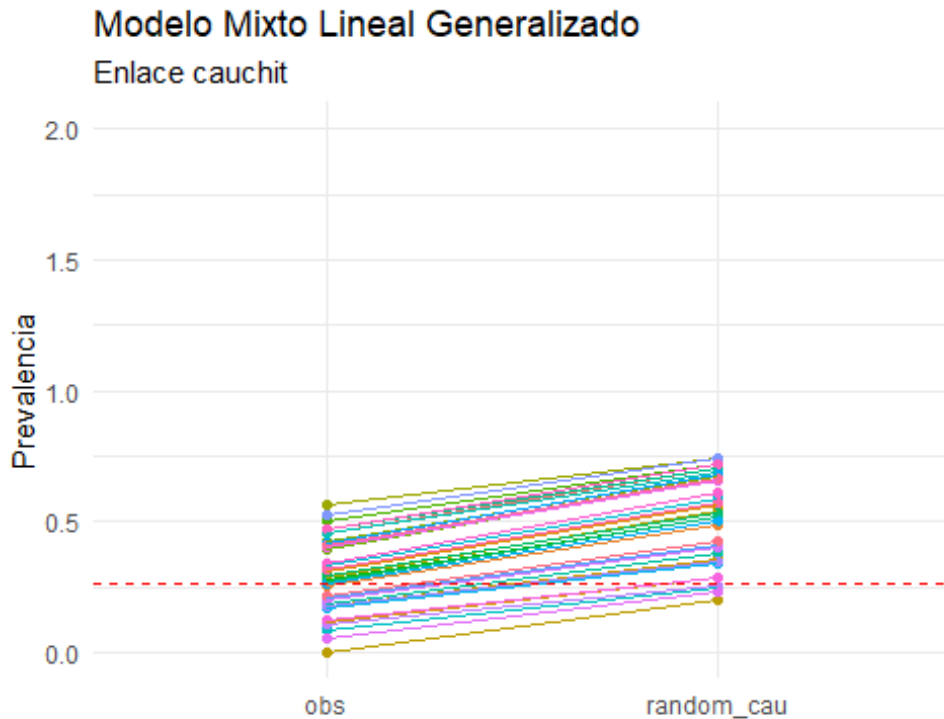
ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact
or(ID)))+
geom_point()+
geom_line()+
ylim(c(0,2))+
geom_hline(yintercept = pnorm(summary(mod_probit)$coefficients[1]),
linetype = "dashed", color = "red")+
labs(y = "Prevalencia",
x = " ",
title = "Modelo Mixto Lineal Generalizado",
subtitle = "Enlace probit")+
theme_minimal()+
theme(legend.position = "none")
```



Enlace cauchit

```
comp.random_1 <- tibble(ID,random_cau, obs) |>
  pivot_longer(cols = random_cau:obs, names_to = "link", values_to = "
Predictions")

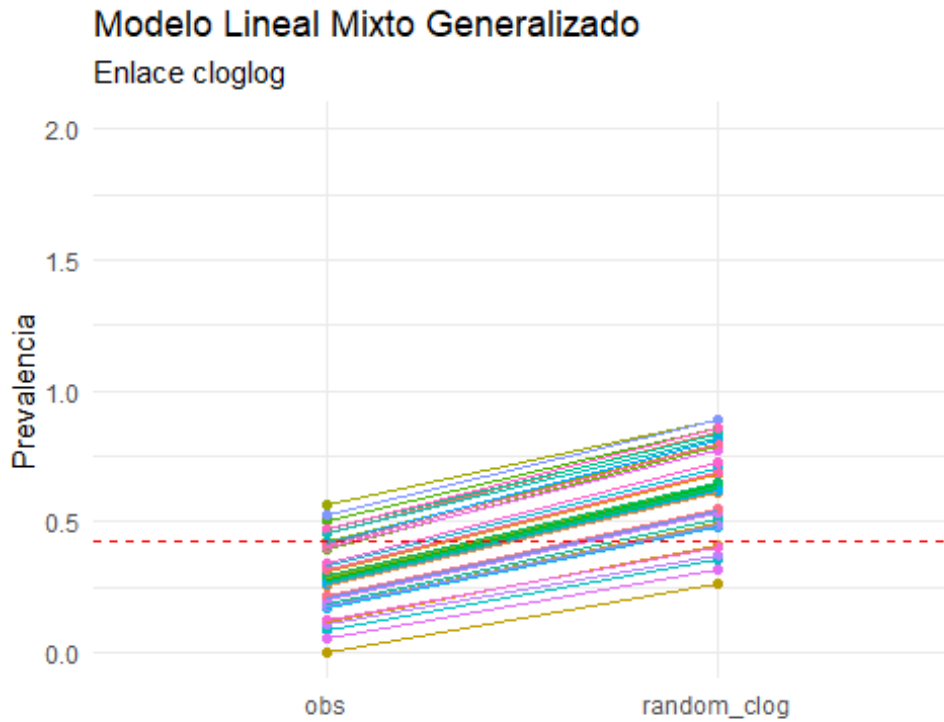
ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact
or(ID)))+
  geom_point()+
  geom_line()+
  ylim(c(0,2))+
  geom_hline(yintercept = pcauchy(summary(mod_cauchitc)$coefficients[1
]), linetype = "dashed", color = "red")+
  labs(y = "Prevalencia",
       x = " ",
       title = "Modelo Mixto Lineal Generalizado",
       subtitle = "Enlace cauchit")+
  theme_minimal()+
  theme(legend.position = "none")
```



Enlace cloglog

```
comp.random_1 <- tibble(ID,random_clog, obs) |>
  pivot_longer(cols = random_clog:obs, names_to = "link", values_to =
"Predictions")

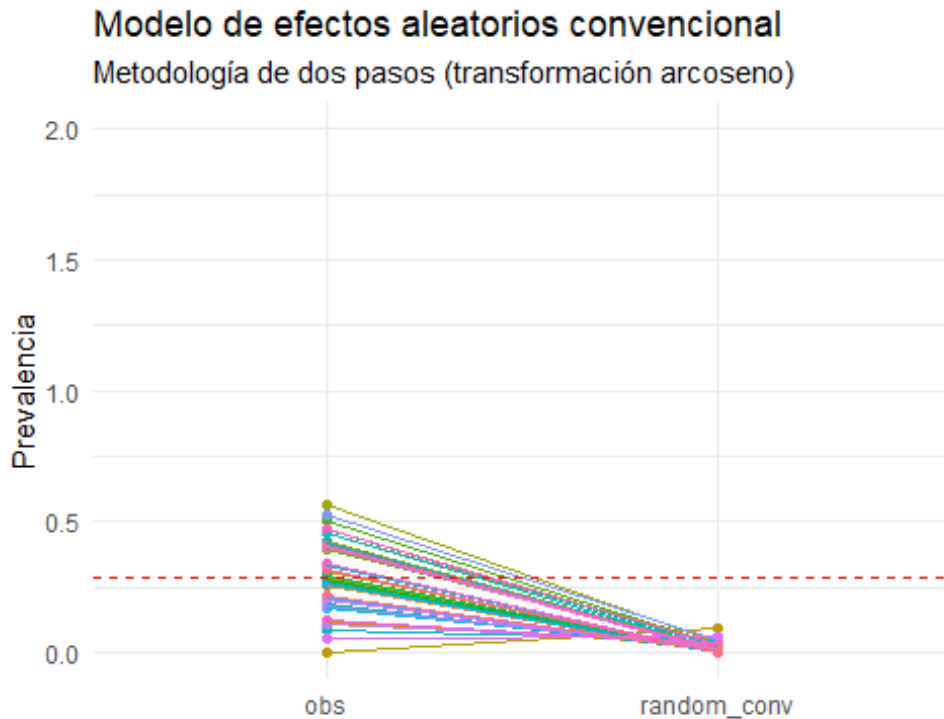
ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact
or(ID)))+
  geom_point()+
  geom_line()+
  ylim(c(0,2))+
  geom_hline(yintercept = (1 - exp(-exp(summary(mod_probit)$coefficien
ts[1]))), linetype = "dashed", color = "red")+
  labs(y = "Prevalencia",
       x = " ",
       title = "Modelo Lineal Mixto Generalizado",
       subtitle = "Enlace cloglog")+
  theme_minimal()+
  theme(legend.position = "none")
```



Convencional

```
comp.random_1 <- tibble(ID,random_conv, obs) |>
  pivot_longer(cols = random_conv:obs, names_to = "link", values_to =
"Predictions")

ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact
or(ID)))+
  geom_point()+
  geom_line()+
  ylim(c(0,2))+
  geom_hline(yintercept = (sin(mu.est))^2, linetype = "dashed", color
= "red")+
  labs(y = "Prevalencia",
       x = " ",
       title = "Modelo de efectos aleatorios convencional",
       subtitle = "Metodología de dos pasos (transformación arcoseno)"
)+
  theme_minimal()+
  theme(legend.position = "none")
```



##Tamaño muestral: análisis de subgrupos

```
n_comp <- abx |>
  pull(N)

quantile(n_comp, c(0.3333, 0.6667))

33.33% 66.67%
  51    82

abx_comp <- abx |>
  mutate(low = if_else(N<=51, 1,0),
         medium = if_else(N>51 & N <= 82, 1,0),
         high = if_else(N>82, 1,0), .after = N)
```

Modelo 1: enlace log

```
mod_log_comp <- glmer(cbind(total, N - total)~ medium + low +(1|ID), data = abx_comp, family = binomial(link = "log"))

sum <- conversion(mod_log_comp, "log");sum

$mu
[1] 0.2855202

$`CI 95%`
[1] 0.6019195 0.1354364

$tau2
[1] 0.1742845
```

```
$tau
[1] 0.417474

a <- summary(mod_log_comp); a

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( log )
Formula: cbind(total, N - total) ~ medium + low + (1 | ID)
Data: abx_comp

      AIC      BIC   logLik deviance df.resid
  280.0   286.5  -136.0   272.0     34

Scaled residuals:
   Min       1Q   Median       3Q      Max
-1.63511 -0.27939 -0.02385  0.24123  0.59530

Random effects:
 Groups Name      Variance Std.Dev.
  ID      (Intercept) 0.1743   0.4175
Number of obs: 38, groups: ID, 38

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2534     0.1288  -9.731  <2e-16 ***
medium        -0.3805     0.1923  -1.978  0.0479 *
low           0.1642     0.1802   0.911  0.3621
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) medium
medium -0.661
low    -0.712  0.477

tibble(Size = c("low", "medium", "high"),
        prev = c(a$coefficients[1]+a$coefficients[3], a$coefficients[1]
+a$coefficients[2], a$coefficients[1])) |>
  mutate(prev = exp(prev))

# A tibble: 3 × 2
  Size    prev
  <chr> <dbl>
1 low    0.336
2 medium 0.195
3 high   0.286
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ medium + low +ID, data = abx_comp,
family = binomial(link = "log"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]
```

```
test <- abx_comp |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(p)) |>
  mutate(vc = (1-p)/(N*p)) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)

Q <- sum(test$wyf);Q

[1] 233.6965

df <- length(abx_comp$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

p val = 2.14216825201191e-30

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.8417

glue("tau2 = {sum$tau2}")

tau2 = 0.174284531009938

glue("tau = {sum$tau}")

tau = 0.417473988423157

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.7916, 0.8797
```

Modelo 2: enlace logit

```
mod_logit_comp <- glmer(cbind(total, N - total)~ low + medium +(1|ID),
  data = abx_comp, family = binomial(link = "logit"))

sum <- conversion(mod_logit_comp, "logit"); sum
```

```
$mu
[1] 0.29264

$`CI 95%`
[1] 0.2031597 0.4016652

$tau2
[1] 0.3589586

$tau
[1] 0.5991315

summary(mod_logit_comp)

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: cbind(total, N - total) ~ low + medium + (1 | ID)
Data: abx_comp

      AIC      BIC   logLik deviance df.resid
  276.2   282.8  -134.1   268.2     34

Scaled residuals:
      Min       1Q   Median       3Q      Max
-1.35108 -0.30322 -0.04032  0.30377  0.81507

Random effects:
Groups Name          Variance Std.Dev.
ID      (Intercept) 0.359     0.5991
Number of obs: 38, groups: ID, 38

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8826     0.1838  -4.803 1.56e-06 ***
low           0.2470     0.2601   0.950  0.342
medium       -0.5233     0.2705  -1.935  0.053 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) low
low   -0.706
medium -0.676  0.479

t <- summary(mod_logit_comp)

tibble(Size = c("low", "medium", "high"),
        prev = c(t$coefficients[1]+t$coefficients[2], t$coefficients[1]
+t$coefficients[3], t$coefficients[1])) |>
  mutate(prev = plogis(prev))

# A tibble: 3 × 2
  Size  prev
<chr> <dbl>
```

```
1 low    0.346
2 medium 0.197
3 high   0.293
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ low + medium + ID, data = abx_comp
, family = binomial(link = "logit"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]
```

```
test <- abx_comp |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(n/(N-n))) |>
  mutate(vc = 1/(N*p*(1-p))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccaу))^2)) |>
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
```

```
[1] 209.5542
```

```
df <- length(abx_comp$ID)-1
```

```
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
```

```
p val = 5.66579134007613e-26
```

```
H2 <- Q/(length(test$ID)-1)
```

```
H <-sqrt(H2)
```

```
I2 <- (H2-1)/H2
```

```
glue("I2 = {round(I2,4)}")
```

```
I2 = 0.8234
```

```
glue("tau2 = {sum$tau2}")
```

```
tau2 = 0.358958554670303
```

```
glue("tau = {sum$tau}")
```

```
tau = 0.599131500315501
```

```
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
```

```
L <- exp(0.5*log(Q/df)-1.96*B)
```

```
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)
```

```
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
```

```
IC95% de I2 = 0.7656, 0.867
```

```
###Modelo 3: enlace probit
```

```
mod_probit_comp <- glmer(cbind(total, N - total)~ medium + low +(1|ID)  
, data = abx_comp, family = binomial(link = "probit"))  
sum <- conversion(mod_probit, "probit"); sum
```

```
$mu  
[1] 0.279117
```

```
$`CI 95%`  
[1] 0.2346941 0.3272721
```

```
$tau2  
[1] 0.1564142
```

```
$tau  
[1] 0.3954924
```

```
b <- summary(mod_probit_comp); b
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace  
Approximation) [glmerMod]  
Family: binomial ( probit )  
Formula: cbind(total, N - total) ~ medium + low + (1 | ID)  
Data: abx_comp
```

AIC	BIC	logLik	deviance	df.resid
275.6	282.1	-133.8	267.6	34

```
Scaled residuals:  
      Min       1Q   Median       3Q      Max  
-1.17602 -0.31653 -0.04717  0.30824  0.86521
```

```
Random effects:  
Groups Name      Variance Std.Dev.  
ID      (Intercept) 0.1259  0.3548  
Number of obs: 38, groups: ID, 38
```

```
Fixed effects:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept)  -0.5392     0.1088  -4.956  7.2e-07 ***  
medium        -0.3057     0.1592  -1.921  0.0547 .  
low           0.1482     0.1545   0.959  0.3376  
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:  
      (Intr) medium  
medium -0.682  
low    -0.704  0.480
```

```
tibble(Size = c("low", "medium", "high"),
       prev = c(b$coefficients[1]+b$coefficients[3], b$coefficients[1]
+b$coefficients[2], b$coefficients[1])) |>
  mutate(prev = pnorm(prev))

# A tibble: 3 × 2
  Size    prev
  <chr> <dbl>
1 low    0.348
2 medium 0.199
3 high   0.295
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ low + medium +ID, data = abx_comp,
family = binomial(link = "probit"))
scau <-summary(cau)
ccau <- scau$coefficients[1]

test <- abx_comp |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = qnorm(p)) |>
  mutate(derivada_cuadrada = 2 * pi * exp(yc^2)) |>
  mutate(var_p = p * (1 - p) / N) |>
  mutate(vc = derivada_cuadrada*var_p)|>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)

Q <- sum(test$wyf);Q

[1] 225.9395

df <- length(abx_comp$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

p val = 5.77159775347735e-29

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.8362

glue("tau2 = {sum$tau2}")

tau2 = 0.15641424359939

glue("tau = {sum$tau}")

tau = 0.395492406500289
```

```
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
```

```
L <- exp(0.5*log(Q/df)-1.96*B)
```

```
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)
```

```
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
```

```
IC95% de I2 = 0.7838, 0.8759
```

###Modelo 4: enlace cauchit

```
mod_cauchit_comp <- glmer(cbind(total, N - total)~ medium + low +(1|ID
), data = abx_comp, family = binomial(link = "cauchit"))
sum <- conversion(mod_cauchit_comp, "cauchit"); sum
```

```
$mu
```

```
[1] 0.2789365
```

```
$`CI 95%`
```

```
[1] 0.5929593 0.1497474
```

```
$tau2
```

```
[1] 0.4142617
```

```
$tau
```

```
[1] 0.6436317
```

```
c <- summary(mod_cauchit_comp); c
```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]

Family: binomial (cauchit)

Formula: cbind(total, N - total) ~ medium + low + (1 | ID)

Data: abx_comp

AIC	BIC	logLik	deviance	df.resid
283.2	289.8	-137.6	275.2	34

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.15986	-0.26486	-0.01013	0.21881	0.58110

Random effects:

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------

ID	(Intercept)	0.4143	0.6436
----	-------------	--------	--------

Number of obs: 38, groups: ID, 38

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.8329	0.2007	-4.150	3.32e-05 ***

```

medium      -0.5784      0.3063  -1.888    0.059 .
low         0.2593      0.2784   0.931    0.352
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
      (Intr) medium
medium -0.634
low    -0.716  0.466

tibble(Size = c("low", "medium", "high"),
        prev = c(c$coefficients[1]+c$coefficients[3], c$coefficients[1]
+c$coefficients[2], c$coefficients[1])) |>
  mutate(prev = pcauchy(prev))

# A tibble: 3 × 2
  Size   prev
<chr> <dbl>
1 low    0.334
2 medium 0.196
3 high   0.279

```

Análisis de la heterogeneidad

```

cau <- glm(cbind(total, N - total)~ low + medium +ID, data = abx_comp,
family = binomial(link = "cauchit"))
scau <-summary(cau)
ccau <- scau$coefficients[1]
test <- abx_comp |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = tan(pi*(p-0.5))) |>
  mutate(vc = (pi*p*(1-p))/(N*cos(pi*(p-0.5))*cos(pi*(p-0.5)))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)

Q <- sum(test$wyf);Q

[1] 1799.485

df <- length(abx_comp$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

p val = 0

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.9794

```

```
glue("tau2 = {sum$tau2}")
tau2 = 0.414261742522001
glue("tau = {sum$tau}")
tau = 0.64363168234791
df <- length(abx$ID)-1
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.9762, 0.9823

###Modelo 5: enlace cloglog

mod_cloglog_comp <- glmer(cbind(total, N - total)~ low + medium +(1|ID
), data = abx_comp, family = binomial(link = "cloglog"))
sum <- conversion(mod_cloglog_comp, "cloglog"); sum

$mu
[1] 0.2891735

$`CI 95%`
[1] 0.2047379 0.3986425

$tau2
[1] 0.2540745

$tau
[1] 0.5040581

d <- summary(mod_cloglog_comp); d

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( cloglog )
Formula: cbind(total, N - total) ~ low + medium + (1 | ID)
Data: abx_comp

      AIC      BIC   logLik deviance df.resid
  277.8   284.3  -134.9   269.8      34

Scaled residuals:
      Min       1Q   Median       3Q      Max
-1.48653 -0.29123 -0.03275  0.27704  0.70847
```

```

Random effects:
Groups Name          Variance Std.Dev.
ID      (Intercept) 0.2541   0.5041
Number of obs: 38, groups:  ID, 38

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.0749     0.1549  -6.942 3.88e-12 ***
low           0.2034     0.2178   0.934  0.3502
medium       -0.4493     0.2293  -1.959  0.0501 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
          (Intr) low
low      -0.710
medium  -0.670  0.479

tibble(Size = c("low", "medium", "high"),
        prev = c(d$coefficients[1]+d$coefficients[2], d$coefficients[1]
+d$coefficients[3], d$coefficients[1])) |>
  mutate(prev = (1 - exp(-exp(prev))))

# A tibble: 3 × 2
  Size  prev
  <chr> <dbl>
1 low   0.342
2 medium 0.196
3 high  0.289

```

Análisis de la heterogeneidad

```

cau <- glm(cbind(total, N - total)~ low + medium + ID, data = abx_comp
, family = binomial(link = "cloglog"))
scau <-summary(cau)
ccau <- scau$coefficients[1]

test <- abx_comp |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(-log(1-p))) |>
  mutate(vc = p/(N*(1-p)*(log(1-p)^2))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)

Q <- sum(test$wyf);Q

[1] 217.5345

df <- length(abx_comp$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

```

```
p_val = 2.0013102373249e-27
H2 <- Q/(length(test$ID)-1)
H <-sqrt(H2)
I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.8299
glue("tau2 = {sum$tau2}")
tau2 = 0.254074543049146
glue("tau = {sum$tau}")
tau = 0.504058075075825
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.7748, 0.8715
```

Modelo convencional

```
test1_comp <- escalc("PAS", xi = total, ni = N, data = abx_comp)
rslt_subgroup <- rma(yi = yi, vi = vi, mods = ~ medium + low, method =
"DL", data = test1_comp)
summary(rslt_subgroup)
```

Mixed-Effects Model (k = 38; tau² estimator: DL)

logLik	deviance	AIC	BIC	AICc
19.4904	103.3993	-30.9807	-24.4304	-29.7686

```
tau^2 (estimated amount of residual heterogeneity): 0.0186 (SE = 0
.0058)
tau (square root of estimated tau^2 value): 0.1362
I^2 (residual heterogeneity / unaccounted variability): 84.94%
H^2 (unaccounted variability / sampling variability): 6.64
R^2 (amount of heterogeneity accounted for): 11.21%
```

Test for Residual Heterogeneity:

```
QE(df = 35) = 232.4782, p-val < .0001
```

```
Test of Moderators (coefficients 2:3):
```

```
QM(df = 2) = 7.7572, p-val = 0.0207
```

```
Model Results:
```

	estimate	se	zval	pval	ci.lb	ci.ub	
intrcpt	0.5772	0.0415	13.8931	<.0001	0.4957	0.6586	***
medium	-0.1098	0.0600	-1.8317	0.0670	-0.2273	0.0077	.
low	0.0556	0.0591	0.9407	0.3469	-0.0603	0.1715	

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
prevalencia_large <- 0.5772
```

```
prevalencia_medium <- 0.5772 - 0.1098
```

```
prevalencia_low <- 0.5772 + 0.0556
```

```
tibble(Size = c("low", "medium", "high"),  
       prev = c(sin(prevalencia_low)^2, sin(prevalencia_medium)^2, sin(  
prevalencia_large)^2))
```

```
# A tibble: 3 × 2
```

	Size	prev
	<chr>	<dbl>
1	low	0.350
2	medium	0.203
3	high	0.298

```
B <- 0.5*((log(67.2914)-log(37))/(sqrt(2*67.2914)-sqrt(2*(37-1))))
```

```
L <- exp(0.5*log(67.2914/37)-1.96*B)
```

```
U <- exp(0.5*log(67.2914/37)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)
```

```
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
```

```
IC95% de I2 = 0.199, 0.6226
```

```
AIC
```

```
AIC(mod_log)
```

```
[1] 284.0112
```

```
AIC(mod_logit)
```

```
[1] 280.0582
```

```
AIC(mod_probit)
```

```
[1] 279.3212
AIC(mod_cauchit)
[1] 287.0299
AIC(mod_cloglog)
[1] 281.7269
AIC.rma(rs1t_subgroup)
[1] -30.98073
```

Estudios con baja heterogeneidad

```
set.seed(42)
n_estudios <- 38
N <- sample(100:500, size = n_estudios, replace = TRUE)
proporciones_individuales <- abs(rnorm(n_estudios, mean = 0.26, sd = 0.03))
casos <- pmax(0, round(N * proporciones_individuales))
datos <- data.frame(
  ID = 1:n_estudios,
  N = N,
  total = casos,
  prev = casos / N
)
datos
```

	ID	N	total	prev
1	1	148	44	0.2972973
2	2	420	133	0.3166667
3	3	252	62	0.2460317
4	4	173	44	0.2543353
5	5	327	68	0.2079511
6	6	245	67	0.2734694
7	7	221	53	0.2398190
8	8	148	41	0.2770270
9	9	227	64	0.2819383
10	10	402	117	0.2910448
11	11	123	30	0.2439024
12	12	426	117	0.2746479
13	13	455	95	0.2087912
14	14	188	44	0.2340426
15	15	264	62	0.2348485
16	16	209	39	0.1866029
17	17	119	31	0.2605042
18	18	469	125	0.2665245
19	19	466	116	0.2489270

```
20 20 486 137 0.2818930
21 21 396 94 0.2373737
22 22 188 41 0.2180851
23 23 382 104 0.2722513
24 24 208 49 0.2355769
25 25 104 32 0.3076923
26 26 311 77 0.2475884
27 27 447 125 0.2796421
28 28 459 124 0.2701525
29 29 358 85 0.2374302
30 30 413 127 0.3075061
31 31 397 111 0.2795970
32 32 123 32 0.2601626
33 33 257 69 0.2684825
34 34 398 112 0.2814070
35 35 498 131 0.2630522
36 36 413 70 0.1694915
37 37 235 63 0.2680851
38 38 391 97 0.2480818
```

Modelo 1: enlace log

```
mod_log <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos, family = binomial(link = "log"))
sum <- conversion(mod_log, "log");sum

$mu
[1] 0.2567469

$`CI 95%`
[1] 0.2462348 0.2677077

$tau2
[1] 0.007074564

$tau
[1] 0.08411043
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "log"))
scau <-summary(cau)
ccau <- scau$coefficients[1]

test <- datos |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(p)) |>
  mutate(vc = (1-p)/(N*p)) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
[1] 61.77231
df <- length(datos$ID)-1
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
p val = 0.00649476686446005
H2 <- Q/(length(test$ID)-1)
H <-sqrt(H2)
I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.401
glue("tau2 = {sum$tau2}")
tau2 = 0.00707456430033374
glue("tau = {sum$tau}")
tau = 0.0841104292007462
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.1224, 0.5912
data_forest <- tibble(datos) |>
  mutate (LL = prev - (1.96 * (sqrt(prev * (1 - prev) / N))),
    UL = prev + (1.96 * (sqrt(prev * (1 - prev) / N)))) |>
  add_row("prev" = sum$mu,
    ID = 39,
    UL = sum$'CI 95%' [2],
    LL = sum$'CI 95%' [1])

plot_data <- data_forest |>
  mutate(CI_texto = as.character(glue("{round(LL,2)}, {round(UL,2)}"))
  )

library(gridExtra)

plot1 <- ggplot(plot_data, aes(y = ID, x = prev)) +
```

```
geom_point(aes(color = ifelse(ID == 39, "red", "black")),
           shape = 15,
           size = 3.5) +
geom_errorbarh(aes(xmin = LL,
                  xmax = UL,
                  color = ifelse(ID == 39, "red", "black")),
              height = 0.25) +
geom_vline(xintercept = sum$mu,
           linetype = "dashed",
           color = "black",
           linewidth = 0.8,
           alpha = 0.8) +
scale_y_continuous(name = "",
                  breaks = plot_data$ID,
                  labels = plot_data$study,
                  trans = "reverse") +
xlab("Prevalencia") +
ylab(" ") +
theme_bw() +
coord_cartesian(xlim = c(min(plot_data$LL) - 0.05, max(plot_data$UL)
+ 0.05)) +
scale_color_identity() +
theme(legend.position = "none",
      panel.border = element_blank(),
      panel.background = element_blank(),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      axis.line.x = element_line(colour = "black"),
      axis.line.y = element_blank(),
      axis.ticks.y = element_blank(),
      axis.text.y = element_text(size = 12, colour = "black", hjust
= 0),
      axis.text.x = element_text(size = 12, colour = "black"),
      axis.title.x = element_text(size = 12, colour = "black"))
```

Warning: Unknown or uninitialised column: `study`.

```
table_base <- ggplot(plot_data, aes(y=study)) +
  ylab(NULL) + xlab(" ") +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        axis.text.x = element_text(color="white", hjust = -3, size = 2
5),
        axis.line = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y = element_blank(),
        legend.position = "none",
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.background = element_blank())
```

```
tab1 <- table_base +
```

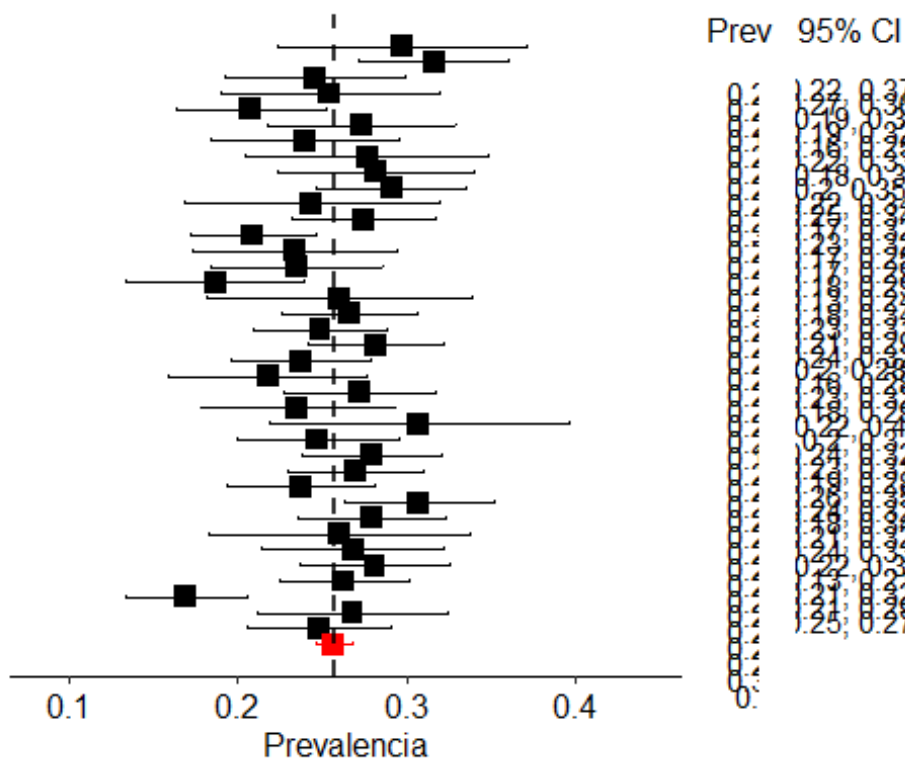
```
geom_text(aes(y = ID), x = 1, label = as.character(round(plot_data$rev,2)), size = 4) +
  ggtitle("Prev")

tab2 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = plot_data$CI_texto), size = 4) +
  ggtitle("95% CI")

lay <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3), nrow = 1)

grid.arrange(plot1, tab1, tab2, layout_matrix = lay)

Warning: Use of `plot_data$CI_texto` is discouraged.
i Use `CI_texto` instead.
```



Modelo 2: enlace logit

```
mod_logit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos, family = binomial(link = "logit"))

sum <- conversion(mod_logit, "logit"); sum

$mu
[1] 0.2568887

$`CI 95%`
[1] 0.2462477 0.2678260

$tau2
[1] 0.01315102
```

```
$tau  
[1] 0.1146779
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "logit"))  
scau <-summary(cau)  
ccau <- scau$coefficients[1]
```

```
test <- datos |>  
  mutate(n = if_else(total==0, 0.5, total)) |>  
  mutate(p = n/N) |>  
  mutate(yc = log(n/(N-n))) |>  
  mutate(vc = 1/(N*p*(1-p))) |>  
  mutate(w = 1/vc) |>  
  mutate(yf = ((yc-(ccau))^2)) |>  
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
```

```
[1] 64.3056
```

```
df <- length(datos$ID)-1
```

```
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
```

```
p val = 0.00355322422366149
```

```
H2 <- Q/(length(test$ID)-1)
```

```
H <-sqrt(H2)
```

```
I2 <- (H2-1)/H2
```

```
glue("I2 = {round(I2,4)}")
```

```
I2 = 0.4246
```

```
glue("tau2 = {sum$tau2}")
```

```
tau2 = 0.0131510174457758
```

```
glue("tau = {sum$tau}")
```

```
tau = 0.114677885600388
```

```
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
```

```
L <- exp(0.5*log(Q/df)-1.96*B)
```

```
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)
```

```
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
```

```
IC95% de I2 = 0.1591, 0.6063
```

```
###Modelo 3: enlace probit
```

```
mod_probit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos,  
family = binomial(link = "probit"))  
sum <- conversion(mod_probit, "probit"); sum
```

```
$mu  
[1] 0.2569428
```

```
$`CI 95%`  
[1] 0.2462555 0.2678666
```

```
$tau2  
[1] 0.004654692
```

```
$tau  
[1] 0.06822531
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "probit"))  
scau <-summary(cau)  
ccaу <- scau$coefficients[1]
```

```
test <- datos |>  
  mutate(n = if_else(total==0, 0.5, total)) |>  
  mutate(p = n/N) |>  
  mutate(yc = qnorm(p)) |>  
  mutate(derivada_cuadrada = 2 * pi * exp(yc^2)) |>  
  mutate(var_p = p * (1 - p) / N) |>  
  mutate(vc = derivada_cuadrada*var_p)|>  
  mutate(w = 1/vc) |>  
  mutate(yf = ((yc-(ccaу))^2)) |>  
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
```

```
[1] 66.16746
```

```
df <- length(datos$ID)-1
```

```
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
```

```
p val = 0.00224716244263084
```

```
H2 <- Q/(length(test$ID)-1)
```

```
H <-sqrt(H2)
```

```
I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.4408

glue("tau2 = {sum$tau2}")

tau2 = 0.00465469228168296

glue("tau = {sum$tau}")

tau = 0.0682253052883089

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

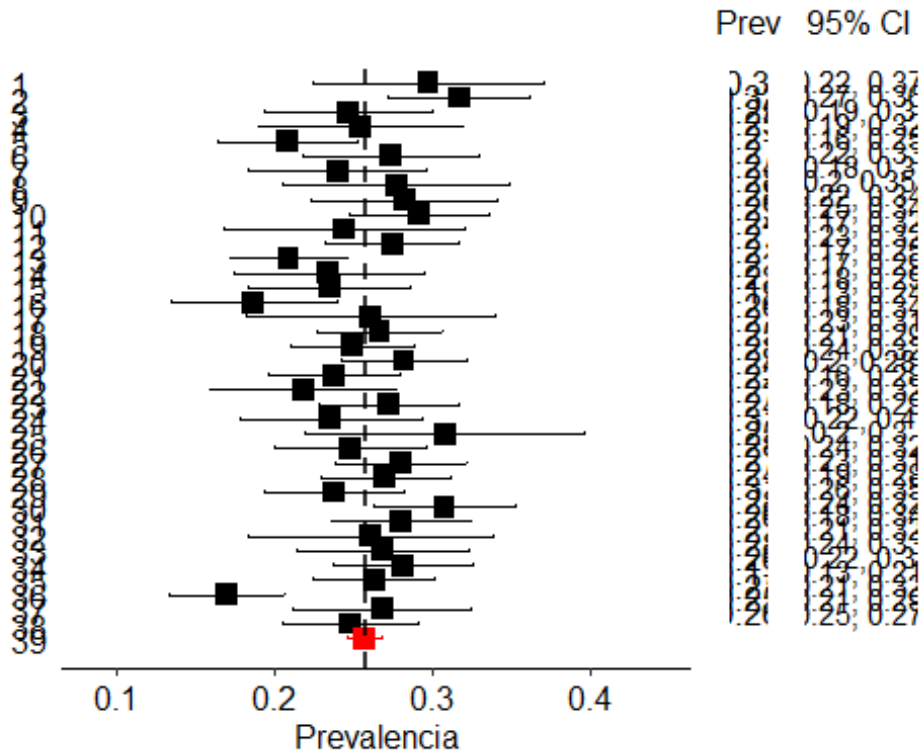
IC95% de I2 = 0.1844, 0.6166

data_forest <- datos |>
  mutate (
    LL = prev - (1.96 * (sqrt(prev * (1 - prev) / N))),
    UL = prev + (1.96 * (sqrt(prev * (1 - prev) / N))) |>
  add_row(prev = sum$mu,
    ID = 39,
    UL = sum$'CI 95%' [2],
    LL = sum$'CI 95%' [1]
  )
plot_data <- data_forest |>
  mutate(CI_texto = as.character(glue("{round(LL,2)}, {round(UL,2)}"))
)

library(gridExtra)

plot1 <- ggplot(plot_data, aes(y = ID, x = prev)) +
  geom_point(aes(color = ifelse(ID == 39, "red", "black")),
    shape = 15,
    size = 3.5) +
  geom_errorbarh(aes(xmin = LL,
    xmax = UL,
    color = ifelse(ID == 39, "red", "black")),
    height = 0.25) +
  geom_vline(xintercept = sum$mu,
    linetype = "dashed",
    color = "black",
    linewidth = 0.8,
    alpha = 0.8) +
  scale_y_continuous(name = "",
```

```
breaks = plot_data$ID,  
labels = plot_data$ID,  
trans = "reverse") +  
xlab("Prevalencia") +  
ylab(" ") +  
labs(title = "hi")+  
theme_bw() +  
coord_cartesian(xlim = c(min(plot_data$LL) - 0.05, max(plot_data$UL)  
+ 0.05)) +  
scale_color_identity() +  
theme(plot.title = element_text(hjust = 0.5, size=12, colour = "white"),  
legend.position = "none",  
panel.border = element_blank(),  
panel.background = element_blank(),  
panel.grid.major = element_blank(),  
panel.grid.minor = element_blank(),  
axis.line.x = element_line(colour = "black"),  
axis.line.y = element_blank(),  
axis.ticks.y = element_blank(),  
axis.text.y = element_text(size = 12, colour = "black", hjust  
= 0),  
axis.text.x = element_text(size = 12, colour = "black"),  
axis.title.x = element_text(size = 12, colour = "black"))  
  
table_base <- ggplot(plot_data, aes(y=ID)) +  
ylab(NULL) + xlab("") +  
theme(plot.title = element_text(hjust = 0.5, size=12),  
axis.text.x = element_text(color="white", hjust = -3, size = 2  
5),  
axis.line = element_blank(),  
axis.text.y = element_blank(),  
axis.ticks = element_blank(),  
axis.title.y = element_blank(),  
legend.position = "none",  
panel.background = element_blank(),  
panel.border = element_blank(),  
panel.grid.major = element_blank(),  
panel.grid.minor = element_blank(),  
plot.background = element_blank())  
  
tab1 <- table_base +  
geom_text(aes(y = rev(ID), x = 1, label = as.character(round(prev,2)  
) , size = 4)) +  
ggtitle("Prev")  
  
tab2 <- table_base +  
geom_text(aes(y = rev(ID), x = 1, label = CI_texto), size = 4) +  
ggtitle("95% CI")  
  
lay <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3), nrow = 1)  
grid.arrange(plot1, tab1, tab2, layout_matrix = lay)
```



###Modelo 4: enlace cauchit

```
mod_cauchit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos,
family = binomial(link = "cauchit"))
sum <- conversion(mod_cauchit, "cauchit"); sum
```

```
$mu
[1] 0.2565943
```

```
$`CI 95%`
[1] 0.2462729 0.2676003
```

```
$tau2
[1] 0.01610451
```

```
$tau
[1] 0.1269035
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(
link = "cauchit"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]
test <- datos |>
mutate(n = if_else(total==0, 0.5, total)) |>
mutate(p = n/N) |>
mutate(yc = tan(pi*(p-0.5))) |>
mutate(vc = (pi*p*(1-p))/(N*cos(pi*(p-0.5))*cos(pi*(p-0.5)))) |>
mutate(w = 1/vc) |>
mutate(yf = ((yc-(ccaу))^2)) |>
```

```
mutate(wyf = w*yf)

Q <- sum(test$wyf);Q
[1] 430.8669

df <- length(datos$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
p val = 1.35622253268471e-68

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.9141

glue("tau2 = {sum$tau2}")
tau2 = 0.0161045068000268

glue("tau = {sum$tau}")
tau = 0.12690353344185

df <- length(abx$ID)-1

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

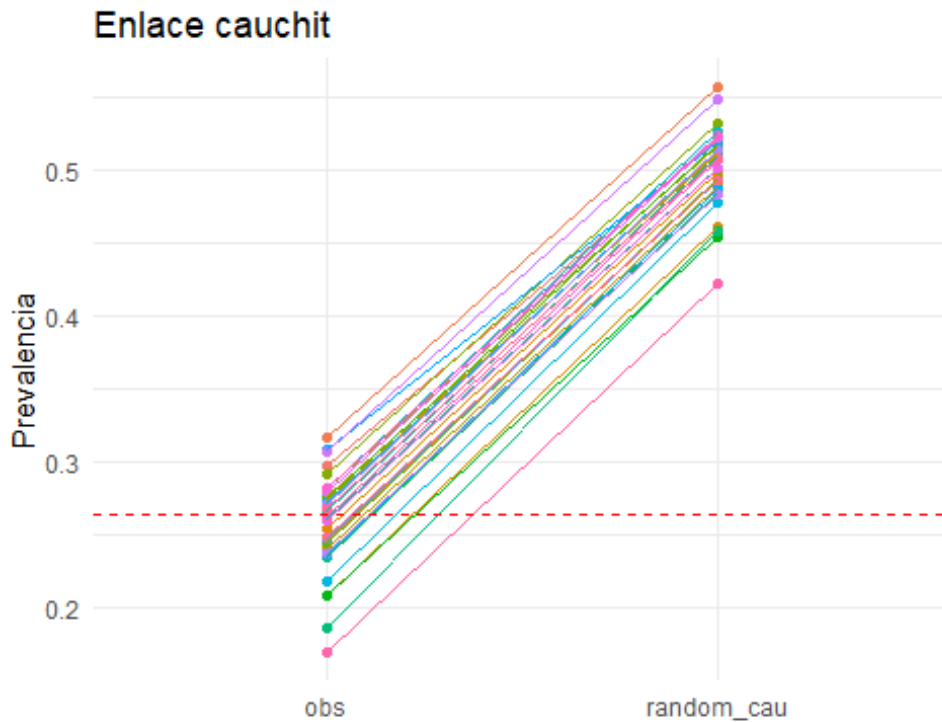
LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.8919, 0.9318

random_cau <-pcauchy(as.data.frame(ranef(mod_cauchit))$condval)
obs <- datos$prev
ID <- datos$ID
comp.random_1 <- tibble(ID,random_cau, obs) |>
  pivot_longer(cols = random_cau:obs, names_to = "link", values_to = "
Predictions")

ggplot(comp.random_1, aes(link, Predictions, group = ID, colour = fact
or(ID)))+
  geom_point()+
  geom_line()+
  geom_hline(yintercept = pcauchy(summary(mod_cauchitc))$coefficients[1
```

```
]), linetype = "dashed", color = "red")+  
  labs(y = "Prevalencia",  
       x = " ",  
       title = "Enlace cauchit")+  
  theme_minimal()+  
  theme(legend.position = "none")
```



###Modelo 5: enlace cloglog

```
mod_cloglog <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos,  
family = binomial(link = "cloglog"))  
sum <- conversion(mod_cloglog, "cloglog"); sum
```

```
$mu  
[1] 0.2568192
```

```
$`CI 95%`  
[1] 0.2462389 0.2677687
```

```
$tau2  
[1] 0.009726679
```

```
$tau  
[1] 0.09862393
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "cloglog"))  
scau <-summary(cau)  
ccaу <- scau$coefficients[1]
```

```
test <- datos |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(-log(1-p))) |>
  mutate(vc = p/(N*(1-p)*(log(1-p)^2))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)

Q <- sum(test$wyf);Q

[1] 63.02276

df <- length(datos$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

p val = 0.00483695298824104

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.4129

glue("tau2 = {sum$tau2}")

tau2 = 0.00972667863344718

glue("tau = {sum$tau}")

tau = 0.0986239252587686

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.1409, 0.5988
```

Modelo convencional

```
test1 <-escalr("PAS",xi = total, ni = N, data = datos, add = 0.5)

rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt
```

Random-Effects Model (k = 38; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.0007 (SE = 0.0003)
tau (square root of estimated tau^2 value): 0.0258
I^2 (total heterogeneity / total variability): 45.02%
H^2 (total variability / sampling variability): 1.82

Test for Heterogeneity:
Q(df = 37) = 67.2914, p-val = 0.0017

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
0.5314	0.0064	83.5087	<.0001	0.5189	0.5439	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)
```

```
(sin(mu.est))^2
```

```
[1] 0.2567909
```

```
(sin(mu.ci))^2
```

```
[1] 0.2459704 0.2677627
```

```
B <- 0.5*((log(67.2914)-log(37))/(sqrt(2*67.2914)-sqrt(2*(37-1))))
```

```
L <- exp(0.5*log(67.2914/37)-1.96*B)
```

```
U <- exp(0.5*log(67.2914/37)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)
```

```
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
```

```
IC95% de I2 = 0.199, 0.6226
```

AIC

```
AIC(mod_log)
```

```
[1] 282.2502
```

```
AIC(mod_logit)
```

```
[1] 281.9806
```

```
AIC(mod_probit)
[1] 281.8699
AIC(mod_cauchit)
[1] 282.6871
AIC(mod_cloglog)
[1] 282.1094
AIC.rma(rs1t)
[1] -136.1543
```

Estudios con alta heterogeneidad

```
set.seed(42)
n_estudios <- 38
N <- sample(100:500, size = n_estudios, replace = TRUE)
proporciones_individuales <- abs(rnorm(n_estudios, mean = 0.26, sd = 0.35))
casos <- pmax(2, round(N * proporciones_individuales))
datos <- data.frame(
  ID = 1:n_estudios,
  N = N,
  total = casos,
  prev = casos / N
)
datos
```

	ID	N	total	prev
1	1	148	101	0.682432432
2	2	420	388	0.923809524
3	3	252	28	0.111111111
4	4	173	29	0.167630058
5	5	327	117	0.357798165
6	6	245	103	0.420408163
7	7	221	8	0.036199095
8	8	148	62	0.418918919
9	9	227	115	0.506607930
10	10	402	250	0.621890547
11	11	123	6	0.048780488
12	12	426	186	0.436619718
13	13	455	155	0.340659341
14	14	188	3	0.015957447
15	15	264	10	0.037878788
16	16	209	122	0.583732057
17	17	119	32	0.268907563

```
18 18 469 156 0.332622601
19 19 466 62 0.133047210
20 20 486 255 0.524691358
21 21 396 2 0.005050505
22 22 188 41 0.218085106
23 23 382 157 0.410994764
24 24 208 5 0.024038462
25 25 104 80 0.769230769
26 26 311 34 0.109324759
27 27 447 219 0.489932886
28 28 459 171 0.372549020
29 29 358 5 0.013966480
30 30 413 335 0.811138015
31 31 397 193 0.486146096
32 32 123 36 0.292682927
33 33 257 92 0.357976654
34 34 398 198 0.497487437
35 35 498 145 0.291164659
36 36 413 325 0.786924939
37 37 235 85 0.361702128
38 38 391 51 0.130434783
```

Modelo 1: enlace log

```
mod_log <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos, family = binomial(link = "log"))

sum <- conversion(mod_log, "log");sum

$mu
[1] 0.226405

$`CI 95%`
[1] 0.1549472 0.3308174

$tau2
[1] 1.386023

$tau
[1] 1.177295
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "log"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]

test <- datos |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(p)) |>
  mutate(vc = (1-p)/(N*p)) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccaу))^2)) |>
```

```
mutate(wyf = w*yf)

Q <- sum(test$wyf);Q
[1] 6712.992

df <- length(datos$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

p val = 0

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.9945

glue("tau2 = {sum$tau2}")

tau2 = 1.38602304183227

glue("tau = {sum$tau}")

tau = 1.17729479818449

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.9939, 0.995
```

Modelo 2: enlace logit

```
mod_logit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos, family = binomial(link = "logit"))

sum <- conversion(mod_logit, "logit"); sum

$mu
[1] 0.2772008

$`CI 95%`
[1] 0.1846509 0.3937374

$tau2
```

```
[1] 2.694034
```

```
$tau
```

```
[1] 1.641351
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "logit"))
scau <-summary(cau)
ccau <- scau$coefficients[1]
```

```
test <- datos |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(n/(N-n))) |>
  mutate(vc = 1/(N*p*(1-p))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccau))^2)) |>
  mutate(wyf = w*yf)
```

```
Q <- sum(test$wyf);Q
```

```
[1] 1858.921
```

```
df <- length(datos$ID)-1
```

```
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
```

```
p val = 0
```

```
H2 <- Q/(length(test$ID)-1)
```

```
H <-sqrt(H2)
```

```
I2 <- (H2-1)/H2
```

```
glue("I2 = {round(I2,4)}")
```

```
I2 = 0.9801
```

```
glue("tau2 = {sum$tau2}")
```

```
tau2 = 2.69403364269673
```

```
glue("tau = {sum$tau}")
```

```
tau = 1.64135116373576
```

```
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))
```

```
L <- exp(0.5*log(Q/df)-1.96*B)
```

```
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
```

```
UL <- ((U^2)-1)/(U^2)
```

```
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.977, 0.9828

data_forest <- datos |>
  mutate (
    LL = prev - (1.96 * (sqrt(prev * (1 - prev) / N))),
    UL = prev + (1.96 * (sqrt(prev * (1 - prev) / N))) |>
  add_row(prev = sum$mu,
    ID = 39,
    UL = sum$'CI 95%'[2],
    LL = sum$'CI 95%'[1]
  )
plot_data <- data_forest |>
  mutate(CI_texto = as.character(glue("{round(LL,2)}, {round(UL,2)}"))
)

library(gridExtra)

plot1 <- ggplot(plot_data, aes(y = ID, x = prev)) +
  geom_point(aes(color = ifelse(ID == 39, "red", "black")),
    shape = 15,
    size = 3.5) +
  geom_errorbarh(aes(xmin = LL,
    xmax = UL,
    color = ifelse(ID == 39, "red", "black")),
    height = 0.25) +
  geom_vline(xintercept = sum$mu,
    linetype = "dashed",
    color = "black",
    linewidth = 0.8,
    alpha = 0.8) +
  scale_y_continuous(name = "",
    breaks = plot_data$ID,
    labels = plot_data$ID,
    trans = "reverse") +
  xlab("Prevalencia") +
  ylab(" ") +
  labs(title = "hi")+
  theme_bw() +
  coord_cartesian(xlim = c(min(plot_data$LL) - 0.05, max(plot_data$UL)
+ 0.05)) +
  scale_color_identity() +
  theme(plot.title = element_text(hjust = 0.5, size=12, colour = "white"),
    legend.position = "none",
    panel.border = element_blank(),
    panel.background = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.line.x = element_line(colour = "black"),
    axis.line.y = element_blank(),
    axis.ticks.y = element_blank(),
```

```
axis.text.y = element_text(size = 12, colour = "black", hjust
= 0),
axis.text.x = element_text(size = 12, colour = "black"),
axis.title.x = element_text(size = 12, colour = "black"))

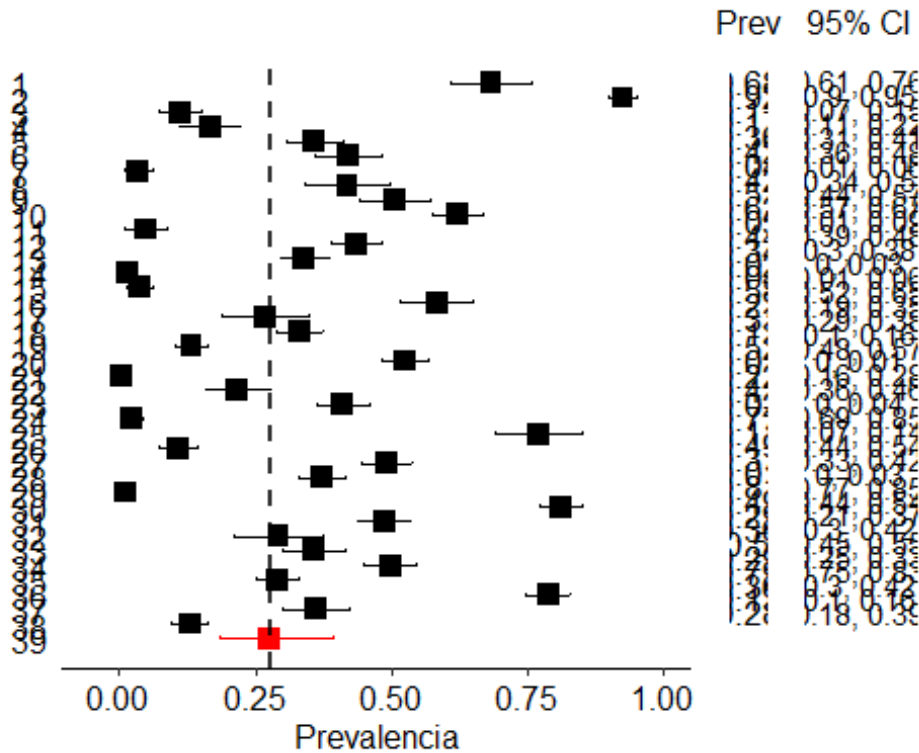
table_base <- ggplot(plot_data, aes(y=ID)) +
  ylab(NULL) + xlab("") +
  theme(plot.title = element_text(hjust = 0.5, size=12),
        axis.text.x = element_text(color="white", hjust = -3, size = 2
5),
        axis.line = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        axis.title.y = element_blank(),
        legend.position = "none",
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.background = element_blank())

tab1 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = as.character(round(prev,2)
), size = 4)) +
  ggtitle("Prev")

tab2 <- table_base +
  geom_text(aes(y = rev(ID), x = 1, label = CI_texto), size = 4) +
  ggtitle("95% CI")

lay <- matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3), nrow = 1)

grid.arrange(plot1, tab1, tab2, layout_matrix = lay)
```



###Modelo 3: enlace probit

```
mod_probit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos,
family = binomial(link = "probit"))
sum <- conversion(mod_probit, "probit"); sum
```

```
$mu
[1] 0.2967703

$`CI 95%`
[1] 0.2054978 0.4031192

$tau2
[1] 0.8108568

$tau
[1] 0.9004759
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "probit"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]

test <- datos |>
mutate(n = if_else(total==0, 0.5, total)) |>
mutate(p = n/N) |>
mutate(yc = qnorm(p)) |>
mutate(derivada_cuadrada = 2 * pi * exp(yc^2)) |>
mutate(var_p = p * (1 - p) / N) |>
```

```
mutate(vc = derivada_cuadrada*var_p)|>
mutate(w = 1/vc) |>
mutate(yf = ((yc-(ccau))^2)) |>
mutate(wyf = w*yf)

Q <- sum(test$wyf);Q
[1] 2339.844

df <- length(datos$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")
p val = 0

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")
I2 = 0.9842

glue("tau2 = {sum$tau2}")
tau2 = 0.810856843010066

glue("tau = {sum$tau}")
tau = 0.900475898072828

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")
IC95% de I2 = 0.9819, 0.9862

###Modelo 4: enlace cauchit

mod_cauchit <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos,
family = binomial(link = "cauchit"))
sum <- conversion(mod_cauchit, "cauchit"); sum

$mu
[1] 0.1573636

$`CI 95%`
[1] 0.1004865 0.3168245
```

```
$tau2  
[1] 13.0879
```

```
$tau  
[1] 3.61772
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "cauchit"))  
scau <-summary(cau)  
ccau <- scau$coefficients[1]  
test <- datos |>  
  mutate(n = if_else(total==0, 0.5, total)) |>  
  mutate(p = n/N) |>  
  mutate(yc = tan(pi*(p-0.5))) |>  
  mutate(vc = (pi*p*(1-p))/(N*cos(pi*(p-0.5))*cos(pi*(p-0.5)))) |>  
  mutate(w = 1/vc) |>  
  mutate(yf = ((yc-(ccau))^2)) |>  
  mutate(wyf = w*yf)  
  
Q <- sum(test$wyf);Q  
  
[1] 51770.86  
  
df <- length(datos$ID)-1  
  
glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")  
  
p val = 0  
  
H2 <- Q/(length(test$ID)-1)  
  
H <-sqrt(H2)  
  
I2 <- (H2-1)/H2  
glue("I2 = {round(I2,4)}")  
  
I2 = 0.9993  
  
glue("tau2 = {sum$tau2}")  
  
tau2 = 13.087900214527  
  
glue("tau = {sum$tau}")  
  
tau = 3.61772030628778  
  
df <- length(abx$ID)-1  
  
B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))  
  
L <- exp(0.5*log(Q/df)-1.96*B)  
U <- exp(0.5*log(Q/df)+1.96*B)
```

```
LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.9993, 0.9993
```

###Modelo 5: enlace cloglog

```
mod_cloglog <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = datos,
family = binomial(link = "cloglog"))
sum <- conversion(mod_cloglog, "cloglog"); sum

$mu
[1] 0.2502184

$`CI 95%`
[1] 0.1690306 0.3610121

$tau2
[1] 1.887329

$tau
[1] 1.373801
```

Análisis de la heterogeneidad

```
cau <- glm(cbind(total, N - total)~ ID, data = datos, family = binomial(link = "cloglog"))
scau <-summary(cau)
ccaу <- scau$coefficients[1]

test <- datos |>
  mutate(n = if_else(total==0, 0.5, total)) |>
  mutate(p = n/N) |>
  mutate(yc = log(-log(1-p))) |>
  mutate(vc = p/(N*(1-p)*(log(1-p)^2))) |>
  mutate(w = 1/vc) |>
  mutate(yf = ((yc-(ccaу))^2)) |>
  mutate(wyf = w*yf)

Q <- sum(test$wyf);Q

[1] 2515.957

df <- length(datos$ID)-1

glue("p val = {pchisq(Q, df, lower.tail = FALSE)}")

p val = 0

H2 <- Q/(length(test$ID)-1)

H <-sqrt(H2)
```

```

I2 <- (H2-1)/H2
glue("I2 = {round(I2,4)}")

I2 = 0.9853

glue("tau2 = {sum$tau2}")

tau2 = 1.88732906688446

glue("tau = {sum$tau}")

tau = 1.37380095606476

B <- 0.5*((log(Q)-log(df))/(sqrt(2*Q)-sqrt(2*(df-1))))

L <- exp(0.5*log(Q/df)-1.96*B)
U <- exp(0.5*log(Q/df)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.9832, 0.9871

```

Modelo convencional

```

test1 <- escalc("PAS",xi = total, ni = N, data = datos, add = 0.5)
rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt

Random-Effects Model (k = 38; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.0916 (SE = 0.0231)
tau (square root of estimated tau^2 value):      0.3026
I^2 (total heterogeneity / total variability):   99.12%
H^2 (total variability / sampling variability):  113.74

Test for Heterogeneity:
Q(df = 37) = 4208.5547, p-val < .0001

Model Results:

estimate      se      zval      pval      ci.lb      ci.ub      ***
  0.6054  0.0494  12.2661  <.0001  0.5087  0.7022

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,

```

```
mu.est + qnorm(1 - 0.05/2) * mu.se)

(sin(mu.est))^2
[1] 0.3238785

(sin(mu.ci))^2
[1] 0.2371909 0.4171383

B <- 0.5*((log(4208.5547)-log(37))/(sqrt(2*4208.5547)-sqrt(2*(37-1))))

L <- exp(0.5*log(4208.5547/37)-1.96*B)
U <- exp(0.5*log(4208.5547/37)+1.96*B)

LL <- ((L^2)-1)/(L^2)
UL <- ((U^2)-1)/(U^2)

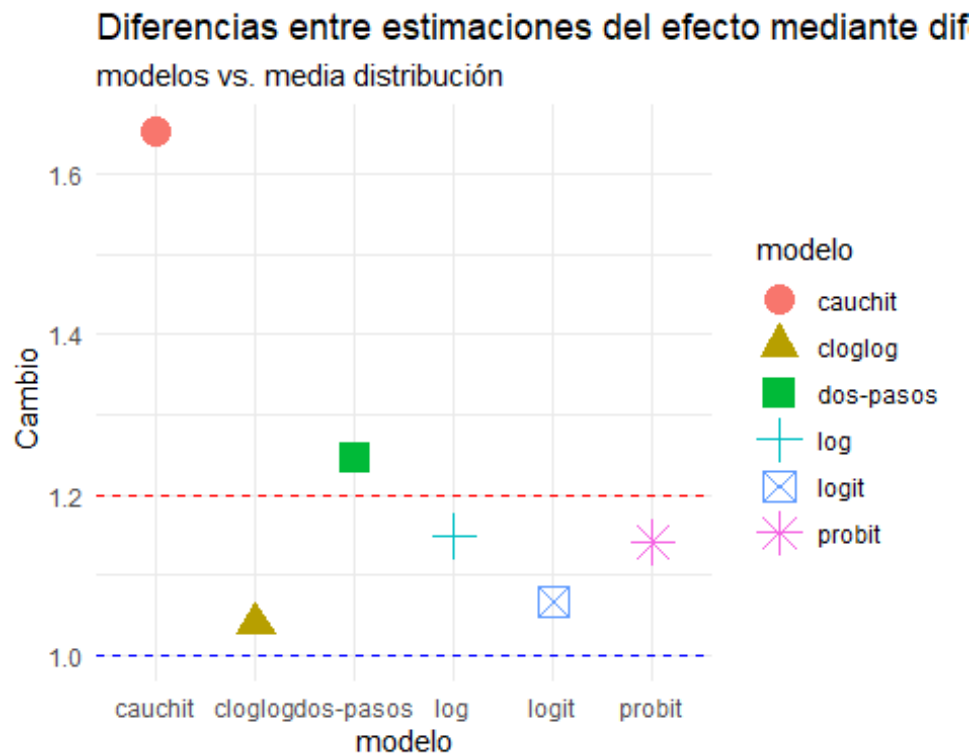
glue("IC95% de I2 = {round(LL,4)}, {round(UL ,4)}")

IC95% de I2 = 0.9902, 0.9921
```

Comparaciones

```
comp <- tibble(modelo = c("log",
                          "logit",
                          "probit",
                          "cauchit",
                          "cloglog",
                          "dos-pasos"),
               estimacion = c(conversion(mod_log, "log")$mu,
                              conversion(mod_logit, "logit")$mu,
                              conversion(mod_probit, "probit")$mu,
                              conversion(mod_cauchit, "cauchit")$mu,
                              conversion(mod_cloglog, "cloglog")$mu,
                              (((sin(mu.est))^2)))) |>
  mutate(ratio = if_else((0.26/estimacion)<1, estimacion/0.26, 0.26/estimacion))

ggplot(comp, aes(modelo, ratio))+
  geom_point(aes(color = modelo, shape = modelo), size = 5)+
  geom_hline(yintercept = 1, colour = "blue", linetype = "dashed")+
  geom_hline(yintercept = 1.2, colour = "red", linetype = "dashed")+
  theme_minimal() +
  labs(
    title = "Diferencias entre estimaciones del efecto mediante difere
ntes modelos",
    subtitle = "modelos vs. media distribución ",
    x = "modelo",
    y = "Cambio "
  )
)
```



AIC

```
AIC(mod_log)
[1] 441.4449
AIC(mod_logit)
[1] 423.3511
AIC(mod_probit)
[1] 421.5565
AIC(mod_cauchit)
[1] 469.1166
AIC(mod_cloglog)
[1] 429.5893
AIC.rma(rs1t)
[1] 20.5983
```

Valores altos de prevalencia

```
abx_1 <- add_row(abx, ID = c(39, 40, 41),
  study = c("Test1", "Test 2", "Test 3" ),
  N = c(146, 74, 51 ),
  total = c(145, 72, 50 ),
```

```

    prev = total/N)
abx_1
# A tibble: 41 × 24
  ID study country period population      N SJS  SJS_TEN TEN  Set
ting Age
  <dbl> <chr> <chr>  <chr>  <chr>      <dbl> <chr> <chr>  <chr> <ch
r>  <chr>
1     1 De B... US      "2000... Adults      131 44   63    24  Inp
ati... 44.9...
2     2 Gleg... US      "1990... Children     51  0    3    48  Inp
ati... 8 (0...
3     3 Olte... Canada "1999... Adults      43  0    7    36  Inp
ati... 54+/-...
4     4 Sing... India  "Jan ... Both       75 48   18    9  Inp
ati... NR
5     5 Thak... India  "Jan ... Unclear    51  7   16   28  Inp
ati... 38.2...
6     6 Abou... Egypt  "Jan ... Adults      83 56   22    5  Inp
ati... 37.9...
7     7 Huye... Vietnam "Jul ... Adults     60 30    0   30  Inp
ati... 52+/-...
8     8 Yang... China  "Jan ... Adults    213 141  19   53  Inp
ati... 48.2...
9     9 Carr... Puerto... "2006... Both      30  3   10   17  Inp
ati... 35.9...
10    10 Chan... Austra... "Jan ... Adults     42 10    6   26  Inp
ati... 56
# i 31 more rows
# i 13 more variables: Female <chr>, Death <dbl>, beta <dbl>, fq <dbl>
,
# sulfa <dbl>, macro <dbl>, tetra <dbl>, vanco <dbl>, other_abx <dbl>
>,
# total <dbl>, prev <dbl>, continente <fct>, n <dbl>

mod_log_t1 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx_1,
family = binomial(link = "log"))
mod_logit_t1 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx_1
, family = binomial(link = "logit"))
mod_probit_t1 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx_
1, family = binomial(link = "probit"))
mod_cauchit_t1 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx
_1, family = binomial(link = "cauchit"))

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$chec
kConv, :
Model failed to converge with max|grad| = 0.0250873 (tol = 0.002, comp
onent 1)

mod_cloglog_t1 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx
_1, family = binomial(link = "cloglog"))

test1 <- escalc("PAS",xi = total, ni = N, data = abx_1, add = 0.5)

rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt

```

```
Random-Effects Model (k = 41; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.0812 (SE = 0.0216)
tau (square root of estimated tau^2 value):      0.2849
I^2 (total heterogeneity / total variability):   96.16%
H^2 (total variability / sampling variability):  26.02

Test for Heterogeneity:
Q(df = 40) = 1040.8886, p-val < .0001

Model Results:

estimate      se      zval      pval      ci.lb      ci.ub      ***
  0.6264  0.0456  13.7359  <.0001  0.5370  0.7158

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)

(sin(mu.est))^2
[1] 0.3436778

(sin(mu.ci))^2
[1] 0.2617188 0.4306189

conversion(mod_log_t1, "log")

$mu
[1] 0.2930809

$`CI 95%`
[1] 0.2416133 0.3555118

$tau2
[1] 0.3498503

$tau
[1] 0.5914815

conversion(mod_logit_t1, "logit")

$mu
[1] 0.3375818

$`CI 95%`
[1] 0.2493744 0.4387523

$tau2
```

```
[1] 1.839635

$tau
[1] 1.356331

conversion(mod_probit_t1, "probit")

$mu
[1] 0.3405638

$`CI 95%`
[1] 0.2558124 0.4342596

$tau2
[1] 0.6061852

$tau
[1] 0.778579

conversion(mod_cauchit_t1, "cauchit")

$mu
[1] 0.3068325

$`CI 95%`
[1] 0.3056505 0.3080208

$tau2
[1] 2.748271

$tau
[1] 1.657791

conversion(mod_cloglog_t1, "cloglog")

$mu
[1] 0.3126674

$`CI 95%`
[1] 0.2451151 0.3934321

$tau2
[1] 0.8139288

$tau
[1] 0.90218
```

Valores bajos

```
abx_2 <- add_row(abx, ID = c(39, 40, 41),
  study = c("Test1", "Test 2", "Test 3" ),
  N = c(150, 200, 190 ),
  total = c(1, 1, 1),
  prev = total/N)
abx_2
```

```
# A tibble: 41 × 24
  ID study country period population      N SJS   SJS_TEN TEN   Set
ting Age
  <dbl> <chr> <chr>   <chr> <chr>      <dbl> <chr> <chr>   <chr> <ch
r> <chr>
1     1 1 De B... US      "2000... Adults      131 44    63     24  Inp
ati... 44.9...
2     2 2 Gleg... US      "1990... Children     51 0     3     48  Inp
ati... 8 (0...
3     3 3 Olte... Canada "1999... Adults      43 0     7     36  Inp
ati... 54+/-...
4     4 4 Sing... India  "Jan ... Both       75 48    18     9   Inp
ati... NR
5     5 5 Thak... India  "Jan ... Unclear    51 7     16    28  Inp
ati... 38.2...
6     6 6 Abou... Egypt  "Jan ... Adults      83 56    22     5   Inp
ati... 37.9...
7     7 7 Huye... Vietnam "Jul ... Adults      60 30     0    30  Inp
ati... 52+/-...
8     8 8 Yang... China  "Jan ... Adults     213 141   19    53  Inp
ati... 48.2...
9     9 9 Carr... Puerto... "2006... Both       30 3     10    17  Inp
ati... 35.9...
10    10 10 Chan... Austra... "Jan ... Adults      42 10     6    26  Inp
ati... 56

# i 31 more rows
# i 13 more variables: Female <chr>, Death <dbl>, beta <dbl>, fq <dbl>
,
#   sulfa <dbl>, macro <dbl>, tetra <dbl>, vanco <dbl>, other_abx <dbl>
>,
#   total <dbl>, prev <dbl>, continente <fct>, n <dbl>

mod_log_t2 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx_2,
family = binomial(link = "log"))
mod_logit_t2 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx_2
, family = binomial(link = "logit"))
mod_probit_t2 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx_
2, family = binomial(link = "probit"))
mod_cauchit_t2 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx
_2, family = binomial(link = "cauchit"))

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$chec
kConv, :
Model failed to converge with max|grad| = 0.0246434 (tol = 0.002, comp
onent 1)

mod_cloglog_t2 <- glmer(cbind(total, N - total)~ 1 +(1|ID), data = abx
_2, family = binomial(link = "cloglog"))

test1 <- escalc("PAS",xi = total, ni = N, data = abx_2, add = 0.5)

rslt <- rma(yi = yi, vi = vi, method = "DL", data = test1, ); rslt

Random-Effects Model (k = 41; tau^2 estimator: DL)
```

```
tau^2 (estimated amount of total heterogeneity): 0.0516 (SE = 0.0141)
tau (square root of estimated tau^2 value):      0.2271
I^2 (total heterogeneity / total variability):   94.51%
H^2 (total variability / sampling variability):  18.22
```

Test for Heterogeneity:

```
Q(df = 40) = 728.8987, p-val < .0001
```

Model Results:

```
estimate      se      zval      pval      ci.lb      ci.ub
  0.5240  0.0368  14.2404  <.0001  0.4518  0.5961  ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sconv <- summary(rslt)
mu.est <- as.numeric(rslt$beta)
mu.se <- rslt$se
mu.ci <- c(mu.est - qnorm(1 - 0.05/2) * mu.se,
mu.est + qnorm(1 - 0.05/2) * mu.se)
```

```
(sin(mu.est))^2
```

```
[1] 0.2503112
```

```
(sin(mu.ci))^2
```

```
[1] 0.1906412 0.3151663
```

```
conversion(mod_log_t2, "log")
```

```
$mu
```

```
[1] 0.2100033
```

```
$`CI 95%`
```

```
[1] 0.1586115 0.2780465
```

```
$tau2
```

```
[1] 0.7633214
```

```
$tau
```

```
[1] 0.8736827
```

```
conversion(mod_logit_t2, "logit")
```

```
$mu
```

```
[1] 0.2252016
```

```
$`CI 95%`
```

```
[1] 0.1691804 0.2932262
```

```
$tau2
```

```
[1] 1.230596
```

```
$tau
[1] 1.109322
conversion(mod_probit_t2, "probit")
$mu
[1] 0.2343517
$`CI 95%`
[1] 0.1789072 0.2981798
$tau2
[1] 0.3695362
$tau
[1] 0.6078949
conversion(mod_cauchit_t2, "cauchit")
$mu
[1] 0.1901745
$`CI 95%`
[1] 0.1896117 0.1907403
$tau2
[1] 2.824769
$tau
[1] 1.680705
conversion(mod_cloglog_t2, "cloglog")
$mu
[1] 0.2176503
$`CI 95%`
[1] 0.1637882 0.2859597
$tau2
[1] 0.9724869
$tau
[1] 0.9861475
```