



Universidad de Valladolid

FACULTAD DE CIENCIAS

GRADO EN ESTADÍSTICA

METODOLOGÍA ESTADÍSTICA UTILIZADA EN EL META-ANÁLISIS.
APLICACIÓN A LA INVESTIGACIÓN SOBRE DERMATITIS ATÓPICA

AUTOR:

LIZA CHENG YIN

Tutoras::

ITZIAR FERNÁNDEZ MARTÍNEZ
YOLANDA LARRIBA GONZÁLEZ

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

CURSO: 2024-2025

Resumen

El meta-análisis es una herramienta estadística fundamental en la investigación médica, ya que permite combinar y analizar los resultados de múltiples estudios para obtener conclusiones más sólidas y generalizables. Este trabajo revisa en profundidad la metodología estadística del meta-análisis, abordando aspectos clave como la estimación del tamaño del efecto, la evaluación de la heterogeneidad y la detección de sesgos.

Además, se presenta una aplicación práctica de esta técnica en el ámbito de la dermatología, evaluando la eficacia de los agentes biológicos y los inhibidores de JAK en la dermatitis atópica moderada a grave en adultos. Los resultados muestran que ambos tipos de tratamientos son eficaces, aunque presentan diferencias en eficacia y perfil de seguridad. El análisis de sesgo de publicación no detectó problemas relevantes, lo que refuerza la validez de las conclusiones. Este trabajo pone de manifiesto la importancia del meta-análisis para la toma de decisiones clínicas basadas en la mejor evidencia disponible.

Abstract

Meta-analysis is a fundamental statistical tool in medical research, as it allows for the combination and analysis of results from multiple studies to obtain more robust and generalizable conclusions. This work provides an in-depth review of the statistical methodology of meta-analysis, addressing key aspects such as effect size estimation, assessment of heterogeneity, and detection of biases.

Additionally, a practical application of this technique is presented in the field of dermatology, evaluating the efficacy of biological agents and JAK inhibitors in moderate to severe atopic dermatitis in adults. The results show that both types of treatments are effective, although they exhibit differences in efficacy and safety profiles. The publication bias analysis did not detect any relevant issues, which reinforces the validity of the conclusions. This work highlights the importance of meta-analysis for clinical decision-making based on the best available evidence.

Tabla de Contenido

1. Introducción	1
1.0.1. Contextualización	1
1.0.2. Objetivos	3
1.0.3. Asignaturas clave	3
2. Diseño y planificación del meta-análisis	4
2.1. Justificación del uso del MA	4
2.2. Pregunta de investigación y selección de estudios	4
2.3. Recogida de datos y plan de análisis	6
2.4. Limitaciones y desafíos metodológicos del MA	7
2.4.1. <i>Garbage in, garbage out</i>	7
2.4.2. Comparaciones inapropiadas (<i>apples and oranges</i>)	7
2.4.3. Sesgo de publicación	8
2.4.4. Sesgos derivados del diseño del análisis	8
2.4.5. Tamaño muestral reducido	8
3. Fundamentos estadísticos del MA	9
3.1. Tamaño de efecto	9
3.1.1. Diferencia de medias	10
3.1.2. Diferencia de medias estandarizadas	10

3.1.3.	Razón de riesgo	11
3.1.4.	Odds ratio	12
3.1.5.	Razón de Tasas de Incidencia	12
3.2.	Modelo de combinación de estudios	13
3.2.1.	Modelo de efectos fijos	14
3.2.2.	Modelo de efectos aleatorios	15
3.3.	Evaluación de la heterogeneidad	16
3.3.1.	Estadístico Q de Cochran	16
3.3.2.	Estadístico I^2 y H^2 de Higgins y Thompson	17
3.3.3.	Estimación de la varianza entre estudios (τ^2)	18
3.4.	Evaluación del sesgo de publicación	19
3.4.1.	Gráfico de embudo (<i>funnel plot</i>)	19
3.4.2.	Prueba de Egger	19
3.4.3.	Prueba de Peters	20
3.4.4.	Método de Recorte y Relleno (<i>trim and fill</i>)	21
3.4.5.	Análisis de subgrupos	21
3.5.	Visualización de resultados	22
3.5.1.	Gráficos de Bosque (<i>Forest Plots</i>)	22
3.5.2.	Diagrama de cortina (<i>Drapery Plots</i>)	22
3.6.	Paquetes de R	25
4.	Aplicación a la Dermatitis Atópica	26
4.1.	Motivación	26
4.2.	Identificación y selección de estudios	27
4.3.	Estimación de tamaños del efecto individuales y pesos	29
4.4.	Estimación del efecto combinado y análisis de la heterogeneidad	30

4.5. Análisis por subgrupos	31
4.6. Sesgo de publicación	34
4.7. Interpretación de los resultados y limitaciones	34
5. Conclusiones	36
Bibliografía	41
6. Anexo A: Código R	42

Capítulo 1

Introducción

1.0.1. Contextualización

El meta-análisis (MA) se ha convertido en una herramienta fundamental en la investigación médica, permitiendo la síntesis y el análisis conjunto de los resultados procedentes de múltiples estudios para obtener conclusiones más robustas y generalizables. Esta técnica estadística, que ha adquirido una notable relevancia en las últimas décadas, ofrece importantes ventajas, como el aumento de la potencia estadística, la mejora en la precisión de las estimaciones y la capacidad para detectar patrones consistentes entre estudios. Su aplicación se extiende a múltiples áreas de la medicina donde la evidencia puede ser especialmente heterogénea o contradictoria. Concretamente, este Trabajo Fin de Grado (TFG) se centra en el uso del MA en el ámbito de la dermatología, con el objetivo de ilustrar su aplicabilidad en la evaluación crítica y cuantitativa de la eficacia de distintos tratamientos frente a la dermatitis atópica (DA), una enfermedad de alta prevalencia e impacto clínico, especialmente en la infancia.

La DA conocida comúnmente como eccema atópico, es una enfermedad dermatológica inflamatoria crónica de la piel que se caracteriza por sequedad, prurito y la presencia de erupciones cutáneas. Aunque su incidencia es mayor en la infancia, también puede persistir o incluso comenzar en la edad adulta. Los síntomas principales incluyen parches eritematosos y secos en la piel, acompañados de intenso prurito y, ocasionalmente, pequeñas vesículas que, al romperse, pueden formar costras [Gómez, 2018].

La DA, especialmente en formas crónicas o severas, puede generar importantes dificultades en la vida diaria y el bienestar psicológico de quienes la padecen. Los síntomas físicos, además de ser molestos o dolorosos, pueden limitar la movilidad y alterar la calidad del sueño debido al prurito nocturno e insomnio. Por otro lado, la elección de la ropa requiere ciertas precauciones, ya que determinados tejidos pueden irritar la piel o agravar los síntomas. La visibilidad de las lesiones en áreas expuestas también puede generar ansiedad y depresión, afectando negativamente la autoestima del individuo.

Actualmente no existe un tratamiento que permita la curación completa de la DA. Por ello, el manejo terapéutico se centra en aliviar los síntomas, disminuir la frecuencia de los brotes y controlar la enfermedad a largo plazo. En fases iniciales, se emplean tratamientos tópicos para reducir la inflamación, aliviar el picor y mantener una adecuada hidratación mediante productos específicos para pieles atópicas. Sin embargo, en los casos más severos, estas medidas pueden ser insuficientes. En tales situaciones, puede recurrirse a tratamientos sistémicos con inmunosupresores convencionales (p.e., ciclosporina) o agentes biológicos (p.e., dupilumab o tralokinumab) o más recientemente, inhibidores de la cinasa Jano o JAK (p.e., abrocitinib, upadacitinib o baricitinib) siendo estos dos últimos los más empleados e investigados actualmente.

En los últimos años, el tratamiento de la DA ha experimentado importantes avances, particularmente en el ámbito de las terapias farmacológicas, con la aparición de nuevos medicamentos dirigidos a mecanismos inmunológicos específicos. Como consecuencia, la literatura científica ha generado una amplia colección de estudios clínicos, muchos de ellos aleatorizados, que evalúan la eficacia y seguridad de estos tratamientos.

No obstante, estos estudios suelen presentar diferencias metodológicas o tamaños muestrales limitados, lo que dificulta extraer conclusiones sólidas de forma individual. En el caso de los inhibidores de JAK, aunque los ensayos clínicos han mostrado mejoras rápidas y significativas frente a placebo, la falta de estudios a largo plazo y la variabilidad en los desenlaces reportados dificultan la generalización de los resultados y la comparación directa [AEMPS, 2022].

En este contexto, el MA se presenta como una herramienta estadística de gran utilidad para integrar, comparar y sintetizar cuantitativamente los resultados obtenidos en múltiples investigaciones independientes. Esta metodología permite abordar las limitaciones de los estudios individuales, combinando sus resultados para obtener una visión más completa y precisa sobre la efectividad de las intervenciones terapéuticas.

A diferencia de otros enfoques, en el MA las unidades de análisis son los estudios en sí mismos, no los sujetos individuales. Cada estudio aporta una estimación del tamaño del efecto junto con una medida de su precisión y se pondera de acuerdo con la fiabilidad de sus resultados. Esta metodología posibilita la identificación de tendencias generales, aumentando la potencia estadística y reduciendo la incertidumbre en la interpretación de los datos. Al integrar información procedente de diversas fuentes, esta metodología proporciona una base más sólida para la toma de decisiones médicas y el diseño de futuras investigaciones.

En este TFG se revisarán los conceptos fundamentales, los problemas metodológicos y los procedimientos clave asociados al MA, como paso previo a su aplicación cuantitativa en la evaluación de la eficacia clínica de distintos tratamientos para la DA. A continuación, se presentarán los fundamentos estadísticos que sustentan esta técnica, incluyendo una descripción de las principales medidas de efecto, modelos de combinación de estudios, estimación del efecto combinado y la evaluación de la heterogeneidad y el sesgo de pu-

blicación. Finalmente, se describirá una aplicación práctica del MA como ilustración de los métodos expuestos, mediante el análisis de la eficacia clínica de diversos tratamientos utilizados en DA.

1.0.2. Objetivos

General

Describir y analizar en profundidad la metodología estadística del MA, con especial énfasis en los modelos de análisis, la evaluación de la heterogeneidad entre estudios, la detección de sesgos y la interpretación de los resultados.

Específicos

- Presentar los fundamentos teóricos del MA y su evolución en el contexto de la investigación biomédica.
- Explicar las principales técnicas para cuantificar y modelizar la heterogeneidad entre estudios.
- Revisar los métodos para identificar y evaluar posibles sesgos en el MA.
- Ilustrar la aplicación práctica del MA mediante la comparación de la eficacia de dos clases de tratamientos (agentes biológicos e inhibidores de JAK) para la DA moderada a grave en población adulta.

1.0.3. Asignaturas clave

Las asignaturas de Modelos Lineales, Regresión y ANOVA constituyen el núcleo metodológico fundamental para comprender y aplicar correctamente las técnicas empleadas en un MA. Junto con Inferencia Estadística, estas materias ofrecen una base sólida para el razonamiento estadístico, el análisis comparado y la construcción e interpretación de modelos.

Asimismo, asignaturas como Computación Estadística, Estadística Descriptiva y Análisis de Datos complementan esta formación al proporcionar herramientas prácticas para el tratamiento de datos reales, la visualización de resultados y el uso eficiente del software estadístico, habilidades esenciales en la implementación rigurosa de un MA.

Capítulo 2

Diseño y planificación del meta-análisis

2.1. Justificación del uso del MA

En el campo de la investigación científica, la toma de decisiones informadas y la búsqueda de la verdad son pilares fundamentales para el avance del conocimiento. Por eso, la creciente cantidad de información disponible en la literatura científica debería ser algo positivo para el avance de la ciencia. Sin embargo, muchas veces nos encontramos con resultados y conclusiones muy variadas que causan incertidumbre en la toma de decisiones.

En este contexto surgió el MA, una técnica estadística que se desarrolló para satisfacer la necesidad de ofrecer una herramienta para llevar a cabo revisiones mediante procedimientos rigurosos y sistemáticos en la que combinamos y resumimos resultados de múltiples estudios individuales. Al combinar múltiples estudios se dispone de un tamaño muestral global mayor, lo que permite obtener estimaciones más precisas y con mayor potencia estadística.

2.2. Pregunta de investigación y selección de estudios

Antes de iniciar el MA, resulta fundamental definir de forma clara y estructurada la pregunta de investigación, así como establecer unos criterios de inclusión y exclusión de estudios bien justificados. Estos criterios permitirán realizar una búsqueda sistemática y rigurosa de la literatura científica, identificando los estudios relevantes y seleccionando únicamente aquellos que cumplan con los requisitos metodológicos establecidos.

Existen guías, estándares y recomendaciones que orientan cada etapa del proceso me-

metodológico. Seguir estas guías ayuda a garantizar que el MA incluya toda la evidencia relevante, sea imparcial y no sesgado y que extraiga conclusiones válidas y aplicables. Sin embargo, no siempre hay una única forma de abordar ciertas decisiones, por lo que es crucial que todas las decisiones adoptadas durante el proceso sean transparentes y que estén debidamente justificadas para que puedan ser reproducibles.

Para definir una buena pregunta de investigación es necesario contar con conocimientos previos sobre el tema y conocer cuestiones relevantes no resueltas en el campo. Una vez formulada la pregunta de investigación, se realiza una revisión de la literatura previa analizando cómo otros estudios anteriores han abordado el tema y si han dejado cuestiones abiertas.

Los criterios FINER (Fiable, Interesante, Novedosa, Ética y Relevante), propuestos por [Cummings et al., 2013], se utilizan para valorar la calidad de la pregunta de investigación. Una pregunta debe ser viable en cuanto a recursos y tiempo, interesante para la comunidad científica, aportar novedad, cumplir principios éticos y tener relevancia para el avance del conocimiento o la práctica clínica.

Posteriormente, es fundamental establecer los criterios de inclusión y exclusión de los estudios. Una herramienta útil es el marco PICO [Mattos and Ruellas, 2015], que ayuda a estructurar la pregunta de investigación en cuatro partes:

- Población (*Population*): Tipo de pacientes o grupo de estudio.
- Intervención (*Intervention*): El tipo de tratamiento que se quiere estudiar.
- Comparador (*Comparison*): Alternativa contra la que se compara la intervención.
- Resultado (*Outcome*): Resultados que se quiere obtener tras la intervención.

Además, es recomendable especificar el rango cultural y lingüístico de los estudios. La mayoría de los estudios se basan en poblaciones *WEIRD* (occidentales, educadas, industrializadas, ricas y democráticas), lo que limita la generalización de resultados a otras culturas [Henrich et al., 2010].

Una vez definidos los criterios de selección, se realiza una búsqueda sistemática y exhaustiva en diversas bases de datos para identificar todos los estudios relevantes. Entre las bases de datos más importantes están PubMed, PsycInfo, Embase y la librería de Cochrane. Tras reunir todos los estudios potencialmente elegibles, el proceso de selección se desarrolla en tres fases:

1. Al buscar en varias fuentes de datos, es necesario realizar un cribado inicial para eliminar los duplicados. Este paso puede realizarse con un software de gestión de referencias como Zotero, Mendeley o EndNote.

2. Se lleva a cabo una revisión de títulos y resúmenes para excluir aquellos estudios que claramente no estén relacionados con la pregunta de investigación.
3. Los estudios restantes pasan a una revisión del texto completo, donde se verifica si cumplen con todos los criterios de inclusión y exclusión previamente establecidos.

Una vez terminado, es fundamental documentar todo el proceso de selección. Para ello, se recomienda utilizar un diagrama de flujo, siguiendo estándares como el diagrama PRISMA [Moher et al., 2009] para garantizar la reproducibilidad y la transparencia.

2.3. Recogida de datos y plan de análisis

Una vez seleccionados los estudios que se van a incluir en el MA, se debe extraer las características de cada estudio y presentarlas en una tabla resumen. Entre los datos imprescindibles se encuentran el autor, el año de publicación y el tamaño muestral. También es recomendable incluir información adicional como la media de edad de los participantes, la proporción de mujeres y hombres, la intervención y los resultados que aporta, ya que no todos los estudios aportarán las mismas variables de interés.

Además, es importante recoger los datos necesarios para calcular el tamaño de efecto, como medias y desviaciones estándar, número de eventos o proporciones, según la medida de efecto elegida.

Por otro lado, es fundamental evaluar y reportar la calidad metodológica y el riesgo de sesgo de los estudios incluidos. Para los ensayos controlados aleatorizados, la herramienta recomendada es la Herramienta de Riesgo de Sesgo desarrollada por Cochrane (RoB 2, *Risk of Bias 2*) [Sterne et al., 2019]. Esta herramienta evalúa cinco dominios principales: sesgo por el proceso de aleatorización, desviaciones de las intervenciones previas, datos faltantes, medición de los resultados y selección de resultados reportados.

Para los estudios no aleatorizados de intervenciones, se emplea la herramienta ROBINS-I (*Risk Of Bias In Non-randomised Studies of Interventions*) [Sterne et al., 2016], que evalúa el riesgo de sesgo en siete dominios: confusión, selección de participantes, clasificación de las intervenciones, desviaciones de las intervenciones, datos faltantes, medición de los resultados y selección de los resultados reportados.

Ambas herramientas permiten clasificar los estudios como de bajo riesgo, riesgo incierto o alto riesgo de sesgo, lo que ayuda a valorar la solidez de la evidencia.

Además de evaluar el riesgo de sesgo, es crucial planificar detalladamente los análisis antes de iniciar el MA. Un plan predefinido garantiza transparencia y reproducibilidad, evitando que las decisiones metodológicas se vean influenciadas por los resultados observados. Este plan debe especificar todos los cálculos y comparaciones previstas, ya que los análisis a priori son considerados más válidos y fiables.

El plan de análisis tiene dos objetivos principales: garantizar la transparencia y permitir la reproducibilidad del MA. Definir los análisis a priori permite a otros investigadores verificar que las decisiones metodológicas no fueron influenciadas por los resultados y facilita que el estudio pueda ser replicado.

Se debe detallar qué información se va a extraer de los estudios incluidos, qué medida del tamaño de efecto se calculará y si se empleará un modelo de efectos fijos o aleatorios, dependiendo de la heterogeneidad esperada entre los estudios.

Además, es importante decidir si se realizarán análisis de subgrupos y especificar los criterios exactos para la clasificación de los estudios en cada subgrupo. Todo esto contribuye a la solidez metodológica y a la validez de las conclusiones del MA.

2.4. Limitaciones y desafíos metodológicos del MA

2.4.1. *Garbage in, garbage out*

El fenómeno conocido como *Garbage in, garbage out* (“Entra basura, sale basura”) hace referencia a la escasa validez de los resultados de un MA cuando los estudios incluidos presentan sesgos o errores sistemáticos importantes. Por ello, es fundamental evaluar adecuadamente el riesgo de sesgo de cada estudio primario. No obstante, si la mayoría de los estudios son de baja calidad, ni siquiera el MA más riguroso podrá compensar estas limitaciones. En tales casos, la principal conclusión metodológica suele ser la necesidad de realizar investigaciones futuras con mayor calidad y rigor.

2.4.2. Comparaciones inapropiadas (*apples and oranges*)

Una crítica frecuente a los MA es la posibilidad de combinar estudios excesivamente heterogéneos, lo que se conoce coloquialmente como “comparar peras con manzanas”. Aunque cierta heterogeneidad es inevitable, lo fundamental es que los estudios compartan suficientes características clave para responder de forma coherente a la pregunta de investigación.

Un MA muy amplio puede carecer de aplicabilidad clínica, mientras que uno demasiado restrictivo puede no aportar información generalizable. Lo esencial es definir criterios claros que respondan a una pregunta de investigación concreta. La heterogeneidad entre estudios no es necesariamente un problema si se aborda y analiza adecuadamente, ya que puede aportar información valiosa sobre la variabilidad de los efectos [Borenstein et al., 2021].

2.4.3. Sesgo de publicación

El sesgo de publicación ocurre cuando solo se publican estudios con resultados positivos o significativos, mientras que los estudios con resultados negativos o nulos suelen quedar sin publicar. Esto distorsiona la evidencia disponible y puede sobrestimar los efectos reales, ya que los estudios negativos están sistemáticamente subrepresentados en la literatura.

Para minimizar este problema, se recomienda realizar búsquedas exhaustivas que incluyan literatura gris y estudios no publicados, además de emplear métodos estadísticos específicos, como los gráficos de embudo (apartado 3.1), que permiten detectar y valorar la presencia de sesgo de publicación.

2.4.4. Sesgos derivados del diseño del análisis

Llevar a cabo un MA implica tomar numerosas decisiones metodológicas que dan a los investigadores un amplio margen de flexibilidad. Esta flexibilidad puede introducir sesgos en la selección, el análisis y la interpretación de los resultados, especialmente si existe conflicto de intereses o familiaridad con la intervención evaluada. Incluso partiendo de los mismos datos, diferentes analistas pueden llegar a conclusiones muy distintas [Silberzahn et al., 2018]. Para reducir este sesgo de diseño, se recomienda registrar y publicar previamente un plan de análisis detallado, lo que favorece la transparencia y la objetividad.

2.4.5. Tamaño muestral reducido

Los estudios incluidos en un MA solo representan una muestra parcial de la población, lo que conlleva un error de muestreo inherente. Cuanto mayor es el tamaño muestral, menor es este error y más precisa la estimación del tamaño del efecto. En el MA, se da mayor peso a los estudios con menor error estándar. Sin embargo, los estudios con muestras pequeñas tienden a sobrestimar el tamaño del efecto, lo que se conoce como sesgo del tamaño de muestra pequeña. Para corregir este sesgo, se puede aplicar la corrección de g de Hedges, que será abordada con más detalle en el apartado 3.1.2.

Capítulo 3

Fundamentos estadísticos del MA

En este capítulo trataremos los fundamentos estadísticos del MA, incluyendo los principales tamaños de efecto como la diferencia de medias, razón de riesgo y odds ratio. También se explicarán los modelos de combinación de estudios (efectos fijos y aleatorios), la evaluación de la heterogeneidad y el sesgo de publicación. Además, se presentarán las herramientas gráficas para la visualización de resultados, proporcionando así las bases necesarias para un análisis riguroso y adecuado del MA.

3.1. Tamaño de efecto

El tamaño de efecto (*Effect Size*) es una medida cuantitativa que permite estimar la magnitud de una diferencia entre grupos o la intensidad de una asociación. En el contexto del MA, su elección depende del tipo de variable resultado y de las características de los estudios incluidos.

El tamaño del efecto verdadero de un estudio k se denota por θ_k , y representa la magnitud real del efecto en la población. Sin embargo, este valor es desconocido. Lo que se observa en la práctica es una estimación muestral, denotada por $\hat{\theta}_k$, que difiere de θ_k debido al error de muestreo.

$$\hat{\theta}_k = \theta_k + \varepsilon_k \tag{3.1}$$

Para que el valor de $\hat{\theta}_k$ sea lo más cercano al tamaño de efecto verdadero se asume que aquellos estudios con menor ε_k ayudarán a estimar θ_k de forma más precisa. Por ello, al combinar los resultados de los estudios se asignarán pesos mayores a aquellos estudios con mayor precisión o lo que es lo mismo menor error de muestreo [Hedges and Olkin, 2014]. El ε_k al igual que θ_k es desconocido, sin embargo, en muchos casos una forma para aproximar

este error es utilizando el error estándar (SE, *Standard Error*), que puede calcularse a partir de los datos disponibles en cada estudio.

En los siguientes apartados describiremos las principales medidas del tamaño de efecto utilizadas en el MA, junto con su interpretación.

3.1.1. Diferencia de medias

Cuando la variable resultado es continua y está medida en la misma escala en todos los estudios, la diferencia de medias entre dos grupos independientes (habitualmente, un grupo experimental y un grupo control) constituye una medida directa del efecto. Esta diferencia está sujeta a variabilidad, ya que se calcula a partir de datos muestrales. La varianza asociada a esta estimación se obtiene sumando las varianzas muestrales de ambos grupos divididas por sus respectivos tamaños muestrales. Esta precisión permite ponderar cada estudio según la precisión de su estimación, dando mayor peso a aquellos con menor variabilidad o mayor tamaño muestral.

3.1.2. Diferencia de medias estandarizadas

Cuando los estudios incluidos en un MA utilizan escalas diferentes para medir una misma variable continua, la diferencia de medias no puede interpretarse directamente. En estos casos, se utiliza la diferencia de medias estandarizada (DME), que expresa la magnitud del efecto en unidades de desviación estándar. La estimación más habitual es la *d* de Cohen (*d*), calculada como la diferencia de medias dividida por la desviación estándar combinada de ambos grupos.

Su varianza se estima como:

$$Var(d) = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)} \quad (3.2)$$

En estudios con tamaños muestrales pequeños, se recomienda aplicar una corrección mediante el estadístico *g* de Hedges, que ajusta el sesgo de estimación con la siguiente expresión:

$$g = d \cdot \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right) \quad (3.3)$$

Y su varianza es:

$$Var(g) = Var(d) \cdot J^2, \quad \text{donde } J = 1 - \frac{3}{4(n_1 + n_2) - 9} \quad (3.4)$$

Para una interpretación clínica más intuitiva de la magnitud del efecto, suele utilizarse la clasificación propuesta por [Cohen, 1988], que establece:

- $d \approx 0,2 \rightarrow$ Efecto *pequeño*
- $d \approx 0,5 \rightarrow$ Efecto *moderado*
- $d \approx 0,8 \rightarrow$ Efecto *grande*

Sin embargo, estos umbrales deben interpretarse con cautela y en función del contexto específico de la variable analizada.

3.1.3. Razón de riesgo

El riesgo relativo (RR, *Risk Ratio*) es una de las medidas estadísticas más utilizadas para variables resultado dicotómicas. Se emplea fundamentalmente en estudios de ensayos clínicos con resultados dicotómicos, donde los datos suelen organizarse en una tabla de contingencia 2x2 (tabla 3.1), donde se registran las frecuencias de aparición del evento de interés y su ausencia en los grupos de tratamiento y control. Esta estructura es fundamental para calcular medidas como el RR.

	Evento	No Evento	Total
Grupo tratamiento	n_{11}	n_{12}	$n_{1.}$
Grupo control	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

Tabla 3.1: Tabla de contingencia 2x2 para el cálculo del RR.

A partir de estos datos, el RR se define como el cociente entre la probabilidad del evento en el grupo tratamiento (p_1) y la del grupo control (p_2):

$$RR = \frac{p_1}{p_2} = \frac{n_{11}/n_{1.}}{n_{21}/n_{2.}} \quad (3.5)$$

Debido a que la distribución del RR es asimétrica, se emplea su transformación logarítmica para los análisis:

$$\log(RR) = \log(p_1) - \log(p_2) \quad (3.6)$$

Y la varianza se calcula con la siguiente fórmula:

$$Var(\log(RR)) = \frac{1}{n_{11}} - \frac{1}{n_{1.}} + \frac{1}{n_{21}} - \frac{1}{n_{2.}} \quad (3.7)$$

Según el valor obtenido, su interpretación será: un RR de 1 indica que el riesgo es idéntico en ambos grupos, mientras que un RR mayor que 1 sugiere un riesgo aumentado en el grupo tratamiento, y un RR menor que 1 indica una reducción del riesgo.

3.1.4. Odds ratio

La odds ratio (OR) es una medida del tamaño utilizada en estudios con respuesta dicotómicas. Se define como el cociente entre la odds del evento en el grupo tratamiento y las odds correspondientes en el grupo control. A diferencia de la razón de riesgos, la OR no compara probabilidades, sino razones entre la probabilidad del evento y la de no evento en cada grupo.

$$OR = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} \quad (3.8)$$

Dado que la distribución de la OR es asimétrica, se utiliza su transformación logarítmica para el análisis estadístico:

$$\log(OR) = \log\left(\frac{n_{11}/n_{12}}{n_{21}/n_{22}}\right) \quad (3.9)$$

La varianza del $\log(OR)$ se calcula de la siguiente forma:

$$Var(\log(OR)) = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \quad (3.10)$$

Según el valor obtenido, su interpretación será que una OR igual a 1 indica que las odds del evento son equivalentes entre grupos. Valores superiores a 1 indican mayor odds en el grupo de tratamiento, y valores inferiores a 1 reflejan menor odds.

3.1.5. Razón de Tasas de Incidencia

La razón de tasas de incidencia (IRR, *Incidence Rate Ratio*) se emplea cuando los estudios comparan tasas de eventos ajustadas por el tiempo de seguimiento, en lugar de proporciones simples. Es especialmente útil cuando los participantes se observan durante

diferentes periodos de tiempo y se producen pérdidas durante el seguimiento (por ejemplo, debido a abandonos, fallecimientos o censura administrativa).

La tasa de incidencia (IR, *Incidence Rate*) se define como el número de eventos ocurridos dividido por el tiempo total de observación en riesgo, habitualmente expresado como persona-tiempo.

$$IR = \frac{\text{Numero de eventos}}{\text{Tiempo-persona Total}} = \frac{E}{T} \quad (3.11)$$

La IRR se calcula como el cociente entre las tasas de incidencia de los dos grupos:

$$IRR = \frac{IR_1}{IR_2} = \frac{E_1/T_1}{E_2/T_2} \quad (3.12)$$

Dado que la distribución de la IRR es asimétrica, se utiliza su transformación logarítmica para el análisis estadístico.

$$\log(IRR) = \log\left(\frac{E_1/T_1}{E_2/T_2}\right) \quad (3.13)$$

La varianza del $\log(IRR)$ se estima bajo el supuesto de que el número de eventos en cada grupo sigue una distribución de Poisson, lo que implica que la varianza es igual a la media. Este supuesto es razonable cuando los eventos son relativamente poco frecuentes y los tiempos de observación son suficientemente largos, como ocurre habitualmente en estudios longitudinales.

$$Var(\log(IRR)) = \frac{1}{E_1} + \frac{1}{E_2} \quad (3.14)$$

Una IRR igual a 1 indica que las tasas de incidencia son iguales en ambos grupos. Valores superiores a 1 sugieren una mayor frecuencia del evento en el grupo de tratamiento, mientras que valores inferiores a 1 indican una menor tasa.

3.2. Modelo de combinación de estudios

La mejor estimación del efecto global requiere considerar la posible heterogeneidad entre los estudios incluidos, es decir, las diferencias que pueden existir en sus características metodológicas, contextuales o poblacionales. Para abordar esta variabilidad, se recurre al uso de modelos estadísticos específicos, conocidos como modelos de efectos.

Los modelos utilizados en el MA para combinar los resultados de diferentes estudios se dividen principalmente en dos categorías: el modelo de efectos fijos y el modelo de efectos aleatorios. Ambos enfoques difieren en sus supuestos sobre la naturaleza del efecto verdadero que subyace a los estudios. Mientras que el modelo de efectos fijos asume que todos los estudios estiman un único y mismo efecto verdadero, el modelo de efectos aleatorios reconoce la existencia de una distribución de efectos verdaderos, producto de la variabilidad inherente entre estudios.

3.2.1. Modelo de efectos fijos

El modelo de efectos fijos asume que todos los estudios incluidos en un MA comparten un único efecto verdadero común, denotado θ . Considera que las diferencias observadas entre las estimaciones de los distintos estudios se deben exclusivamente al error aleatorio asociado al muestreo ε_k . Así, la estimación del efecto en el estudio k , $\hat{\theta}_k$:

$$\hat{\theta}_k = \theta + \varepsilon_k \quad (3.15)$$

Bajo este modelo, se asume que cualquier diferencia entre los resultados de los estudios no se debe a variaciones reales entre ellos, sino únicamente al error de muestreo, es decir, a la variabilidad que ocurre por azar al seleccionar diferentes muestras. Para los estudios con menor varianza muestral (mayor precisión) habrá mayor influencia en la estimación del efecto combinado.

En el modelo de efectos fijos el efecto combinado $\hat{\theta}$ se calcula como una media ponderada de los efectos individuales θ_k , donde los pesos ω_k son el inverso de la varianza de cada estudio:

$$\omega_k = \frac{1}{s_k^2} \quad (3.16)$$

$$\hat{\theta} = \frac{\sum_{k=1}^K \hat{\theta}_k \omega_k}{\sum_{k=1}^K \omega_k} \quad (3.17)$$

Y la varianza del efecto combinado es:

$$Var(\hat{\theta}) = \frac{1}{\sum_{k=1}^K \omega_k} \quad (3.18)$$

3.2.2. Modelo de efectos aleatorios

El modelo de efectos aleatorios es más flexible que el de efectos fijos y resulta más apropiado cuando existe evidencia de heterogeneidad sustancial entre estudios. La estimación obtenida puede interpretarse como una media de los efectos verdaderos distribuidos en una población de estudios, lo que permite generalizar los resultados más allá del conjunto concreto de estudios incluidos.

El modelo de efectos aleatorios parte del supuesto de que los estudios no comparten un efecto común θ , sino que cada estudio tiene su propio efecto verdadero θ_k . Debido a esto, se introduce una nueva fuente de variación ζ_k , que representa la desviación del efecto verdadero del estudio k respecto a la media global μ :

$$\theta_k = \mu + \zeta_k \quad (3.19)$$

donde ζ_k es una variable aleatoria con media 0 y varianza τ^2 , que cuantifica la heterogeneidad entre los efectos verdaderos de los diferentes estudios. La estimación observada en cada estudio puede expresarse como:

$$\hat{\theta}_k = \mu + \zeta_k + \varepsilon_k \quad (3.20)$$

Sabiendo que ε_k es el error aleatorio de muestreo, que puede estimarse con s_k^2 . Así, la varianza total de cada estimación observada es:

$$Var(\hat{\theta}_k) = s_k^2 + \tau^2 \quad (3.21)$$

y los pesos utilizados bajo este modelo se ajustan a esta varianza total:

$$\omega_k^* = \frac{1}{s_k^2 + \tau^2} \quad (3.22)$$

y el efecto global se calcula como una media ponderada:

$$\hat{\theta} = \frac{\sum_{k=1}^K \hat{\theta}_k \omega_k^*}{\sum_{k=1}^K \omega_k^*} \quad (3.23)$$

La varianza del efecto combinado es:

$$Var(\hat{\theta}) = \frac{1}{\sum_{k=1}^K \omega_k^*} \quad (3.24)$$

Existen diversos métodos para estimar τ^2 , entre los que destacan: el estimador de DerSimonian-Laird [DerSimonian and Laird, 1986], de máxima verosimilitud [Viechtbauer, 2005], Paule-Mandel [Paule and Mandel, 1982], Empirical Bayes [Sidik and Jonkman, 2019] y Sidik-Jonkman [Sidik and Jonkman, 2005]. En el apartado 3.3 se abordará con mayor detalle la evaluación de la heterogeneidad entre estudios.

3.3. Evaluación de la heterogeneidad

En un MA, la heterogeneidad se refiere a las diferencias que existen entre los resultados de los estudios incluidos. Estas variaciones pueden deberse a diferencias metodológicas, poblacionales o contextuales. Aunque cierto grado de heterogeneidad es esperable, una heterogeneidad excesiva puede afectar la validez y aplicabilidad de las conclusiones del análisis conjunto. Cuando esta heterogeneidad resulta estadísticamente significativa, suele ser apropiado emplear un modelo de efectos aleatorios.

Una primera aproximación a la heterogeneidad puede obtenerse de forma visual mediante los diagramas de bosque (apartado 3.5.1), donde cada estudio se representa como una estimación puntual del efecto y su intervalo de confianza. Cuando los intervalos muestran poca o nula superposición, se sugiere la presencia de heterogeneidad estadística. Sin embargo, esta evaluación visual debe ser siempre complementada con medidas cuantitativas específicas [Chandler et al., 2019].

3.3.1. Estadístico Q de Cochran

El estadístico Q de Cochran se emplea para distinguir entre el error muestral ε_k y la heterogeneidad entre estudios ζ_k . Este se calcula como la suma de cuadrados ponderada, que mide la desviación de cada efecto estimado respecto al efecto global:

$$Q = \sum_{k=1}^K \omega_k (\hat{\theta}_k - \hat{\theta})^2 \quad (3.25)$$

donde ω_k los pesos del modelo, $\hat{\theta}_k$ es la estimación del estudio k y $\hat{\theta}$ es la estimación global combinada.

El valor de Q se compara con los valores de una distribución chi-cuadrado con k-1 grados de libertad. Cuando el valor calculado supera significativamente el valor esperado bajo el supuesto de homogeneidad, esto proporciona evidencia estadística para concluir que existe heterogeneidad significativa entre los estudios.

El estadístico Q de Cochran presenta importantes limitaciones. Su potencia estadística

depende críticamente del número de estudios incluidos y del tamaño muestral de estos, lo que significa que en MA con muchos estudios o de gran tamaño puede detectar heterogeneidad incluso cuando las diferencias reales entre estudios son mínimas. Por esta razón, no se recomienda utilizar únicamente su significación estadística para decidir entre modelos de efectos fijos o aleatorios.

3.3.2. Estadístico I^2 y H^2 de Higgins y Thompson

El estadístico I^2 propuesto por Higgins y Thompson [Higgins and Thompson, 2002], es otra medida utilizada en MA para cuantificar la heterogeneidad entre estudios. Representa el porcentaje de variabilidad total en los resultados que no se debe al azar, sino a diferencias reales entre los estudios. Se calcula a partir de Q mediante la siguiente expresión:

$$I^2 = \frac{Q - (k - 1)}{Q} \quad (3.26)$$

Este índice toma valores entre 0 y 1, donde valores más altos indican mayor heterogeneidad [Higgins et al., 2003]. A diferencia de Q , I^2 es una medida estandarizada que permite comparaciones entre MA de diferente tamaño o número de estudios.

- $I^2 < 25\% \rightarrow$ Heterogeneidad baja
- $25\% \leq I^2 < 50\% \rightarrow$ Heterogeneidad moderada
- $50\% \leq I^2 < 75\% \rightarrow$ Heterogeneidad sustancial
- $I^2 \geq 80\% \rightarrow$ Heterogeneidad considerable

Aún así, la interpretación del estadístico I^2 requiere cierta cautela. Su precisión depende del tamaño y número de estudios incluidos, pudiendo ser inestable y subestimar la heterogeneidad real en MA con pocos estudios o tamaños muestrales pequeños. Además, un I^2 alto no indica necesariamente un problema, sino simplemente que los estudios varían más de lo esperado por el azar.

El estadístico H^2 también propuesto por Higgins y Thompson [Higgins and Thompson, 2002] se deriva directamente del Q de Cochran y describe la relación entre la variación observada y la variación esperada solo por el error de muestreo.

$$H^2 = \frac{Q}{K - 1} \quad (3.27)$$

Un valor de $H^2 = 1$ indica ausencia de heterogeneidad, mientras que valores crecientes reflejan mayor variabilidad entre estudios.

A diferencia del estadístico I^2 , existen puntos de corte ampliamente aceptados para interpretar los valores de H^2 .

3.3.3. Estimación de la varianza entre estudios (τ^2)

En el modelo de efectos aleatorios, la varianza total asociada a la estimación de cada estudio incorpora tanto el error muestral como la varianza entre estudios τ^2 , que cuantifica la heterogeneidad real entre los efectos verdaderos subyacentes. La estimación precisa de τ^2 resulta crucial para ponderar adecuadamente los estudios en el análisis combinado.

El método más utilizado para estimar τ^2 es el propuesto por DerSimonian y Laird (1986) [DerSimonian and Laird, 1986], debido a su simplicidad computacional. Este estimador se basa en el estadístico Q de Cochran y se calcula como:

$$\hat{\tau}^2 = \frac{Q - (K - 1)}{\sum(w_k) - \frac{\sum(w_k^2)}{\sum(w_k)}}, \text{ si } Q > (K - 1) \quad (3.28)$$

donde K es el número de estudios, y $w_k = 1/s_k^2$ son los pesos derivados del modelo de efectos fijos. Este cálculo solo sirve cuando $Q > (K - 1)$. En caso contrario ($Q \leq K - 1$), se asume $\tau^2 = 0$, indicando ausencia de heterogeneidad detectable.

Existen métodos alternativos para estimar τ^2 , como la máxima verosimilitud (ML), la máxima verosimilitud restringida (REML) o enfoques bayesianos. Estos métodos pueden ofrecer estimaciones más precisas o ajustadas en situaciones específicas, pero requieren mayor complejidad computacional. En la práctica, el método de DerSimonian y Laird es muy utilizado y proporciona resultados adecuados en la mayoría de los casos.

Una vez estimado τ^2 , este valor se incorpora a la varianza total de cada estudio:

$$Var(\hat{\theta}_k) = s_k^2 + \hat{\tau}^2 \quad (3.29)$$

y se utiliza para calcular los pesos del modelo de efectos aleatorios:

$$w_k = 1/(s_k^2 + \hat{\tau}^2) \quad (3.30)$$

$$w_k = \frac{1}{s_k^2 + \hat{\tau}^2} \quad (3.31)$$

Este ajuste permite tener en cuenta la heterogeneidad entre estudios al combinar los resultados, otorgando menos peso a los estudios cuya precisión se ve reducida por la variabilidad entre efectos.

3.4. Evaluación del sesgo de publicación

El objetivo de la evaluación del sesgo de publicación en un MA es identificar si los resultados publicados reflejan fielmente toda la evidencia disponible o si están distorsionados porque los estudios con resultados negativos o no concluyentes tienden a no publicarse. Este sesgo puede llevar a sobreestimar la magnitud del efecto y a extraer conclusiones incorrectas, por lo que su detección es fundamental para la validez de los resultados del MA.

3.4.1. Gráfico de embudo (*funnel plot*)

El gráfico de embudo (*funnel plot*) es una herramienta visual que ayuda a detectar el sesgo de publicación. Consiste en un diagrama de dispersión que muestra visualmente la distribución de la medida del efecto estimada en función de la precisión de cada estudio, siendo especialmente útil para detectar posibles efectos de estudios pequeños y sesgos de publicación.

En un gráfico de embudo (Figura 3.1A) sin sesgo de publicación, los estudios tanto de muestras grandes como pequeñas se distribuyen de manera simétrica y homogénea a ambos lados del resultado del MA, formando un embudo simétrico e invertido. En cambio, un gráfico de embudo donde los estudios se distribuyen de forma asimétrica (Figura 3.1B) indica que probablemente faltan estudios y que el resultado del MA podría estar sesgado. Sin embargo, la interpretación visual es subjetiva, por lo que se recomienda complementarla con pruebas estadísticas.

3.4.2. Prueba de Egger

La prueba de Egger [Egger et al., 1997] es un método estadístico que evalúa si existe asimetría en el gráfico de embudo, lo que sugeriría la presencia de sesgo de publicación.

Esta prueba evalúa si la relación entre el tamaño del efecto y la precisión de los estudios es la esperada en ausencia de sesgo. Si el resultado es significativo, indica que los estudios pequeños tienden a mostrar efectos diferentes a los estudios grandes, lo que puede deberse a la no publicación de estudios con resultados negativos.

Es importante tener en cuenta que la prueba de Egger puede perder potencia cuando el número de estudios es bajo y puede verse afectada por la heterogeneidad entre ellos, por lo que se recomienda utilizarla junto con otros métodos.

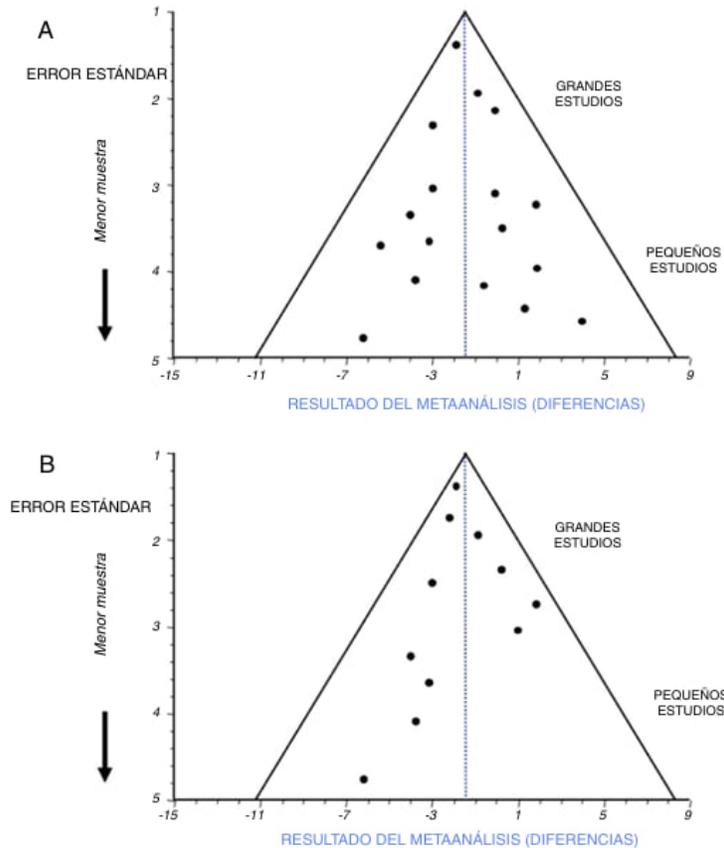


Figura 3.1: *Funnel plot* para examinar la existencia de sesgo de publicación.

3.4.3. Prueba de Peters

La prueba de Peters [Peters et al., 2006] es una alternativa a la prueba de Egger, especialmente recomendada cuando se analizan medidas de efecto binarias, como el riesgo relativo u odds ratio. El modelo de la prueba de Peters emplea una regresión ponderada que relaciona el efecto en escala logarítmica, donde ψ_k representa la medida de efecto del estudio, con el inverso del tamaño muestral del estudio k .

$$\log \psi_k = \beta_0 + \beta_1 \frac{1}{n_k} + \varepsilon_K \quad (3.32)$$

$$\omega_k = \frac{1}{\left(\frac{1}{n_{11}+n_{21}} + \frac{1}{n_{12}+n_{22}}\right)} \quad (3.33)$$

La lógica es que, si hay sesgo de publicación, los estudios pequeños y con menos eventos tenderán a mostrar efectos diferentes a los estudios grandes. Si la prueba detecta una relación significativa, sugiere la presencia de asimetría y, por tanto, de sesgo de publicación.

La prueba de Peters es menos propensa a falsos positivos en estudios con variables binarias y es recomendada en estos contextos.

3.4.4. Método de Recorte y Relleno (*trim and fill*)

El método de Recorte y Relleno (*trim and fill*) [Duval and Tweedie, 2000] sirve para estimar el tamaño real del efecto en un MA corrigiendo la asimetría del gráfico de embudo, un signo que puede indicar tanto efectos de estudios pequeños como posible sesgo de publicación.

La idea central del método es añadir los efectos faltantes hasta que el gráfico de embudo se vuelva simétrico. Esto se logra a través de un algoritmo que consta de dos pasos principales:

1. Recorte (*Trimming*): El método identifica y recorta los estudios atípicos en el gráfico de embudo, es decir, aquellos pequeños estudios que se encuentran dispersos en un lado del gráfico. Una vez identificados, estos estudios se eliminan del análisis y se recalcula el efecto combinado sin ellos.
2. Relleno (*Filling*): Por cada estudio recortado, se añade un estudio “reflejado” en el lado opuesto del gráfico. Finalmente, el efecto promedio se recalcula incluyendo todos los estudios originales y los imputados, obteniendo así una estimación corregida del efecto global.

Una limitación clave del método de recorte y relleno es que pierde fiabilidad cuando existe una alta heterogeneidad entre los estudios (por ejemplo, valores elevados de I^2). En estas situaciones, incluso los estudios grandes pueden alejarse mucho del efecto promedio, lo que puede provocar que el proceso de recorte y relleno introduzca datos estimados de manera incorrecta y distorsione la validez de los resultados corregidos.

3.4.5. Análisis de subgrupos

El análisis de subgrupos es una técnica estadística que permite investigar patrones de heterogeneidad entre los estudios incluidos y determinar si ciertas características de los estudios o de los participantes (p.e., edad, gravedad de la enfermedad, tipo de intervención) modifican el efecto observado. Su objetivo es identificar si existen diferencias significativas en el efecto entre subgrupos definidos por estas características, proporcionando así una interpretación más precisa y útil de los resultados globales.

Este enfoque parte de la hipótesis de que los estudios pueden agruparse en subgrupos definidos por una característica específica y que cada subgrupo puede tener su propio

efecto global. Habitualmente, se utiliza un modelo de efectos aleatorios para combinar los resultados dentro de cada subgrupo, reflejando la variabilidad interna, mientras que los subgrupos se consideran efectos fijos. El análisis de subgrupos se desarrolla en dos fases principales:

1. Agrupación de efectos dentro de subgrupos: Se calcula un efecto global para cada subgrupo, generalmente utilizando un modelo de efectos aleatorios, reflejando la variabilidad interna.
2. Comparación entre subgrupos: Se emplean pruebas estadísticas, como la prueba Q entre subgrupos, para evaluar si existen diferencias significativas entre los efectos combinados de los subgrupos, lo que sugeriría que la característica analizada explica parte de la heterogeneidad

El análisis de subgrupos debe planificarse antes de realizar el MA para evitar resultados debidos al azar. Además, es recomendable que los subgrupos cuenten con un tamaño suficiente de estudios para garantizar la validez estadística de los resultados y evitar conclusiones erróneas por falta de potencia.

3.5. Visualización de resultados

3.5.1. Gráficos de Bosque (*Forest Plots*)

Los gráficos de bosque (*Forest Plots*), como el ilustrado en la figura 3.2, son la herramienta visual central en los MA para presentar y sintetizar los resultados de múltiples estudios sobre un mismo tema. Permiten analizar de manera clara y rápida tanto el efecto de cada estudio individual como el efecto combinado global.

El gráfico de bosque representa visualmente el tamaño del efecto de cada estudio mediante un cuadrado en el centro del diagrama, cuyo tamaño refleja el peso del estudio en el MA. Una línea horizontal atraviesa los cuadrados, mostrando el intervalo de confianza de los efectos individuales de cada estudio. Además, el gráfico incluye los datos numéricos del tamaño del efecto para facilitar la replicación. En la parte inferior, un rombo ilustra el efecto combinado y su intervalo de confianza. Una línea vertical marca el punto de ausencia de efecto en el eje x, ayudando a interpretar si los resultados son estadísticamente significativos.

3.5.2. Diagrama de cortina (*Drapery Plots*)

Los diagramas de cortina *Drapery Plots* [Rücker and Schwarzer, 2021], como el ilustrado en la figura 3.3, muestran las curvas de p-valores de cada uno de los estudios individuales

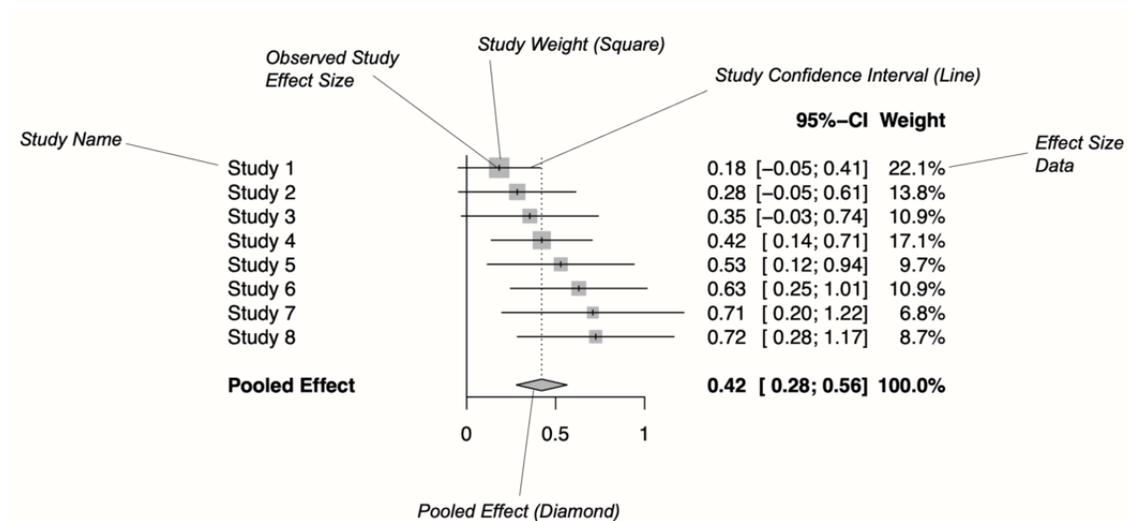


Figura 3.2: *Forest Plots*

incluidos en un MA, las estimaciones combinadas o agrupadas y una región de predicción para un estudio futuro individual. De manera similar al *forest plot*, el *drapery plot* permite visualizar la información principal de un MA de comparaciones pareadas, integrando tanto los resultados individuales como el efecto global y la incertidumbre sobre futuros estudios.

La posición de un estudio respecto al eje de precisión indica cuánto peso tiene en el MA, mientras que si un estudio aparece fuera de las bandas de confianza, sugiere que contribuye significativamente a la heterogeneidad de los resultados. Los diagramas de cortina ayudan así a interpretar de manera visual la robustez y coherencia de los resultados combinados en un MA.

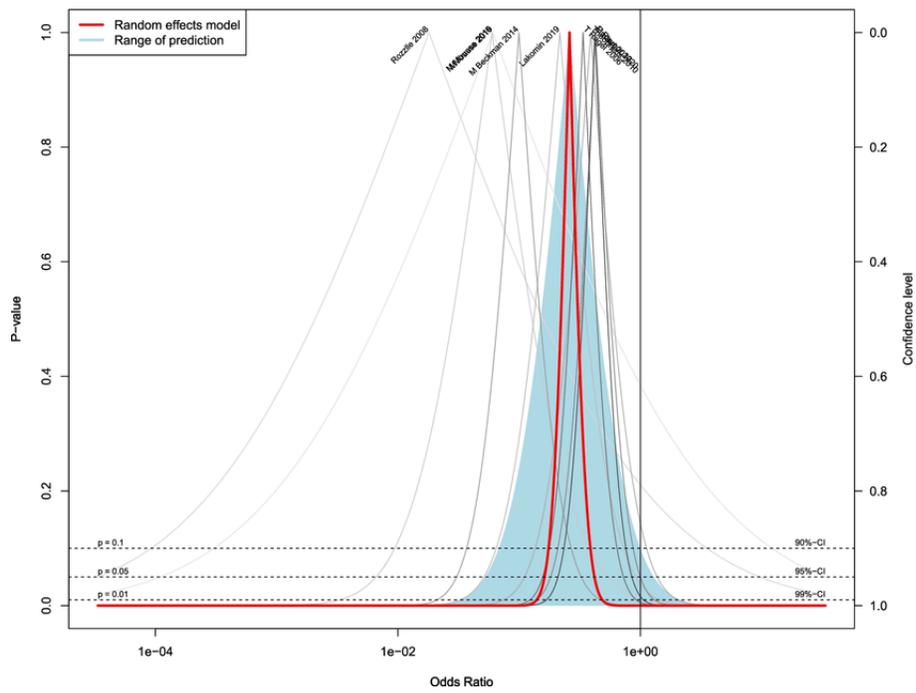


Figura 3.3: Diagrama de cortina (*Drapery Plots*)

3.6. Paquetes de R

En el desarrollo de la metodología del MA se pueden utilizar distintas librerías especializadas del entorno R. A continuación, se proporciona una breve descripción de su funcionamiento y aplicaciones principales. La información presentada se ha extraído de la documentación oficial de cada paquete:

- **{tidyverse}** [Wickham et al., 2019] es una colección de paquetes disponibles en R y orientados a la manipulación, importación, exploración y visualización de datos y que se utiliza exhaustivamente en ciencia de datos. El uso de tidyverse permite facilitar el trabajo estadístico y la generación de trabajos reproducibles. Está compuesto de los siguientes paquetes: readr, dplyr, ggplot2, tibble, tidyr, purr, stringr, forcats.
- **{meta}** [Balduzzi et al., 2019] es un paquete diseñado específicamente para la realización de MA clásicos, tanto de variables dicotómicas como continuas. Ofrece funciones para el cálculo de tamaños de efecto, análisis de subgrupos, gráficos de bosque y gráficos de embudo, entre otros.
- **{metafor}** [Viechtbauer, 2010] es un paquete avanzado para MA que permite ajustar modelos de efectos fijos, aleatorios y mixtos, calcular diferentes tamaños de efecto y realizar análisis de meta-regresión, sensibilidad y visualización avanzada de resultados.
- **{dmeta}** [Harrer et al., 2019] es un paquete diseñado para facilitar la realización de MA, especialmente en combinación con los paquetes meta y metafor. Fue desarrollado como apoyo a la guía “Doing Meta-Analysis in R” [Harrer et al., 2021] y proporciona funciones adicionales que simplifican etapas clave del proceso. El paquete está orientado a investigadores en biomedicina y psicología, aunque sus herramientas son aplicables a otras áreas. Muchas funciones de dmeta interactúan directamente con las funciones de meta y metafor para mejorar el flujo de trabajo y la interpretación de resultados.

El empleo de R y sus paquetes especializados asegura la reproducibilidad y trazabilidad del MA, ya que todo el proceso, desde la importación y preparación de los datos hasta el análisis estadístico y la visualización de los resultados, puede ser documentado y ejecutado mediante *scripts*. Esta metodología permite que otros investigadores puedan replicar el análisis, verificar los resultados obtenidos y adaptar el código a futuras investigaciones, cumpliendo así con los estándares internacionales de transparencia y buenas prácticas científicas.

Capítulo 4

Aplicación a la Dermatitis Atópica

4.1. Motivación

La DA es una enfermedad inflamatoria crónica de la piel que afecta a pacientes de todas las edades y presenta una elevada heterogeneidad clínica y terapéutica. En la última década, los avances en la comprensión de la fisiopatología de la DA han permitido el desarrollo de tratamientos, entre los que destacan los agentes biológicos y los inhibidores de JAK. Los agentes biológicos, como el dupilumab, el lebrikizumab y el tralokinumab, son anticuerpos monoclonales que bloquean selectivamente las vías inflamatorias clave, especialmente las vías de señalización de interleucinas como la IL-4 y la IL-13. Por otro lado, los inhibidores de JAK, como el baricitinib, el upadacitinib y el abrocitinib, actúan a nivel intracelular bloqueando la señalización de múltiples citoquinas implicadas en la inflamación y el prurito, lo que les confiere un inicio de acción rápido y una eficacia significativa en DA moderada a grave.

Ambas clases terapéuticas han demostrado eficacia significativa en el control de los síntomas, aunque no curativos, proporcionando opciones eficaces para pacientes que no responden o no toleran las terapias convencionales. Sin embargo, existen diferencias en la magnitud de la respuesta, la rapidez de acción y el perfil de seguridad entre estos tratamientos y frente a placebo. Además, la heterogeneidad en los diseños de los ensayos, las poblaciones estudiadas y los desenlaces evaluados dificulta la comparación directa entre terapias, lo que justifica la necesidad de un MA riguroso.

4.2. Identificación y selección de estudios

Se realizó una búsqueda sistemática en la base de datos PubMed, abarcando los últimos 10 años hasta mayo de 2025. La búsqueda se enfocó en identificar ensayos clínicos aleatorizados (RCTs, *Randomized Controlled Trials*) que evaluaran la eficacia y seguridad de estas intervenciones. La estrategia de búsqueda se diseñó conforme al marco PICO, con los siguientes componentes:

- P (Población): Adultos con dermatitis atópica moderada a grave.
- I (Intervención): Agentes biológicos (dupilumab, tralokinumab, etc.) o inhibidores de JAK (upadacitinib, baricitinib, etc.) en monoterapia.
- C (Comparador): Placebo.
- O (Outcome - Desenlaces):
 - Principal: Proporción de pacientes con EASI-75.
 - Secundario: IGA 0/1 y frecuencia de eventos adversos.

Para la eficacia clínica, se consideró como variable principal el EASI-75, que corresponde al número de pacientes que lograron al menos un 75 % de mejora en el EASI (*Eczema Area and Severity Index*) que mide la gravedad y extensión del eccema en un rango de 0 a 72. Y como variable secundaria, la IGA (*Investigator Global Assessment*) con una escala de 0 a 4, centrándose en los pacientes que alcanzaron una puntuación de 0 (clara) o 1 (casi clara). En cuanto a la seguridad, se recogió el número de eventos adversos desde el inicio del tratamiento (TEAE, *Treatment emergent adverse events*). Se incluyeron estudios que reportaban resultados principales entre las 12 y 16 semanas y al menos uno de los desenlaces de interés.

La búsqueda sistemática se llevó a cabo en PubMed utilizando combinaciones de términos clave y operadores booleanos relacionados con “dermatitis atópica”, “eczema”, “biológicos”, “inhibidores JAK”, “placebo”, “EASI”, “IGA” y “TEAE”. Se identificaron un total de 136 estudios a través de la base de datos PubMed. Durante la fase inicial de cribado, se eliminaron únicamente 2 estudios repetidos. Después, por el título y el resumen de los estudios se excluyeron 92 estudios: 80 por no ser RCTs y 12 por tratar temas no relacionados. Finalmente, 42 estudios pasaron a revisión a texto completo de los cuales se excluyeron 15 estudios por el uso complementario de corticoides tópicos (TCS), 10 estudios por la duración de los tratamientos y 4 estudios por no reportar ninguna de las variables deseadas. Este proceso de identificación y selección de estudios se puede observar en la figura 4.1.

Finalmente, se incluyeron 13 estudios originales, aunque el análisis final incorporó 18 RCTs debido a que cinco de los estudios originales reportaron dos RCTs independientes

Estudio	Edad	Periodo	N	Intervenciones	Duración	Desenlaces
[Guttman-Yassky et al., 2021]	Todas las edades	2018-2020	1683	Upadacitinib 15mg Upadacitinib 30mg	16 semanas	EASI75,IGA TEAE
[Simpson et al., 2021]	≥ 18 años	2018-2019	440	Baricitinib 1mg Baricitinib 2mg	16 semanas	EASI75,IGA TEAE
[Bieber et al., 2021]	≥ 18 años	2018-2019	837	Dupilumab 300mg Abrocitinib 200mg Abrocitinib 100mg	16 semanas	EASI75,IGA TEAE
[Simpson et al., 2020b]	Todas las edades	2017-2019	387	Abrocitinib 200mg Abrocitinib 100mg	12 semanas	EASI75,IGA TEAE
[Schmid-Grendelmeier et al., 2024]	≥ 12 años	NA	444	Abrocitinib 200mg Abrocitinib 100mg	12 semanas	EASI75,IGA TEAE
[Ständer et al., 2023]	≥ 18 años	NA	837	Abrocitinib 200mg Abrocitinib 100mg	12 semanas	EASI75,IGA
[Simpson et al., 2016]	39 ± 11,5 en control 38 ± 10 en trat	NA	1379	Dupilumab 300mg Abrocitinib 100mg	16 semanas	EASI75,IGA TEAE
[Silverberg et al., 2023]	≥ 12 años	2019-2021	851	Lebrikizumab 250mg	52 semanas	EASI75,IGA TEAE
[Blauvelt et al., 2023]	33 ± 16,6 en control 36,1 ± 17 en trat	2019-2022	291	Lebrikizumab 250mg	52 semanas	EASI75,IGA
[Wollenberg et al., 2021]	≥ 18 años	2017-2018	1593	Tralokinumab	52 semanas	EASI75,IGA TEAE
[Simpson et al., 2020a]	≥ 12 años	2017	251	Dupilumab 300mg	16 semanas	EASI75,IGA TEAE
[Paller et al., 2023]	12-17 años	2018-2021	289	Tralokinumab 150mg Tralokinumab 300mg	52 semanas	EASI75,IGA TEAE
[Guttman-Yassky et al., 2020]	≥ 18 años	2018-2019	280	Lebrikizumab 250mg	16 semanas	EASI75,IGA TEAE

Tabla 4.1: Características de los estudios incluidos

NA: Dato no disponible en los estudios originales.

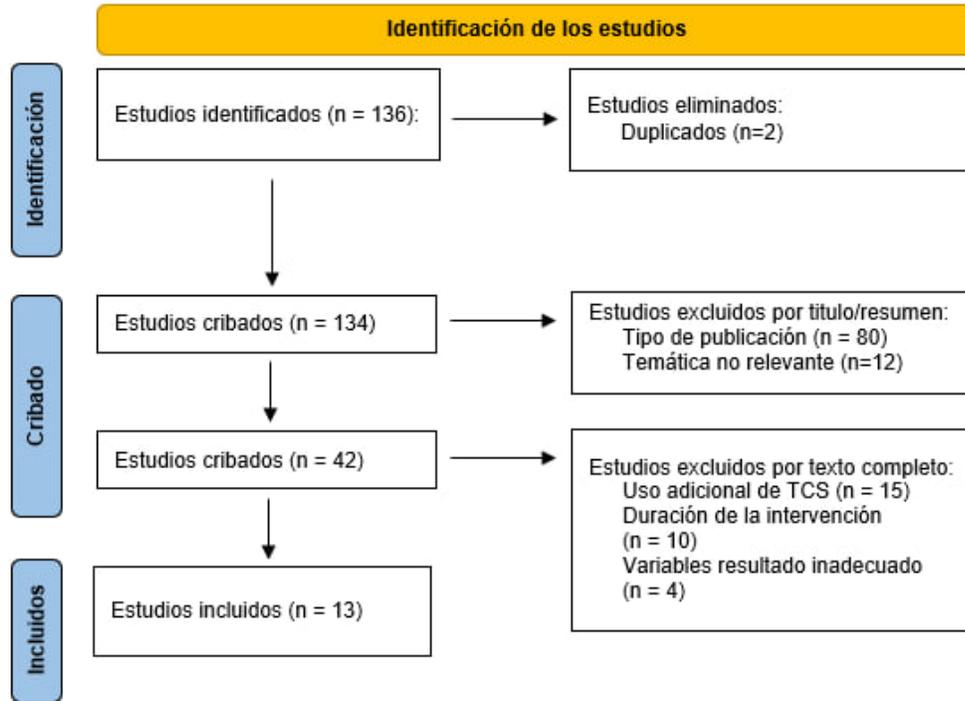


Figura 4.1: Diagrama de flujo para la identificación y selección de estudios relevantes

cada uno. Estos fueron diferenciados mediante la denominación “Estudio a” y “Estudio b” para mantener la trazabilidad de los datos. Además, en aquellos estudios que comparaban varias dosis del medicamento frente al mismo grupo placebo, todas las dosis fueron agrupadas. De este modo, se sumaron tanto los eventos como los tamaños muestrales correspondientes, permitiendo así evaluar el efecto global del medicamento frente al placebo de manera más robusta y evitando la duplicación de controles. Un total de 8,446 pacientes participaron en los estudios incluidos, cuyas características se encuentran resumidas en la tabla 4.1.

4.3. Estimación de tamaños del efecto individuales y pesos

Los tamaños de efecto individuales se han calculado con ayuda del paquete {meta} en R, se empleará el OR como medida de efecto para los desenlaces dicotómicos (EASI-75, IGA 0/1 y TEAE).

Para la estimación del efecto combinado de los OR y sus intervalos de confianza del 95 %, se empleó el método de varianza inversa, en el que los estudios se ponderan según el inverso de la varianza de sus estimaciones, tal y como se describió en el apartado 3.2. Además, dada la limitada cantidad de estudios y la posible heterogeneidad clínica, se

optó por un modelo de efectos aleatorios, lo que permite obtener una estimación más conservadora y generalizable del efecto combinado.

Los pesos asignados a cada estudio son inversamente proporcionales a la varianza de su estimación del OR, de modo que los estudios más precisos contribuyen en mayor medida al cálculo del efecto combinado.

4.4. Estimación del efecto combinado y análisis de la heterogeneidad

La Figura 4.2 presenta mediante un *forest plot*, los OR individuales y sus intervalos de confianza del 95 % para la variable EASI-75 en cada uno de los 18 RTCs incluidos (N=8,446 pacientes), junto con los pesos asignados en el MA. El análisis bajo el modelo de efectos aleatorios mostró un OR global de 5.77 (IC 95 %: 4.51-7.38) indicando que los tratamientos evaluados aumentan significativamente la probabilidad de alcanzar una mejoría del 75 % en el índice EASI respecto al control, con un $p < 0.0001$. Sin embargo, se detectó una alta heterogeneidad global ($I^2 = 76.5\%$, $\tau^2 = 0.19$, $Q = 72.41$; $p < 0.0001$), lo que sugiere una considerable variabilidad entre los estudios incluidos.

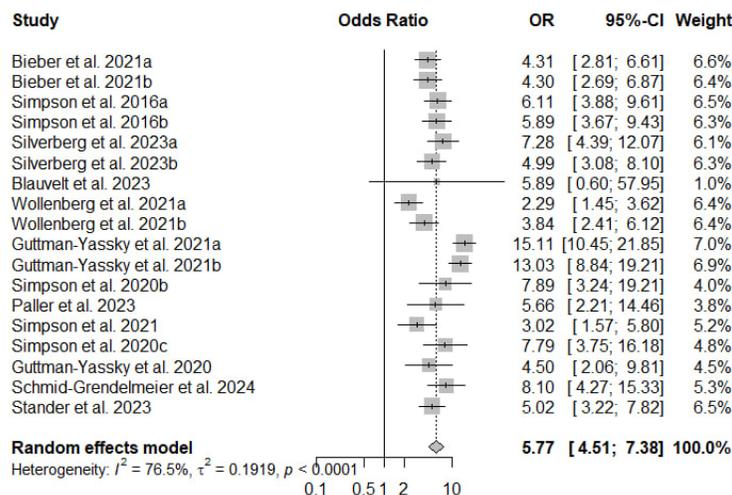


Figura 4.2: *Forest plot* de la proporción de pacientes que alcanzaron EASI-75

La Figura 4.3 muestra el *forest plot* para la variable IGA 0/1, donde el análisis de los mismos 18 estudios reveló un OR agrupado de 4.51 (IC 95 %: 3.22-6.32; $p < 0.0001$), lo que equivale a un aumento de 4.5 veces en la probabilidad de alcanzar remisión clínica con los tratamientos activos frente al control. La heterogeneidad fue aún mayor que para EASI-75 ($I^2 = 81.0\%$, $\tau^2 = 0.41$, $Q = 89.49$; $p < 0.0001$).

La Figura 4.4 presenta el *forest plot* del número de TEAE. El análisis de 16 RTCs (N=7,765 pacientes) no mostró una asociación significativa entre los tratamientos para

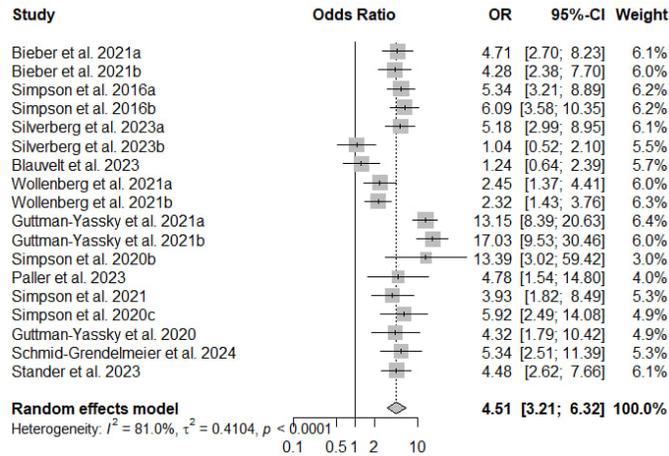


Figura 4.3: *Forest plot* de la proporción de pacientes que alcanzaron IGA 0/1 (remisión clínica)

DA y un mayor riesgo global de TEAE (OR 1.11, IC 95 %: 0.94-1.31; $p = 0.24$). Sin embargo, se observó heterogeneidad moderada-alta entre los estudios ($I^2 = 61.2\%$, $\tau^2 = 0.069$, $p = 0.0007$).

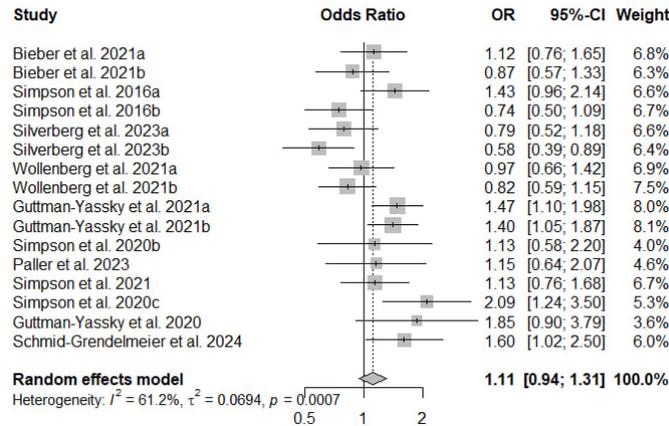


Figura 4.4: *Forest plot* de la proporción de pacientes que obtuvieron al menos un TEAE

4.5. Análisis por subgrupos

Con el objetivo de explorar posibles fuentes de heterogeneidad y obtener una comprensión más detallada de los resultados, se realizó un análisis de subgrupos diferenciando entre el tipo de medicamento utilizado, específicamente si se trataba de un agente biológico o un inhibidor de JAK.

Para la variable EASI-75, ambos subgrupos mostraron una eficacia significativamen-

te superior al comparador, aunque se observaron diferencias en la magnitud del efecto. Los agentes biológicos presentaron un OR de 4.84 (IC 95 %: 3.81–6.14), mientras que los inhibidores de JAK alcanzaron un OR de 7.13 (IC 95 %: 4.38–10.77), como se detalla en el *forest plot* de la Figura 4.5. Sin embargo, la diferencia entre ambos subgrupos no fue significativa ($p = 0.1336$).

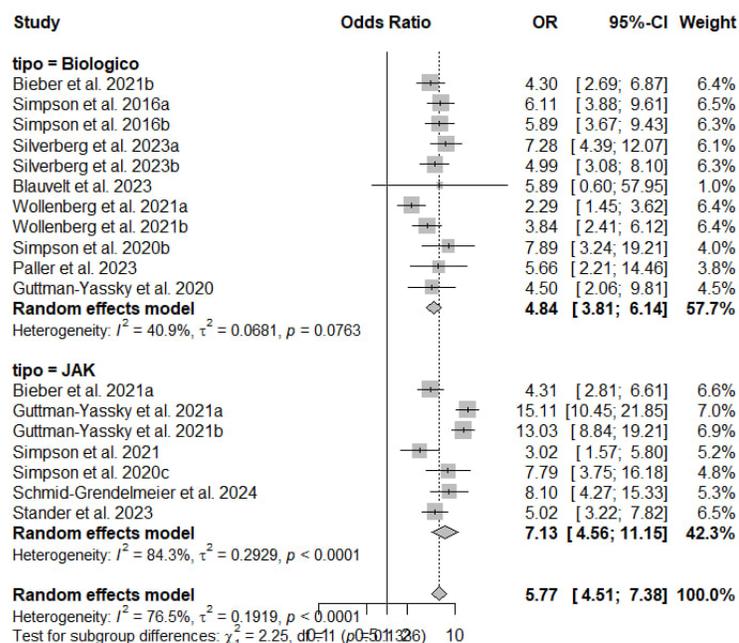


Figura 4.5: *Forest plot* del subanálisis por tipo de medicamento sobre la proporción de pacientes que alcanzaron EASI-75.

En el caso de la variable IGA 0/1, los resultados del *forest plot* de la Figura 4.6 demostraron que los inhibidores de JAK mostraron una mayor eficacia en comparación con los agentes biológicos, con un OR de 3.34 (IC 95 %: 2.28–5.06) frente a un OR de 6.88 (IC 95 %: 4.38–10.77) para los agentes biológicos. En este caso, la diferencia entre subgrupos sí fue significativa ($p = 0.0213$), lo que sugiere una posible superioridad de los inhibidores de JAK en este desenlace.

Respecto a la seguridad, evaluada mediante la TEAE, los agentes biológicos no mostraron un aumento significativo del riesgo respecto al control (OR = 0.93; IC 95 %: 0.77–1.12), mientras que los inhibidores de JAK sí presentaron un incremento significativo (OR = 1.40; IC 95 %: 1.20–1.62), como se visualiza en la Figura 4.7. La diferencia entre subgrupos fue significativa ($p = 0.0008$), lo que indica un perfil de seguridad menos favorable para los inhibidores de JAK.

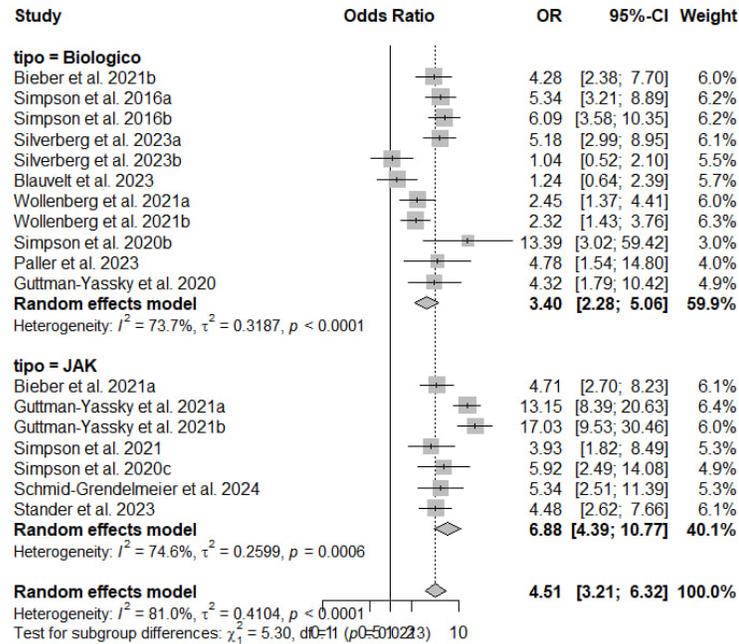


Figura 4.6: *Forest plot* del subanálisis por tipo de medicamento sobre la proporción de pacientes que alcanzaron IGA 0/1.

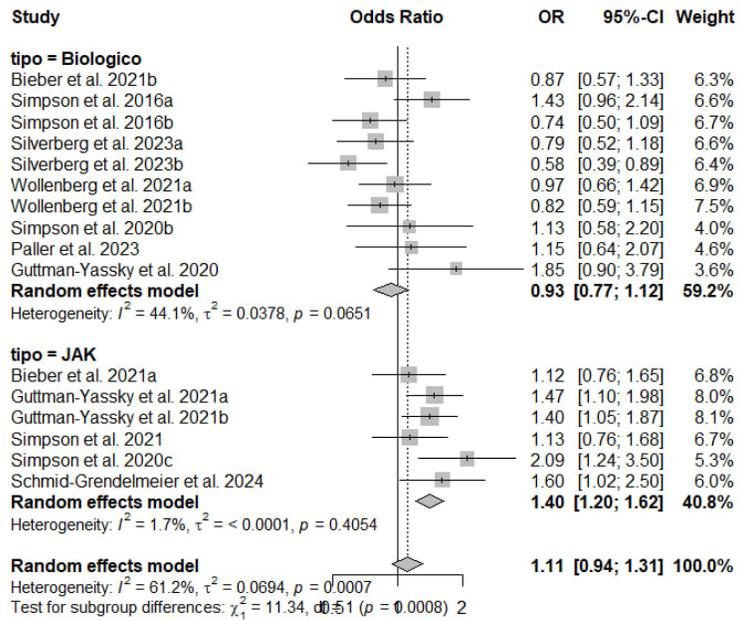


Figura 4.7: *Forest plot* del subanálisis por tipo de medicamento sobre la proporción de pacientes que alcanzaron al menos un TEAE.

4.6. Sesgo de publicación

La prueba de Egger no mostró evidencia de asimetría significativa en el *funnel plot* de la figura 4.8 ($t = -0.75$, $p = 0.4618$), lo que indica ausencia de sesgo de publicación según este método. De manera similar, la prueba de Peters tampoco detectó asimetría significativa ($t = -0.56$, $p = 0.5860$).

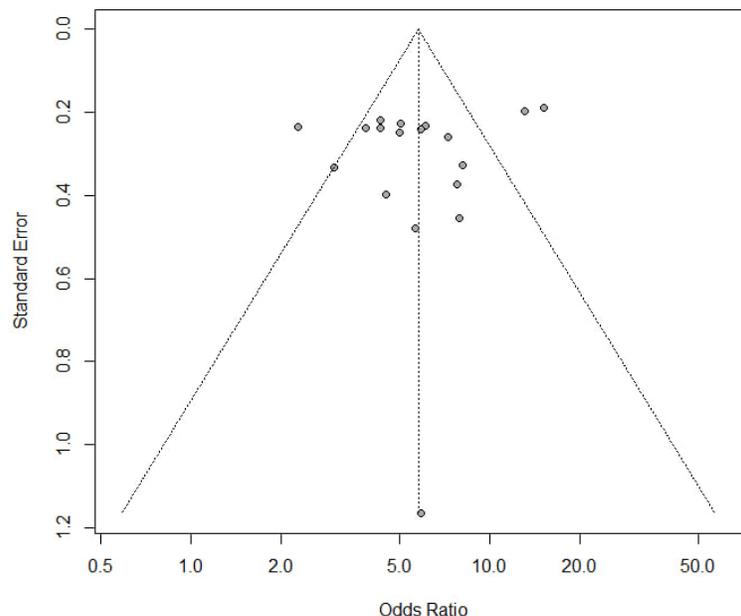


Figura 4.8: *Funnel plot* de la proporción de pacientes que alcanzaron IGA 0/1 (remisión clínica)

4.7. Interpretación de los resultados y limitaciones

Los resultados de este MA muestran que tanto los agentes biológicos como los inhibidores de JAK son eficaces para el tratamiento de la DA, logrando mejoras significativas en los desenlaces clínicos principales, como EASI75 e IGA 0/1, en comparación con el grupo control. Al analizar los subgrupos, se observó que los inhibidores de JAK presentan una mayor eficacia que los agentes biológicos, especialmente para lograr una piel clara o casi clara (IGA 0/1), diferencia que resultó significativa.

Sin embargo, este beneficio en eficacia de los inhibidores de JAK se acompaña de un mayor riesgo de acontecimientos adversos emergentes del tratamiento (TEAE), en comparación con los agentes biológicos. Mientras que los agentes biológicos no mostraron un aumento significativo en el riesgo de TEAE respecto al control, los inhibidores de JAK sí presentaron un incremento significativo en este desenlace.

Además, se realizó una evaluación del sesgo de publicación para comprobar si los resultados podrían estar influidos por la falta de estudios con resultados negativos o no concluyentes. Las pruebas estadísticas aplicadas no detectaron un sesgo de publicación relevante, lo que da mayor confianza en la validez de los resultados obtenidos. Sin embargo, es importante tener en cuenta que la alta variabilidad entre los estudios y algunas limitaciones de los datos pueden influir en la interpretación final.

Por otro lado, aunque los resultados aportan evidencia relevante, este MA presenta algunas limitaciones que deben tenerse en cuenta al interpretar sus conclusiones. En primer lugar, se detectó una heterogeneidad moderada-alta entre los estudios, por lo que se utilizó un modelo de efectos aleatorios para tener en cuenta esta variabilidad tanto dentro de los estudios como entre ellos. En segundo lugar, se agruparon múltiples dosis de abrocitinib, upadacitinib y baricitinib, lo que podría generar heterogeneidad y estimaciones menos precisas del efecto si existieran diferencias importantes entre dosis. En tercer lugar, en algunos estudios, los resultados se recogieron a las 12 semanas, mientras que en otros se evaluaron a las 16 semanas.

Capítulo 5

Conclusiones

Este TFG ha tenido como objetivo central revisar y aplicar en profundidad la metodología del MA, centrándose especialmente en la combinación de resultados de estudios clínicos con medidas de efecto dicotómicas, mediante modelos de efectos fijos y aleatorios. Para ello, se ha llevado a cabo una aplicación práctica sobre la eficacia de distintos tratamientos sistémicos aprobados para la DA moderada o grave en adultos.

Desde el punto de vista metodológico, el trabajo ha permitido explorar las bases formales del MA, prestando especial atención a la estimación puntual del tamaño del efecto, el cálculo de la varianza, la asignación de pesos mediante el método de varianza inversa y la combinación de estudios bajo diferentes supuestos de heterogeneidad. Asimismo, se han revisado y aplicado herramientas para evaluar la heterogeneidad entre estudios (como los estadísticos Q , y I^2) y para valorar la posible existencia de sesgo de publicación, tanto de forma gráfica como mediante pruebas estadísticas como la de Egger.

La aplicación práctica ha incluido 18 RCTs, extraídos conforme a una estrategia estructurada bajo el modelo PICO. En ellos se ha evaluado la eficacia de varios tratamientos —biológicos (dupilumab, tralokinumab) e inhibidores de JAK (baricitinib, abrocitinib, upadacitinib)— en comparación con placebo, considerando como desenlace principal la proporción de pacientes que alcanzan EASI-75. Los resultados muestran que todos los tratamientos incluidos presentan un beneficio estadísticamente significativo respecto al grupo control, si bien con diferencias sustanciales en la magnitud del efecto.

El análisis de subgrupos indica que el tratamiento con mayor eficacia estimada es upadacitinib, seguido de baricitinib y dupilumab, lo cual puede relacionarse con el mecanismo de acción y la potencia antiinflamatoria de los inhibidores de JAK. No obstante, estas diferencias no pueden interpretarse como comparaciones directas entre tratamientos activos, ya que todos los análisis se han realizado con respecto a placebo. Además, el trabajo no ha incluido análisis sobre seguridad, calidad de vida ni coste-efectividad, aspectos clave para una evaluación clínica completa.

Como principales limitaciones del estudio destacan la no inclusión de análisis comparativos directos entre tratamientos, la falta de evaluación sistemática del riesgo de sesgo en los estudios primarios, y la exclusión de desenlaces secundarios clínicamente relevantes. Asimismo, aunque se han aplicado técnicas clásicas de detección de sesgo de publicación, estas poseen sensibilidad limitada en muestras reducidas, por lo que los resultados deben interpretarse con precaución.

Este trabajo ha permitido integrar los fundamentos teóricos y prácticos del MA en el contexto de un problema real de investigación biomédica. La experiencia adquirida pone de relieve tanto el potencial de esta herramienta estadística como las decisiones críticas que requiere su aplicación rigurosa.

Para investigaciones futuras, se propone ampliar el enfoque metodológico adoptado en este trabajo mediante el uso de MA en red (*network meta-analysis*, NMA), que permite comparar de forma simultánea múltiples intervenciones, incluso cuando no existen comparaciones directas entre todos los tratamientos. Esta metodología sería especialmente útil en contextos como el de la DA, donde coexisten diversos fármacos aprobados, pero los ensayos disponibles suelen comparar únicamente cada uno frente a placebo.

Además, sería recomendable incorporar modelos jerárquicos bayesianos, que ofrecen una mayor flexibilidad en la incorporación de incertidumbre, permiten trabajar con estructuras complejas de datos y facilitan la estimación de parámetros en presencia de escasa información o alta heterogeneidad. Estos modelos también permitirían integrar evidencia externa, como estudios observacionales, mediante la inclusión de distribuciones a priori informativas, lo que incrementa la robustez del análisis cuando los datos disponibles son limitados o heterogéneos.

Otra línea de desarrollo consistiría en ampliar el análisis a desenlaces múltiples o longitudinales, especialmente relevantes en el contexto clínico. Por ejemplo, además de la eficacia a corto plazo (como EASI-75 a las 16 semanas), podría analizarse la evolución de los pacientes en el tiempo, la persistencia de la respuesta terapéutica, la incidencia acumulada de eventos adversos, o indicadores de calidad de vida. El uso de modelos multivariantes o de MA multivariante permitiría integrar simultáneamente varios desenlaces correlacionados, proporcionando una visión más completa del perfil beneficio-riesgo de cada tratamiento.

Bibliografía

- [AEMPS, 2022] AEMPS (2022). Informe de posicionamiento terapéutico de upadacitinib (rinvoq®) en dermatitis atópica. Informe de Posicionamiento Terapéutico IPT 31/2022, Ministerio de Sanidad, Servicios Sociales e Igualdad, España.
- [Balduzzi et al., 2019] Balduzzi, S., Rücker, G., and Schwarzer, G. (2019). How to perform a meta-analysis with r: a practical tutorial. *BMJ Ment Health*, 22(4):153–160.
- [Bieber et al., 2021] Bieber, T., Simpson, E. L., Silverberg, J. I., Thaçi, D., Paul, C., Pink, A. E., Kataoka, Y., Chu, C.-Y., DiBonaventura, M., Rojo, R., et al. (2021). Abrocitinib versus placebo or dupilumab for atopic dermatitis. *New England Journal of Medicine*, 384(12):1101–1112.
- [Blauvelt et al., 2023] Blauvelt, A., Thyssen, J. P., Guttman-Yassky, E., Bieber, T., Serra-Baldrich, E., Simpson, E., Rosmarin, D., Elmaraghy, H., Meskimen, E., Natalie, C. R., et al. (2023). Efficacy and safety of lebrikizumab in moderate-to-severe atopic dermatitis: 52-week results of two randomized double-blinded placebo-controlled phase iii trials. *British Journal of Dermatology*, 188(6):740–748.
- [Borenstein et al., 2021] Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2021). *Introduction to meta-analysis*. John wiley & sons.
- [Chandler et al., 2019] Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. (2019). *Cochrane handbook for systematic reviews of interventions*. Hoboken: Wiley, 4.
- [Cohen, 1988] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ, 2nd edition.
- [Cummings et al., 2013] Cummings, S. R., Browner, W. S., and Hulley, S. B. (2013). Conceiving the research question and developing the study plan. *Designing clinical research*, 4:14–22.
- [DerSimonian and Laird, 1986] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188.
- [Duval and Tweedie, 2000] Duval, S. and Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2):455–463.

- [Egger et al., 1997] Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *bmj*, 315(7109):629–634.
- [Gómez, 2018] Gómez, N. J. (2018). Dermatitis atópica:: definición, consejos prácticos y tratamiento. *El farmacéutico: profesión y cultura*, (562):27–32.
- [Guttman-Yassky et al., 2020] Guttman-Yassky, E., Blauvelt, A., Eichenfield, L. F., Paller, A. S., Armstrong, A. W., Drew, J., Gopalan, R., and Simpson, E. L. (2020). Efficacy and safety of lebrikizumab, a high-affinity interleukin 13 inhibitor, in adults with moderate to severe atopic dermatitis: a phase 2b randomized clinical trial. *JAMA dermatology*, 156(4):411–420.
- [Guttman-Yassky et al., 2021] Guttman-Yassky, E., Teixeira, H. D., Simpson, E. L., Papp, K. A., Pangan, A. L., Blauvelt, A., Thaçi, D., Chu, C.-Y., Hong, H. C.-h., Katoh, N., et al. (2021). Once-daily upadacitinib versus placebo in adolescents and adults with moderate-to-severe atopic dermatitis (measure up 1 and measure up 2): results from two replicate double-blind, randomised controlled phase 3 trials. *The Lancet*, 397(10290):2151–2168.
- [Harrer et al., 2021] Harrer, M., Cuijpers, P., Furukawa, T., and Ebert, D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC.
- [Harrer et al., 2019] Harrer, M., Cuijpers, P., Furukawa, T., and Ebert, D. D. (2019). dmetar: companion r package for the guide ‘doing meta-analysis in r’. *R package version 0.0*, 9000:2019.
- [Hedges and Olkin, 2014] Hedges, L. V. and Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.
- [Henrich et al., 2010] Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302):29–29.
- [Higgins and Thompson, 2002] Higgins, J. P. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11):1539–1558.
- [Higgins et al., 2003] Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *bmj*, 327(7414):557–560.
- [Mattos and Ruellas, 2015] Mattos, C. T. and Ruellas, A. C. d. O. (2015). Systematic review and meta-analysis: what are the implications in the clinical practice? *Dental press journal of orthodontics*, 20(1):17–19.
- [Moher et al., 2009] Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). PRISMA Statement. <https://www.prisma-statement.org/>. Accessed: 2024-06-07.
- [Paller et al., 2023] Paller, A. S., Flohr, C., Cork, M., Bewley, A., Blauvelt, A., Hong, H. C.-h., Imafuku, S., Schuttelaar, M. L., Simpson, E. L., Soong, W., et al. (2023). Efficacy and safety of tralokinumab in adolescents with moderate to severe atopic dermatitis: the phase 3 ecztra 6 randomized clinical trial. *JAMA dermatology*, 159(6):596–605.

- [Paule and Mandel, 1982] Paule, R. C. and Mandel, J. (1982). Consensus values and weighting factors. *Journal of research of the National Bureau of Standards*, 87(5):377.
- [Peters et al., 2006] Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama*, 295(6):676–680.
- [Rücker and Schwarzer, 2021] Rücker, G. and Schwarzer, G. (2021). Beyond the forest plot: the drapery plot. *Research Synthesis Methods*, 12(1):13–19.
- [Schmid-Grendelmeier et al., 2024] Schmid-Grendelmeier, P., Gooderham, M. J., Hartmann, K., Konstantinou, G. N., Fellmann, M., Koulias, C., Clibborn, C., Biswas, P., and Brunner, P. M. (2024). Efficacy and safety of abrocitinib in patients with moderate-to-severe atopic dermatitis and comorbid allergies. *Allergy*, 79(1):174–183.
- [Sidik and Jonkman, 2005] Sidik, K. and Jonkman, J. N. (2005). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54(2):367–384.
- [Sidik and Jonkman, 2019] Sidik, K. and Jonkman, J. N. (2019). A note on the empirical bayes heterogeneity variance estimator in meta-analysis. *Statistics in medicine*, 38(20):3804–3816.
- [Silberzahn et al., 2018] Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahnik, Š., Bai, F., Bannard, C., Bonnier, E., et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in methods and practices in psychological science*, 1(3):337–356.
- [Silverberg et al., 2023] Silverberg, J. I., Guttman-Yassky, E., Thaçi, D., Irvine, A. D., Stein Gold, L., Blauvelt, A., Simpson, E. L., Chu, C.-Y., Liu, Z., Gontijo Lima, R., et al. (2023). Two phase 3 trials of lebrikizumab for moderate-to-severe atopic dermatitis. *New England Journal of Medicine*, 388(12):1080–1091.
- [Simpson et al., 2016] Simpson, E. L., Bieber, T., Guttman-Yassky, E., Beck, L. A., Blauvelt, A., Cork, M. J., Silverberg, J. I., Deleuran, M., Kataoka, Y., Lacour, J.-P., et al. (2016). Two phase 3 trials of dupilumab versus placebo in atopic dermatitis. *New England Journal of Medicine*, 375(24):2335–2348.
- [Simpson et al., 2021] Simpson, E. L., Forman, S., Silverberg, J. I., Zirwas, M., Mavarakis, E., Han, G., Guttman-Yassky, E., Marnell, D., Bissonnette, R., Waibel, J., et al. (2021). Baricitinib in patients with moderate-to-severe atopic dermatitis: results from a randomized monotherapy phase 3 trial in the united states and canada (breeze-ad5). *Journal of the American Academy of Dermatology*, 85(1):62–70.
- [Simpson et al., 2020a] Simpson, E. L., Paller, A. S., Siegfried, E. C., Boguniewicz, M., Sher, L., Gooderham, M. J., Beck, L. A., Guttman-Yassky, E., Pariser, D., Blauvelt, A.,

- et al. (2020a). Efficacy and safety of dupilumab in adolescents with uncontrolled moderate to severe atopic dermatitis: a phase 3 randomized clinical trial. *JAMA dermatology*, 156(1):44–56.
- [Simpson et al., 2020b] Simpson, E. L., Sinclair, R., Forman, S., Wollenberg, A., Aschoff, R., Cork, M., Bieber, T., Thyssen, J. P., Yosipovitch, G., Flohr, C., et al. (2020b). Efficacy and safety of abrocitinib in adults and adolescents with moderate-to-severe atopic dermatitis (jade mono-1): a multicentre, double-blind, randomised, placebo-controlled, phase 3 trial. *The Lancet*, 396(10246):255–266.
- [Ständer et al., 2023] Ständer, S., Kwatra, S. G., Silverberg, J. I., Simpson, E. L., Thyssen, J. P., Yosipovitch, G., Zhang, F., Cameron, M. C., Cella, R. R., Valdez, H., et al. (2023). Early itch response with abrocitinib is associated with later efficacy outcomes in patients with moderate-to-severe atopic dermatitis: subgroup analysis of the randomized phase iii jade compare trial. *American journal of clinical dermatology*, 24(1):97–107.
- [Sterne et al., 2016] Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., et al. (2016). Robins-i: a tool for assessing risk of bias in non-randomised studies of interventions. *bmj*, 355.
- [Sterne et al., 2019] Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., et al. (2019). Rob 2: a revised tool for assessing risk of bias in randomised trials. *bmj*, 366.
- [Viechtbauer, 2005] Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3):261–293.
- [Viechtbauer, 2010] Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of statistical software*, 36:1–48.
- [Wickham et al., 2019] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. *Journal of open source software*, 4(43):1686.
- [Wollenberg et al., 2021] Wollenberg, A., Blauvelt, A., Guttman-Yassky, E., Worm, M., Lynde, C., Lacour, J.-P., Spelman, L., Katoh, N., Saeki, H., Poulin, Y., et al. (2021). Tralokinumab for moderate-to-severe atopic dermatitis: results from two 52-week, randomized, double-blind, multicentre, placebo-controlled phase iii trials (ecztra 1 and ecztra 2). *British journal of dermatology*, 184(3):437–449.

Capítulo 6

Anexo A: Código R

```
#Librerias

install.packages(c("tidyverse", "meta", "metafor"))
library("tidyverse")
library("meta")
library("metafor")
library("data.table")

# Preparación de los datos
file <-"Datos_estudios.csv"
datos <- fread(file)

datos$study_id <- as.factor(datos$study_id)
datos$medicamento <- as.factor(datos$medicamento)
datos$outcome <- as.factor(datos$outcome)
datos$tipo <- as.factor(datos$tipo)

# Analisis variable EASI-75

ma_resultados_EASI <- metabin(e_treat, n_treat, e_contr, n_contr,
                             studlab = study_id,
                             data = datos[outcome == "EASI75"],
                             sm = "OR",
                             common = FALSE,
                             random = TRUE)

summary(ma_resultados_EASI)
forest(ma_resultados_EASI, leftcols = "studlab")
```

```

ma_resultados_EASI <- metabin(e_treat, n_treat, e_contr, n_contr,
                             studlab = study_id,
                             data = datos[outcome == "EASI75"],
                             subgroup = datos[outcome == "EASI75"]$tipo,
                             sm = "OR",
                             common = FALSE,
                             random = TRUE)

summary(ma_resultados_EASI)
forest(ma_resultados_EASI, leftcols = "studlab")

```

```
# Analisis variable IGA 0/1
```

```

ma_resultados_IGA <- metabin(e_treat, n_treat, e_contr, n_contr,
                             studlab = study_id,
                             data = datos[outcome == "IGA"],
                             sm = "OR",
                             common = FALSE,
                             random = TRUE)

```

```

summary(ma_resultados_IGA)
forest(ma_resultados_IGA, leftcols = "studlab")

```

```

ma_resultados_IGA <- metabin(e_treat, n_treat, e_contr, n_contr,
                             studlab = study_id,
                             data = datos[outcome == "IGA"],
                             subgroup = datos[outcome == "IGA"]$tipo,
                             sm = "OR",
                             common = FALSE,
                             random = TRUE)

```

```

summary(ma_resultados_IGA)
forest(ma_resultados_IGA, leftcols = "studlab")

```

```
# Analisis variable TEAE
```

```

ma_resultados_TEAE <- metabin(e_treat, n_treat, e_contr, n_contr,
                              studlab = study_id,
                              data = datos[outcome == "TEAE"],
                              sm = "OR",
                              common = FALSE,
                              random = TRUE)

```

```

summary(ma_resultados_TEAE)
forest(ma_resultados_TEAE, leftcols = "studlab")

ma_resultados_TEAE <- metabin(e_treat, n_treat, e_contr, n_contr,
                             studlab = study_id,
                             data = datos[outcome == "TEAE"],
                             subgroup = datos[outcome == "TEAE"]$tipo,
                             sm = "OR",
                             common = FALSE,
                             random = TRUE)

summary(ma_resultados_TEAE)
forest(ma_resultados_TEAE, leftcols = "studlab")

# Sesgo de publicación

funnel(ma_resultados_EASI)

# Test de Egger
metabias(ma_resultados_EASI, method.bias = "egger")

# Test de Peters
metabias(ma_resultados_EASI, method.bias = "peters")

```