



## TALLER: REVISIONES SISTEMÁTICAS CON METAANÁLISIS EN RED Y COMPARACIONES INDIRECTAS: REVISIÓN CRÍTICA DE LOS TRATAMIENTOS PARA ESCLEROSIS MÚLTIPLE

Revisiones sistemáticas, comparaciones indirectas, metaanálisis en red. Claves para su análisis y lectura crítica

### Comparaciones indirectas

**M<sup>a</sup> Dolores Fraga Fuentes**

*Coordinadora de Farmacia.*

*Dirección General de Planificación, Ordenación e Inspección Sanitaria y Farmacia.*

*Consejería de Sanidad de Castilla-La Mancha*



¿Por qué estamos hablando de comparaciones indirectas?

# ¿Por qué estamos hablando de comparaciones indirectas?



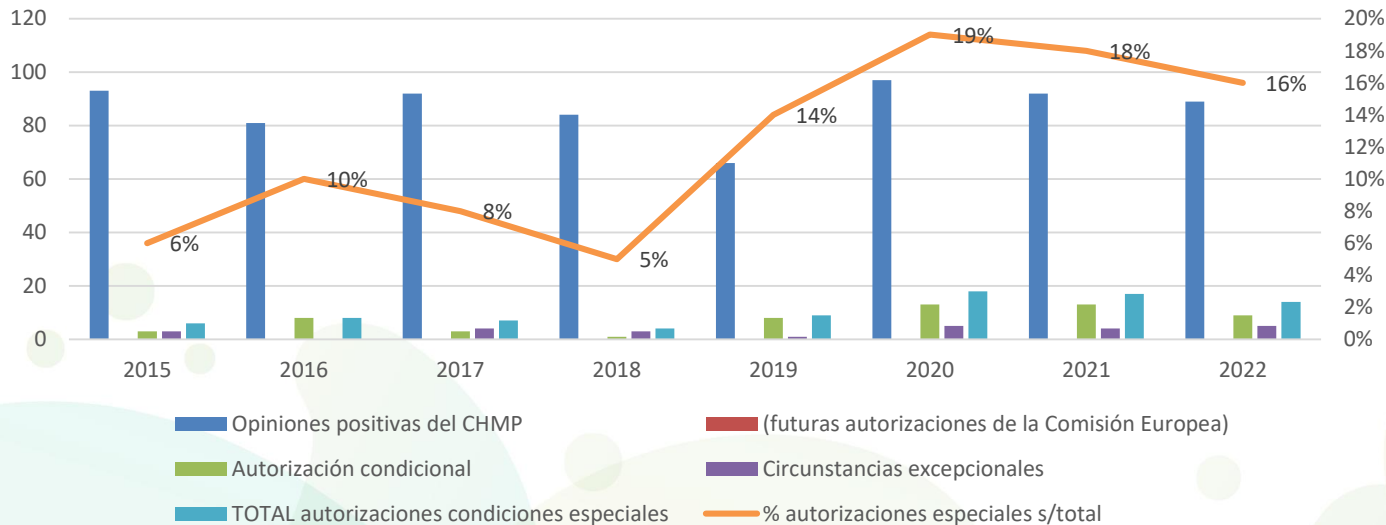
A la hora de evaluar un nuevo medicamento para establecer su lugar en terapéutica, es necesario realizar una comparación con el resto de las alternativas disponibles para la indicación estudiada. Los ensayos clínicos comparativos con asignación aleatoria y las revisiones sistemáticas y/o metaanálisis se consideran las mejores fuentes de evidencia, a pesar de sus limitaciones.

Incremento en opciones de tratamiento que requieren efectividad comparativa. Escasez de ensayos comparativos directos  
Necesidad de toma de decisiones ante varias alternativas terapéuticas

**Una comparación indirecta permite comparar dos tratamientos cuando no hay un ensayo que los compara, pero sí hay estudios de cada uno de ellos frente a una misma alternativa.**

# ¿Por qué estamos hablando de comparaciones indirectas?

## AUTORIZACIONES EMA 2015-2022



	Single arm	Multi-arm uncontrolled	RCT	Other	Total (n)	Total (%)
Phase I	1	0	0	3	4	4
Phase I/II	3	2	0	2	7	7
Phase II	50	8	0	0	58	60
Phase II/III	1	2	0	0	3	3
Phase III	6	4	2	2	14	15
Other	3	0	0	7	10	10
N/R	0	0	0	0	0	0
<b>Total (n)</b>	<b>64</b>	<b>16</b>	<b>2</b>	<b>14</b>	<b>96</b>	<b>100</b>
<b>Total (%)</b>	<b>67</b>	<b>17</b>	<b>2</b>	<b>15</b>	<b>100</b>	

Si no hay comparaciones directas,  
podemos realizar siempre CI?

# ¿Por qué estamos hablando de comparaciones indirectas?

Figura 1. Valoración de los comparadores en los estudios pivoteles de nuevos fármacos/indicaciones con respecto al tratamiento estándar en la indicación europea. Muestra: 186 IPT (50% de los publicados entre 2013 y marzo 2022).

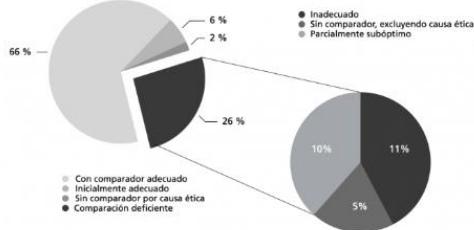
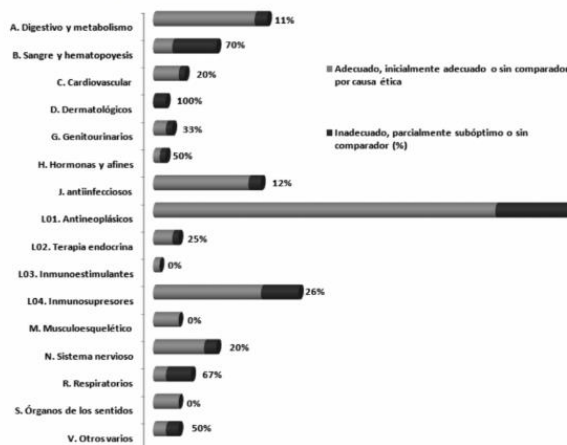


Figura 2. Proporción de comparadores de calidad suficiente (adecuados, inicialmente adecuados, o sin comparador por causa ética) en autorizaciones europeas por año de publicación del IPT. Se omiten 2013 y 2014 por contar con una muestra N<5.



Figura 3. Proporción de comparadores de calidad suficiente (adecuado, inicialmente adecuado o sin comparador por causa ética) o insuficiente (inadecuados, parcialmente subóptimos, o sin comparador excluida causa ética) por grupo terapéutico. Autorizaciones europeas con IPT publicado entre 2013 y marzo 2022. La longitud de la barra indica el tamaño de la muestra de cada grupo. El porcentaje se refiere a la carencia de una comparación de calidad suficiente.



Para paliar la falta de una comparación adecuada se recurre en la evaluación a comparaciones indirectas ajustadas. Con frecuencia esto se encuentra dificultado por ausencia de las condiciones necesarias, como son la presencia de un comparador común en los ensayos clínicos respectivos y la similitud en el diseño de los mismos.

## Evaluación de los comparadores empleados en los estudios pivoteles de nuevos medicamentos e indicaciones

ALIGRE-DEL-REY EJ, FÉNIX-CABALLERO S, MORENO RAMOS C, BORRERO-RUBIO JM  
Servicio de Farmacia, Hospital Universitario Puerto Real (Cádiz, España)

Fecha de recepción: 27/01/2023 - Fecha de aceptación: 27/02/2023

DOI: <http://dx.doi.org/10.4321/S1699-714X2023000300002>

# ¿Por qué estamos hablando de comparaciones indirectas?

Farm Hosp. 2012;36(4):176-179



ORIGINAL

## Comparaciones indirectas en los informes de evaluación de medicamentos en la web del grupo GENESIS de la SEFH\*

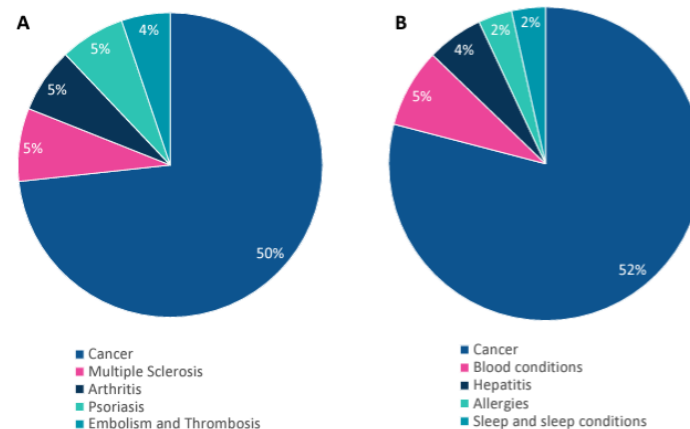
Tabla 1 Disponibilidad de ensayos clínicos comparativos del medicamento estudiado en los informes incluidos en el estudio

Disponibilidad de estudios comparativos del medicamento evaluado en el informe	Número de informes	Porcentaje
Frente al tratamiento de referencia para la indicación estudiada	169	50,1
Frente a placebo	95	28,2
Frente a otro tratamiento diferente al tratamiento de referencia	47	13,9
No hay estudios comparativos	15	4,4
Frente a «no tratamiento»	11	3,3
Total	337	100

En el 95% de los informes existe un comparador activo del medicamento evaluado, sin embargo sólo en el 50% hay un estudio comparativo directo frente al comparador de referencia

Ortega A et al. Farm Hosp. 2012;36(4):176-9

Figure 1. Five most frequently reported disease areas for TAs (A) included indirect comparisons (N=167), and (B) did not include indirect comparisons (N=132)

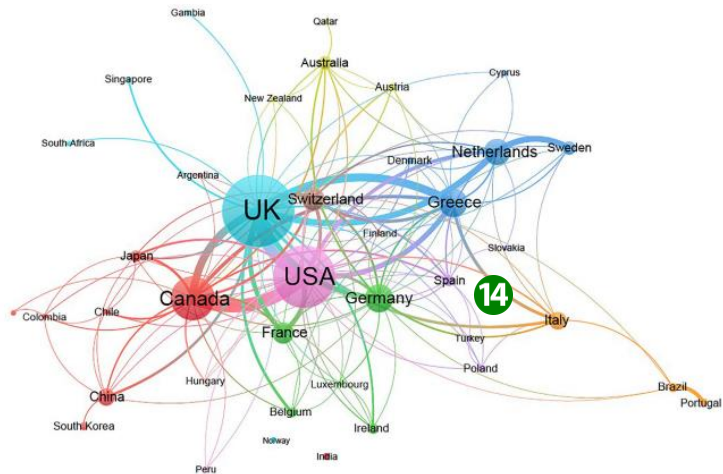


Un poco más de la mitad de las TA valoradas por el NICE describen una CI (167/299;56%), siendo el 5% en esclerosis múltiple

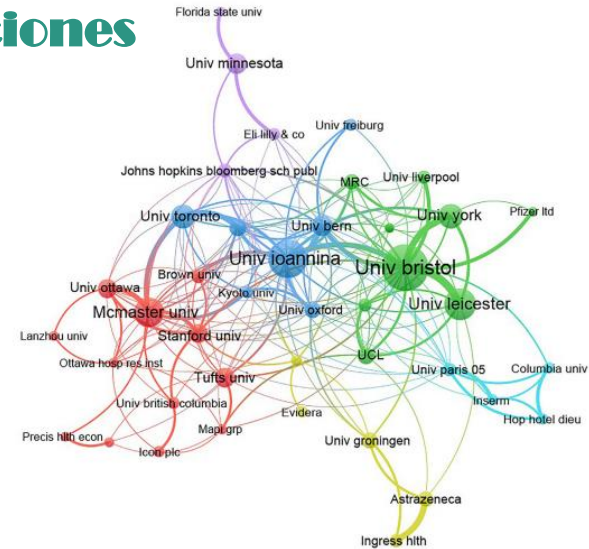
Gonçalves-Bradley DC et al. Indirect Comparisons of Clinical Effectiveness in NICE HTA Submissions: Software Use Trends, 2000-2022. Presented at ISPOR 2022 Vienna Annual Meeting.  
[https://www.ispor.org/docs/default-source/euro2022/isporeu22langfordhta240-pdf.pdf?sfvrsn=ca49ecb6\\_0](https://www.ispor.org/docs/default-source/euro2022/isporeu22langfordhta240-pdf.pdf?sfvrsn=ca49ecb6_0)



# Dónde estamos..... en comparaciones indirectas



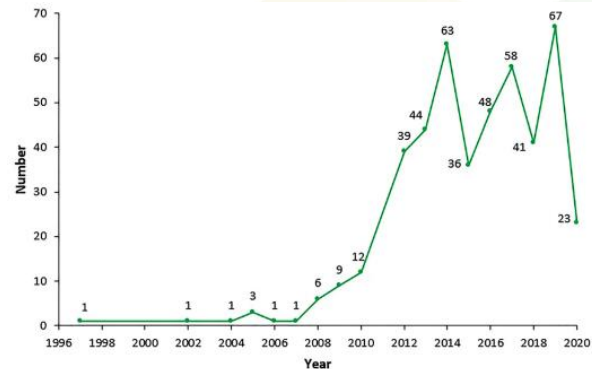
The network map of countries for NMA methodological studies



The network map of institutions for NMA methodological studies

Tian J, Gao Ya, Zhang J, et al. Progress and challenges of network meta-analysis. J Evid Based Med. 2021;14:218–231. <https://doi.org/10.1111/jebm.12443>

Citados:  
 PRISMA-NMA 2015 (Huton et al)  
 Modelo bayesiano (2004) Lu et al  
 Bucher et al 1997



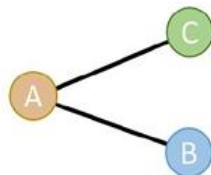
Publication years of NMA methodological studies



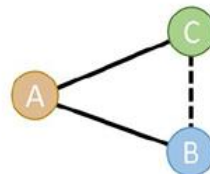
# Tipos de comparaciones entre intervenciones



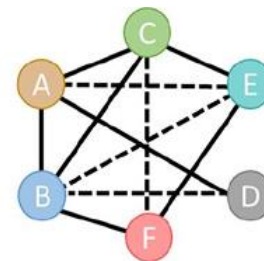
Comparación directa  
Head-to-head comparison



Meta-análisis tradicional  
Comparación directa de  
dos intervenciones frente  
a un comparador común



Comparaciones indirectas  
Comparación indirecta de  
dos o más intervenciones  
usando un comparador o  
múltiples comparadores  
Un sólo ensayo de A vs B y  
uno de A vs C. Combinados  
estiman B vs C  
Método Bucher en red de  
estrella simple  
Uno de los brazos o ambos  
son metanálisis (RCT,  
observacionales, IPD, ...)



Network meta-análisis  
Comparación directa e  
indirecta de tres o más  
intervenciones usando  
múltiples comparadores.  
Generalización del meta-  
análisis para analizar una  
red de evidencia  
compleja, que incluye  
tanto evidencia directa  
como indirecta. Métodos  
bayesianos,  
frecuentistas...

Redes conectadas  
Redes desconectadas



# Tipos de comparaciones entre intervenciones



## Comparación directa entre dos tratamientos

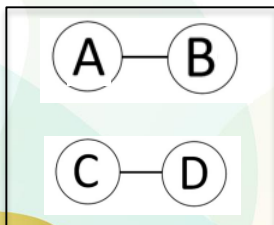
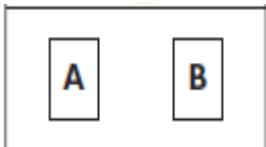
Comparados mediante:

Ensayo clínico

Varios ensayos clínicos

Un metaanálisis

Uno o varios estudios observacionales



## Comparación indirecta naïve o no ajustada

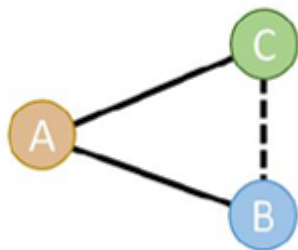
Se comparan directamente los resultados de los brazos individuales de varios estudios como si se tratara del mismo estudio.

NO ADECUADO, riesgo elevado de no proporcionar evidencia relevante

Red de evidencia desconectada



## Tipos de comparaciones entre intervenciones



### Comparación indirecta ajustada

Se comparan los efectos relativos de los tratamientos

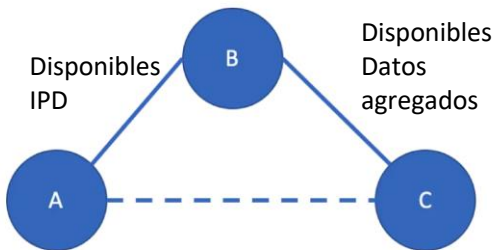
B y C no están comparados directamente  
Ambos están comparados con un comparador común A

A partir de ECAs que valoren por separado la eficacia de dos tratamientos de interés (B y C) frente a un comparador o control común A (B vs A y C vs A), se puede comparar de manera indirecta la eficacia de B vs C, en este caso se **preserva la asignación aleatoria original de los pacientes incluidos en los diferentes ECAs**. La validez de los resultados de este método implica asumir que el efecto de ambos tratamientos es coherente entre los ECAs de ambos grupos, así los resultados de B y C vs A deberían ser intercambiables.

# Tipos de comparaciones entre intervenciones

Métodos ajustados a población para comparaciones indirectas:  
 método para comparaciones indirectas en las que los datos individuales de los pacientes (IPD) de uno o más ensayos son usados para ajustar las características relevantes de la población que difieren entre los estudios para estimar un efecto del tratamiento

La técnica MAIC está diseñada para hacer comparables las poblaciones de estudios, haciendo coincidir los criterios de inclusión/exclusión y las características basales de los pacientes entre los diferentes ensayos. Este ajuste se realiza dando un mayor peso a los datos a nivel del paciente.



Método híbrido

Normalmente los IPD no están disponibles y no pueden ser reconstruidos con todos los detalles requeridos. Si se dispone de todos los datos pueden ser de utilidad para el análisis de subgrupos



## A methodological review of Population-Adjusted Indirect Comparisons reveals inconsistent reporting and suggests publication bias

A. Serres-Lamande, B. Zenati, A. Dechartres, J. Lambert, D. Hojage, 2023

### Context & Objectives

Population-Adjusted Indirect Comparisons (PAICs) allow for comparisons between two treatments evaluated in different trials while accounting for differences in patients characteristics, when Individual Patient Data (IPD) are only available for one trial and published Aggregated Data (AgD) for the other.

Such comparisons gained popularity recently in market access applications, but their implementations had not been evaluated in the medical literature.

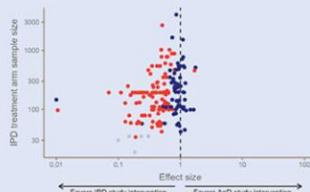


**Bibliographical Trends**  
288 PAICs collected, among 133 publications in PubMed (until April 1st, 2022).  
Increase in publications in recent years.

### Methodology and Reporting



### Published Results



### Conclusion

The methodology and reporting of these studies were heterogeneous and overall insufficient regarding key criteria such as primary outcome definition and covariates selection for adjustment.

Published results suggest a major publication and reporting bias. Improving the methodology and reporting standards is required to increasing trust in PAICs results.



ScienceDirect

Contents lists available at [sciencedirect.com](http://sciencedirect.com)  
Journal homepage: [www.elsevier.com/locate/jval](http://www.elsevier.com/locate/jval)

### Systematic Literature Review

## A Comparison of Relative-Efficacy Estimate(S) Derived From Both Matching-Adjusted Indirect Comparisons and Standard Anchored Indirect Treatment Comparisons: A Review of Matching-Adjusted Indirect Comparisons

Owen Cassidy, MSc, Marie Harte, MSc, Lea Trela-Larsen, PhD, Cathal Walsh, PhD, Arthur White, PhD, Laura McCullagh, PhD, Joy Leahy, PhD

### ABSTRACT

**Objectives:** We present an empirical comparison of relative-efficacy estimate(s) from matching-adjusted indirect comparisons (MAICs) with estimates from corresponding standard anchored indirect treatment comparisons.

**Methods:** A total of 80 comparisons were identified from 17 publications through a systematic rapid review. A standardized metric that used reported relative treatment efficacy estimates and their associated uncertainty was used to compare the methods across different treatment indications and outcome measures.

**Results:** On aggregate, MAICs presented for connected networks tended to report a more favorable relative-efficacy estimate for the treatment for which individual-level patient data were available relative to the reported indirect treatment comparison estimate.

**Conclusions:** Although we recognize the importance of MAIC and other population adjustment methods in certain situations, we recommend that results from these analyses are interpreted with caution. Researchers and analysts should carefully consider if MAICs are appropriate where presented and whether MAICs would have added value where omitted.

**Keywords:** indirect treatment comparisons, matching-adjusted indirect comparison, network meta analysis, population adjusted indirect comparisons, relative efficacy, treatment efficacy.

VALUE HEALTH. 2023; ■■■■■■

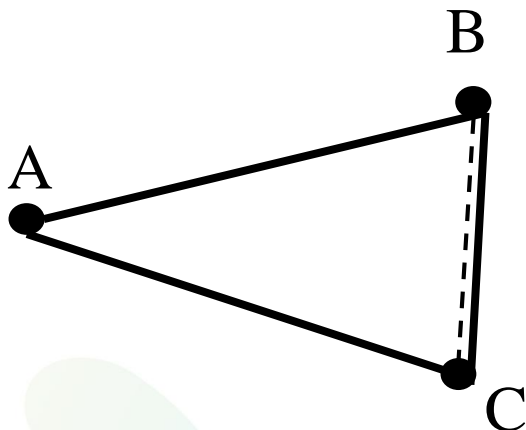
Journal of Clinical Epidemiology DOI: (10.1016/j.jclinepi.2023.09.004)

Copyright © 2023 [Terms and Conditions](#)



Precaución en la interpretación

## Comparaciones Mixtas



Se llaman comparaciones mixtas cuando hay algún bucle cerrado, alguna comparación se puede calcular de forma directa e indirecta, en este caso existe un ensayo que compara B y C, pero B y C también lo podemos calcular indirectamente a través de la comparación de B con A y C con A.

<https://gruposdetrabajo.sefh.es/genesis/enlaces-de-interes/utilidades>

CALCULATOR FOR INDIRECT COMPARISONS (for absolute effects)

Direct comparisons	Effect size	95% CI		z	p
		Lower	Upper		
A vs B (B-A)	2,00	1,00	3,00		
A vs C (C-A)	2,00	1,00	3,00		

Indirect comparison	Effect size	95% CI		z	p
		Lower	Upper		
B vs C (C-B)	0,00	-1,41	1,41	0,00	1,0000
C vs B (B-C)	0,00	-1,41	1,41	0,00	1,0000

CALCULATOR FOR MIXED COMPARISONS (for absolute effects)

A vs B (B-A)	Effect size	95% CI		z	p
		Lower	Upper		
Direct comparison	2,00	1,00	3,00		
Indirect comparison	2,00	1,00	3,00		

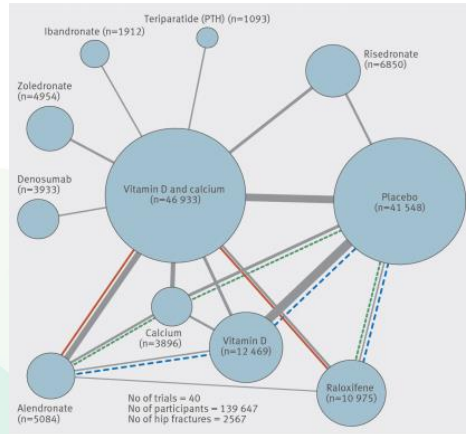
A vs B (B-A)	Effect size	95% CI		z	p
		Lower	Upper		
Mixed comparison	2,00	1,29	2,71	5,54	0,0000


  

Inconsistency factor	IF	95% CI		z	p
		Lower	Upper		
	0,00	0,24	1,41	0,00	1,0000

# Geometría de la red

Un elemento clave para comprender la red de tratamientos es la evaluación de la **geometría de su red de evidencia, es decir, todos los estudios seleccionados y su relación**. Veamos un ejemplo, mostramos las generalidades, aunque pueden existir diferentes gráficos en función del programa utilizado:



Los nodos  representan los distintos tratamientos

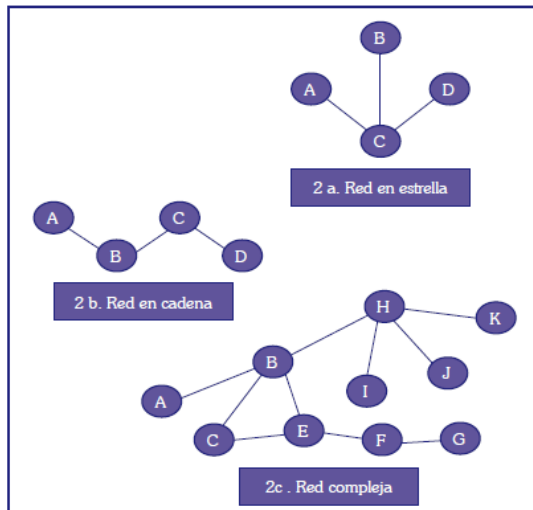
El tamaño de los nodos es proporcional al tamaño de muestra (número de participantes asignados al tratamiento)

El grosor de las líneas es proporcional al número de ensayos usados para la comparación

Las líneas continuas son comparaciones directas y sin línea o líneas discontinuas indirectas

**Puhan MA et al. BMJ 2014;349:g5630**

Algunos ejemplos:



En la figura 2a tenemos la red en estrella, que contiene estudios que comparten un comparador común (C), los tratamientos (A,B,C) sólo están comparados con el comparador común, no entre ellos. En este caso sólo podemos realizar comparaciones indirectas entre ellos. Esto puede reducir la confianza en los efectos, sobre todo si el comparador es placebo y hay un número limitado de ensayos, pacientes y eventos.

La 2b es una red en cadena, para comparar de forma indirecta A vs D se necesita enlazarlos vía dos comparadores

La 2c es una red compleja con múltiples combinaciones

*Fraga Fuentes MD et al. Boletín  
Farmacoterapéutico de Castilla-La  
Mancha. 2016*



# Geometría de la red



La diversidad y la fuerza de la red de evidencia está determinada por el número de intervenciones y comparaciones diferentes que están disponibles, cómo están representadas en la red y cuanta evidencia existe de las mismas.



Desequilibrios importantes en la cantidad de evidencia para cada intervención pueden afectar a la potencia y fiabilidad del análisis global, por lo que puede favorecer a unos tratamientos frente a otros.

La evidencia proporcionada por pequeños ensayos puede estar más sesgada y pequeños ensayos pueden mostrar falsamente efectos de tratamientos mayores. La valoración de la certidumbre de la evidencia es crucial para poder interpretar resultados y trasladar la evidencia a la decisión.



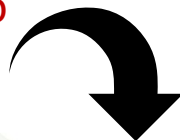
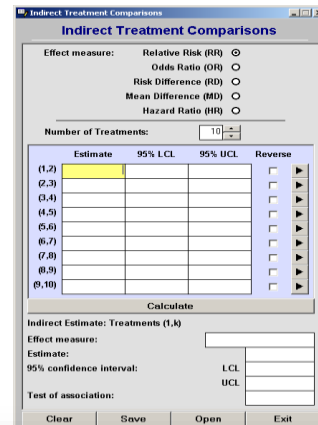
La combinación de resultados sesgados puede llevar a estimaciones poco fiables en la red y a más incertidumbre a la hora de la toma de decisiones .

Existen métodos sencillos para realizar comparaciones indirectas

ajustadas como por ejemplo el **método Bucher**

Generalmente los intervalos de confianza con que se estiman las comparaciones indirectas (ej IC 95%) son más amplios que en las comparaciones directas, es decir hay más incertidumbre en las estimaciones

El método de Bucher sólo es adecuado cuando se comparan dos tratamientos que están comparados frente a un comparador común

**Indirect Treatment Comparisons**

Effect measure: Relative Risk (RR)   
 Odds Ratio (OR)   
 Risk Difference (RD)   
 Mean Difference (MD)   
 Hazard Ratio (HR)

Number of Treatments:

	Estimate	95% LCL	95% UCL	Reverse
(1,2)				<input type="checkbox"/>
(2,3)				<input type="checkbox"/>
(3,4)				<input type="checkbox"/>
(4,5)				<input type="checkbox"/>
(5,6)				<input type="checkbox"/>
(6,7)				<input type="checkbox"/>
(7,8)				<input type="checkbox"/>
(8,9)				<input type="checkbox"/>
(9,10)				<input type="checkbox"/>

Calculate

Indirect Estimate: Treatments (1,4)

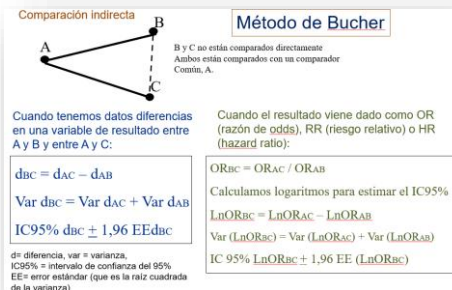
Effect measure:

Estimate:

95% confidence interval: LCL  UCL

Test of association:

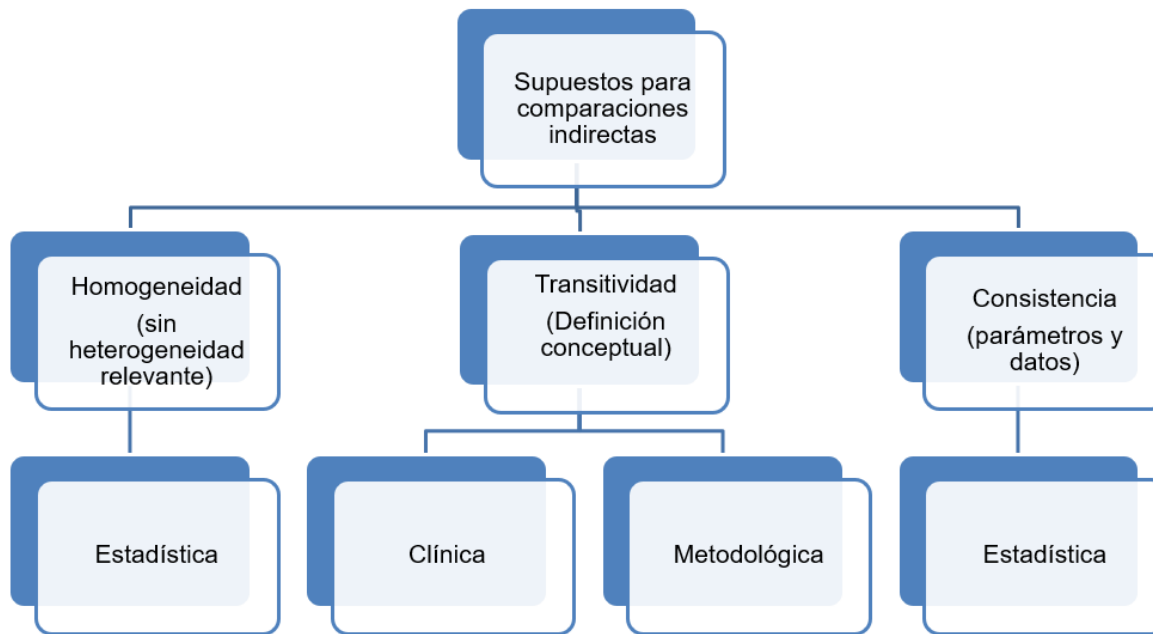
Clear Save Open Exit



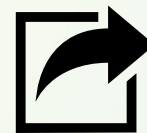
Puede ser usado, por ejemplo, para la red en estrella, siempre que los ensayos disponibles sean de dos brazos

Network metanálisis: programas más complejos

# Supuestos a considerar en las comparaciones indirectas



**Transitividad (similitud)**  
**Homogeneidad**  
**Consistencia (coherencia)**



**Intercambiabilidad**

# Supuestos a considerar en las comparaciones indirectas

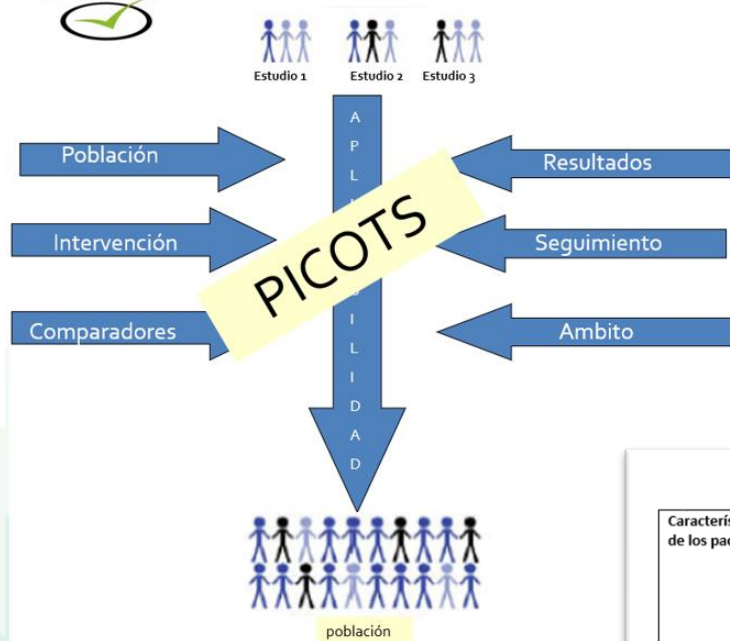
Existen **tres preguntas básicas** que debemos realizar ante una comparación indirecta:

1. ¿Los estudios son suficientemente similares para poder ser comparados? (supuesto de transitividad/similitud). *Todos los ensayos incluidos deben ser comparables en términos de potenciales factores modificadores de efectos.*
2. ¿Los ensayos disponibles para las distintas comparaciones son suficientemente homogéneos para combinarlos? (supuesto de homogeneidad). *No debe existir una heterogeneidad relevante entre los resultados de los ensayos en las comparaciones por pares.*
3. Si existe evidencia directa e indirecta, ¿los hallazgos son suficientemente consistentes para que ambas evidencias puedan combinarse? (supuesto de consistencia). *No debe existir discrepancia o inconsistencia relevante entre la evidencia directa e indirecta.*

# Supuestos a considerar en las comparaciones indirectas



## Puntos críticos



Estudios comparables con respecto a los factores modificadores del efecto

Solo 17 de 511 revisiones de CI (3,3%) incluyen una descripción de los procesos de selección para FME y FP

Freitag A et al. J Comp Eff Res. 2023 Oct;12(10):e230046. doi: 10.57264/ceer-2023-0046.

Study	Probabilidad de Respuesta			Efecto relativo	
	A	B	C	RAR	RR
Ensayo 1	40%		20%	20%	2
Ensayo 2		30%	10%	20%	3

Características del estudio y de los pacientes	Edad Sexo Gravedad de la enfermedad Duración de la enfermedad Duración del estudio Comorbilidades Líneas de tratamiento Respuesta previa a tratamiento
Características de la intervención y el comparador	Tipo de intervención Dosis Vía de administración Tratamientos concomitantes Duración del tratamiento
Características de los resultados	Definición de la variable de resultado Valores basales Escala de medición de la variable resultado (RR, RAR) Duración del seguimiento

# Supuestos a considerar en las comparaciones indirectas

## Transitividad/Similitud

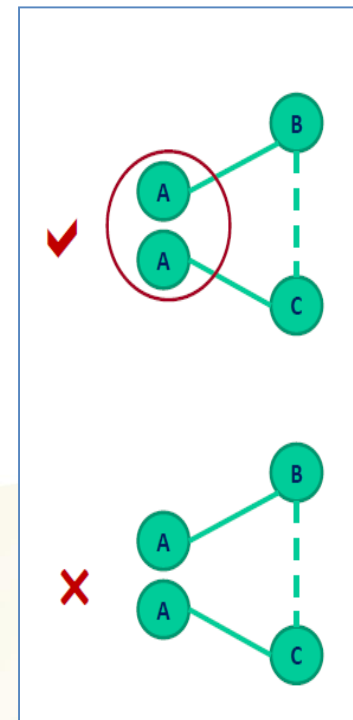
No se puede probar estadísticamente pero puede ser evaluada conceptual y epidemiológicamente

Grupo control similar

Ensayos AC y AB no difieran en cuanto a la distribución de los factores modificadores del efecto

Se asume para la media de los tamaños de los efectos y no para los estudios individuales

Los ensayos tienen que ser similares en cuanto a los factores que modifican el efecto.



Placebo oral vs intravenoso  
Comparadores no apropiados



Comparador A se da a diferentes dosis, pero no hay diferencia en la media de dosis de A entre AC y AB

## Transitividad/Similitud

Para poder combinar los estudios y no obtener resultados con sesgos: Es importante que:

1. La **selección de los estudios sea adecuada** (detectar sesgo publicación).  
No excluir ninguna alternativa
2. El **grupo control** sea el mismo en estudios que se comparan,  
**TRANSITIVIDAD**
3. Las **poblaciones sean similares** en cuanto a los factores que pueden afectar al efecto (**modificadores del efecto**), TRANSITIVIDAD
4. Las **variables de respuesta sean similares** entre los estudios, TRANSITIVIDAD, y sean las relevantes (incluir eficacia y seguridad y no dejarse llevar por los datos)
5. La **calidad de los estudios** sea adecuada (ciego, seguimiento, ..)
6. La **metodología** con la que están hechas sea **adecuada**

## ¿ Qué hacer si no hay Transitividad/Similitud?

Si no hay similitud en los factores modificadores del efecto se puede:

- ✓ Analizar subgrupos
- ✓ Hacer meta-regresión

Si estos análisis son posibles →

CI o NMA puede  
ayudar en la  
decisión

Para que sean posibles tiene que haber un número suficiente de estudios, y/o resultados fiables por subgrupos

Pero muchas veces no son posibles y  
**NO se puede hacer la comparación indirecta**





## Heterogeneidad Clínica



Variación en las características de la población, condiciones coexistentes, co-intervenciones y resultados evaluados en los estudios que pueden influir o modificar la magnitud del efecto de la medida de la intervención

## Heterogeneidad Metodológica



Diferencias entre estudios en los tamaños estimados del efecto que pueden ser atribuidos a la variabilidad y calidad y a los tipos de diseños de los estudios

## Heterogeneidad Estadística



Variabilidad en los efectos observados de los tratamientos más allá de los esperado por el error aleatorio

## Medida del efecto



Expresión de las medidas de la variable (OR, RR, DR, RAR, HR...)

# Heterogeneidad

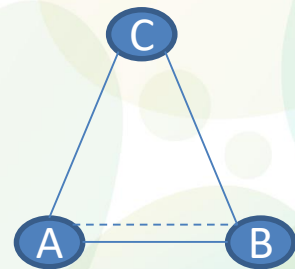
- Estudios en diferentes lugares, tiempos, poblaciones..resultados diferentes
- Estadística, test  $\chi^2$ :  $p < 0,1$  heterogeneidad
- *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.4 (updated August 2023). [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook). Test  $I^2$ :
  - 0% a 40%: puede no ser importante
  - 30% a 60%: puede representar heterogeneidad moderada
  - 50% a 90%: puede representar heterogeneidad sustancial
  - 75% a 100%: heterogeneidad considerable
- Los autores deben indicar los motivos de la heterogeneidad
- Test  $I^2$ : Cuantifica el porcentaje de la variabilidad de las estimaciones del efecto del tratamiento que se debe a la variabilidad entre estudios, ver forest plot, funnel plot
- Uso de modelo de efectos fijos o aleatorios debe ser adecuadamente valorada
- Análisis de subgrupos, meta-regresión
- Análisis de sensibilidad

Cuando hay comparaciones mixtas (dos tratamientos se pueden comparar de forma directa e indirecta)

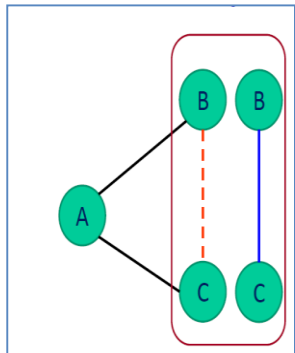
Debo comprobar si hay **CONSISTENCIA** entre comparaciones directas e indirectas, es decir, si hay coherencia entre la evidencia que he analizado

Si están comparados directamente los tratamientos pero existe otra evidencia disponible, mediante una comparación indirecta o un metaanálisis en red, puedo ver si existen discrepancias entre lo estimado mediante comparación directa e indirecta

Debe existir un bucle cerrado para analizarla



## Consistencia



Analizar la coherencia entre las comparaciones directas e indirectas  
Se pueden ver las discrepancias entre lo estimado mediante comparación directa e indirecta

Si existen discrepancias entre la comparación directa y la indirecta, es decir existe **INCONSISTENCIA**, es importante estudiar la causa y podemos no encontrarla

Se puede deber a:

que exista evidencia sesgada (no refleje la realidad)

que se hayan combinado estudios que no se pueden combinar  
(poblaciones diferentes, intervenciones diferentes, variables diferentes,..)

En general las comparaciones directas son de mayor fiabilidad que las indirectas, pero puede haber sesgos en las directas

### Box 3. Potential Reasons for Incoherence Between the Results of Direct and Indirect Comparisons

Chance

Genuine diversity

Differences in enrolled participants (eg, entry criteria, clinical setting, disease spectrum, baseline risk, selection based on prior response)

Differences in the interventions (eg, dose, duration of administration, prior administration [second-line treatment])

Differences in background treatment and management (eg, evolving treatment and management in more recent years)

Differences in definition or measurement of outcomes

Bias in head-to-head (direct) comparisons

Optimism bias with unconcealed analysis

Publication bias

Selective reporting of outcomes and analyses

Inflated effect size in early stopped trials and in early evidence

Limitations in allocation concealment, blinding, loss to follow-up, analysis as randomized

Bias in indirect comparisons

Each of the biasing issues above can affect the results of the direct comparisons on which the indirect comparisons are based

*Mills EJ et al JAMA 2012*

RESEARCH

Open Access



## Prevalence of evidence of inconsistency and its association with network structural characteristics in 201 published networks of interventions

Arete Angeliki Veroniki<sup>1,2\*</sup>, Sofia Tsokani<sup>3</sup>, Ian R. White<sup>4</sup>, Guido Schwarzer<sup>5</sup>, Gerta Rücker<sup>5</sup>, Dimitris Mavridis<sup>3,6</sup>, Julian P. T. Higgins<sup>7</sup> and Georgia Salanti<sup>8</sup>

### Abstract

**Background:** Network meta-analysis (NMA) has attracted growing interest in evidence-based medicine. Consistency between different sources of evidence is fundamental to the reliability of the NMA results. The purpose of the present study was to estimate the prevalence of evidence of inconsistency and describe its association with different NMA characteristics.

**Methods:** We updated our collection of NMAs with articles published up to July 2018. We included networks with randomised clinical trials, at least four treatment nodes, at least one closed loop, a dichotomous primary outcome, and available arm-level data. We assessed consistency using the design-by-treatment interaction (DBT) model and testing all the inconsistency parameters globally through the Wald-type chi-squared test statistic. We estimated the prevalence of evidence of inconsistency and its association with different network characteristics (eg, number of studies, interventions, intervention comparisons, loops). We evaluated the influence of the network characteristics on the DBT *p*-value via a multivariable regression analysis and the estimated Pearson correlation coefficients. We also evaluated heterogeneity in NMA (consistency) and DBT (inconsistency) random-effects models.

**Results:** We included 201 published NMAs. The *p*-value of the design-by-treatment interaction (DBT) model was lower than 0.05 in 14% of the networks and lower than 0.10 in 20% of the networks. Networks including many studies and comparing few interventions were more likely to have small DBT *p*-values (less than 0.10), which is probably because they yielded more precise estimates and power to detect differences between designs was higher. In the presence of inconsistency (DBT *p*-value lower than 0.10), the consistency model displayed higher heterogeneity than the DBT model.

**Conclusions:** Our findings show that inconsistency was more frequent than what would be expected by chance, suggesting that researchers should devote more resources to exploring how to mitigate inconsistency. The results of this study highlight the need to develop strategies to detect inconsistency (because of the relatively high prevalence of evidence of inconsistency in published networks), and particularly in cases where the existing tests have low power.

14% de 201

## Consistencia

### Método Bucher

	A vs C	B vs C		
Estimador (RR) directo:	1,29	1,53		
Límite inferior IC 95% estimador directo:	1,18	1,30		
Límite superior IC 95% estimador directo:	1,40	1,81		
	IC95%		z	p
Estimador (RR) ajustado indirecto A vs B:	0,84	0,70a 1,02	-1,80	0,07257

- Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ 2003;326:219.

### Métodos bayesianos: modelos de inconsistencia y Node splitting

	Indirecto A vs B			
	Directo A vs B	B		
Estimador (RR):	1,18	0,84		
Límite inferior IC 95% estimador:	1,09	0,70		
Límite superior IC 95% estimador:	1,27	1,02		
	IC95%		z	p
Estimador (RR) combinado:	1,12	1,05a 1,21	3,25	0,00115
	IC95%		z	p
Discrepancia entre estimador directo e indirecto: (Razón de riesgos relativos)	1,40	1,15a 1,72	3,28	0,00104

**INCONSISTENCIA**

El método usado para el test de inconsistencia debe ser claramente identificado y justificada su elección, con referencia a la estructura de la red

# ¿Para qué no sirven las comparaciones indirectas?

No sirven para generar mejor evidencia de la que existe  
No son una fuente de evidencia

Ayudan a:

- manejar de forma objetiva la evidencia
  - tomar la decisión con la evidencia que existe
- Pero:
- no generan nueva evidencia

**RESEARCH**

 OPEN ACCESS

## Living network meta-analysis compared with pairwise meta-analysis in comparative effectiveness research: empirical study

Adriani Nikolakopoulou,<sup>1</sup> Dimitris Mavridis,<sup>2,3</sup> Toshi A Furukawa,<sup>4</sup> Andrea Cipriani,<sup>5,6</sup> Andrea C Tricco,<sup>7,8</sup> Sharon E Straus<sup>7,9</sup> *et al.*

**ABSTRACT**  
**OBJECTIVE**  
To examine whether the continuous updating of prospectively planned randomised controlled trials (RCTs) ("living" network meta-analysis) provides strong evidence against the null hypothesis in comparative effectiveness research earlier than the conventional, pairwise meta-analysis.

**DESIGN**  
Empirical study of the accumulating evidence on the comparative effectiveness of clinical interventions.

**DATA SOURCES**  
Database of network meta-analyses of through searches of Medline, Embase, Cochrane Database of Systematic Reviews, and April 2015.

**ELIGIBILITY CRITERIA FOR STUDY SELECTION**  
Network meta-analyses published after 2000 that compared at least five treatments in at least 20 RCTs. Clinical experts were asked to identify each network the treatment comparison of clinical interest. Comparisons were excluded if direct and indirect evidence disagreed or node, splitting test ( $P < 0.10$ ).

**OUTCOMES AND ANALYSIS**  
Cumulative pairwise and network meta-analyses were performed for each selected comparison. Monitoring boundaries of statistical significance were constructed and the evidence against the null hypothesis was considered to be strong if monitoring boundaries were crossed.

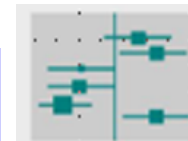
### WHAT IS ALREADY KNOWN ON THIS TOPIC

Network meta-analysis is an extension to conventional meta-analysis, which includes both direct and indirect evidence on the comparative effectiveness of multiple treatments. Network meta-analysis might produce strong evidence against the null hypothesis on the comparative effectiveness of treatments earlier than standard pairwise meta-analysis but requires more assumptions and advanced statistical methods. Sequential methods for analysis of "living" network meta-analysis of prospectively planned randomised controlled trials have recently become available, allowing the continuous updating of evidence.

### WHAT THIS STUDY ADDS

Network meta-analysis was 20% more likely to provide strong evidence against the null hypothesis of treatment differences than pairwise meta-analysis. Network meta-analysis provided strong evidence against the null hypothesis four years earlier than pairwise meta-analysis. Prospectively planned living network meta-analysis can facilitate timely recommendations and contribute to reduce research waste by providing strong evidence against the null hypothesis earlier than living pairwise meta-analysis.

# Puntos clave en la valoración crítica



1º punto fundamental TRANSPARENCIA

Datos gráficos y narrativos

Evidencia

Transitividad

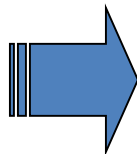
Homogeneidad

Consistencia

Impacto clínico

Generalización

Aplicabilidad



Validez interna

Validez externa

Aplicabilidad

Puntos fuertes y débiles

Relevancia y credibilidad de  
cada uno de los puntos

Robustez de la evidencia

Fuentes de sesgo



## Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review

Wasifa Zarin<sup>1</sup>, Areti Angeliki Veroniki<sup>1</sup>, Vera Nincic<sup>1</sup>, Afshin Vafaei<sup>1</sup>, Emily Reynen<sup>1</sup>, Sanobar S. Motiwala<sup>1</sup>, Jesmin Antony<sup>1</sup>, Shannon M. Sullivan<sup>1</sup>, Patricia Rios<sup>1</sup>, Caitlin Daly<sup>1</sup>, Joycelynne Ewusie<sup>1</sup>, Maria Petropoulou<sup>2</sup>, Adriani Nikolakopoulou<sup>2,3</sup>, Anna Chaimani<sup>2,3,4</sup>, Georgia Salanti<sup>2,3,4</sup>, Sharon E. Straus<sup>1,5</sup> and Andrea C. Tricco<sup>1,6\*</sup>

### Abstract

**Background:** Network meta-analysis (NMA) has become a popular method to compare more than two treatments. This scoping review aimed to explore the characteristics and methodological quality of knowledge synthesis approaches underlying the NMA process. We also aimed to assess the statistical methods applied using the Analysis subdomain of the ISPOR checklist.

**Methods:** Comprehensive literature searches were conducted in MEDLINE, PubMed, EMBASE, and Cochrane Database of Systematic Reviews from inception until April 14, 2015. References of relevant reviews were scanned. Eligible studies compared at least four different interventions from randomised controlled trials with an appropriate NMA approach. Two reviewers independently performed study selection and data abstraction of included articles. All discrepancies between reviewers were resolved by a third reviewer. Data analysis involved quantitative (frequencies) and qualitative (content analysis) methods. Quality was evaluated using the AMSTAR tool for the conduct of knowledge synthesis and the ISPOR tool for statistical analysis.

**Results:** After screening 3538 citations and 877 full-text papers, 456 NMAs were included. These were published between 1997 and 2015, with 95% published after 2006. Most were conducted in Europe (51%) or North America (31%), and approximately one-third reported public sources of funding. Overall, 84% searched two or more electronic databases, 62% searched for grey literature, 58% performed duplicate study selection and data abstraction (independently), and 62% assessed risk of bias. Seventy-eight (17%) NMAs relied on previously conducted systematic reviews to obtain studies for inclusion in their NMA. Based on the AMSTAR tool, almost half of the NMAs incorporated quality appraisal results to formulate conclusions, 36% assessed publication bias, and 16% reported the source of funding. Based on the ISPOR tool, half of the NMAs did not report if an assessment for consistency was conducted or whether they accounted for inconsistency when present. Only 13% reported heterogeneity assumptions for the random-effects model.

**Conclusions:** The knowledge synthesis methods and analytical process for NMAs are poorly reported and need improvement.

**Keywords:** Mixed-treatment, Multiple treatments, Research reporting, ISPOR, AMSTAR



# Puntos clave en la valoración crítica

Table 1

Tools and checklists to aid in systematic review conduct, to assess the reporting, quality of conduct, or the risk of bias in a review

Tool purpose	Examples of tools or checklists	Description of an example tool	Targeted users	Available tool for reviews with NMA
Guidance for conducting systematic reviews	MECIR [15]	Detailed guidance for the conduct of systematic reviews of: interventions, diagnostic test accuracy, individual patient data, public health and health promotion.	<ul style="list-style-type: none"> <li>• Review authors</li> <li>• Journal editors</li> </ul>	No
Assess the quality of conduct of reviews	AMSTAR-2 [16, 17], OQAQ [18]	AMSTAR-2 is a critical appraisal tool to assess the conduct of intervention reviews including RCTs.	<ul style="list-style-type: none"> <li>• Review authors</li> <li>• Stakeholders</li> </ul>	No
Assess the risk of bias of published reviews	ROBIS [23]	ROBIS is a tool for assessing the risk of bias in reviews. It is aimed at four broad categories of reviews mainly within health care settings: interventions, diagnosis, prognosis, and etiology.	<ul style="list-style-type: none"> <li>• Stakeholders</li> </ul>	Not at present, but in process <b>RoB NMA tool</b>
Assess the certainty in the evidence and the strength of recommendations in health care	GRADE [19]	The GRADE working group defined the certainty of a body of evidence as the extent to which one can be confident that a pooled effect estimate is close to the true effect of the intervention. Five domains assessed: risk of bias, inconsistency, indirectness, imprecision, and publication bias.	<ul style="list-style-type: none"> <li>• Review authors</li> </ul>	GRADE-NMA [20, 21], CINeMA [22], Threshold method [23]
Guidelines for the complete reporting published reviews	PRISMA Update [24]	PRISMA focuses on the reporting of already published reviews evaluating RCTs of interventions. PRISMA can determine whether a review is well described and transparently reported.	<ul style="list-style-type: none"> <li>• Stakeholders</li> <li>• Journal editors</li> </ul>	PRISMA-NMA [13]

AMSTAR-2: A Measurement Tool to Assess Systematic Reviews 2; CINeMA: Confidence in Network Meta-Analysis; Cochrane RoB 2: Cochrane risk-of-bias tool for randomized trials Version 2.0; GRADE: (Grading of Recommendations Assessment, Development and Evaluation; GRADE-NMA: (Grading of Recommendations Assessment, Development and Evaluation for Network Meta-Analysis; MECIR: Methodological Expectations of Cochrane Intervention Reviews; N/A: not applicable; OQAQ: Overview Quality Assessment Questionnaire ; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT: Randomised Controlled Trial; ROBIS:

Lunny C et al. Methodological review of NMA bias concepts provides groundwork for the development of a list of concepts for potential inclusion in a new risk of bias tool for network meta-analysis (RoB NMA Tool). <https://assets.researchsquare.com/files/rs-2010358/v1/9a133e8f-78fd-44ff-a1be-6237b1e0ce30.pdf?c=1661971510/>

## NMA risk of bias assessment

A risk of bias assessment would evaluate limitations in the way in which the NMA analysis was planned, analysed and presented. If these methods are inappropriate, the validity of the findings can be compromised.<sup>39</sup> Our tool aims either/or to assess the biases in the individual results of the NMA, and the authors' conclusions.

## Bias in results of an NMA

NMA of effect estimates from primary studies can result in overestimation or underestimation of the effects of specific intervention comparisons.<sup>39 40</sup> For example, Chaimani *et al* conducted a network meta-epidemiological study and found that, in the majority of the 32 networks they analysed, small studies tended to exaggerate the true effect estimate of the intervention, possibly due to small-study effects and publication bias.<sup>41</sup> Our tool will focus on the results of an NMA (eg, network characteristics (including geometry, effect modifiers)).<sup>42</sup> This is the approach taken in tools such as the Cochrane Risk of Bias 2 tool for assessing risk of bias in randomised trials.<sup>43</sup>

## Bias in the conclusions of an NMA

Bias may be introduced when interpreting the NMA results to draw conclusions. Conclusions may include 'spin' (eg, biased misrepresentation of the evidence, perhaps to facilitate publication) or (erroneous) misinterpretation of the evidence.<sup>44</sup> Ideally, potential biases identified in the results of the NMA might be addressed appropriately when drawing conclusions. Similarly, a well-conducted systematic review draws conclusions that are appropriate to the included evidence and can therefore be free of bias even when

the primary studies included in the review have high risk of bias.

Lunny C, Veroniki AA, Hutton B, et al. *BMJ Evidence-Based Medicine* 2023;28:58–67.

# Puntos clave en la valoración crítica

Para ver de manera sencilla la **validez** y **aplicabilidad** de una comparación indirecta disponemos en el programa MADRE de estas listas de comprobación

## Análisis de validez interna de una comparación indirecta

Referencia de la comparación indirecta evaluada:

Tabla 5.2.b.1.2 CUESTIONARIO SOBRE LA VALIDEZ DE UNA COMPARACIÓN INDIRECTA		
	SÍ/NO	JUSTIFICAR
-¿Es adecuado el método utilizado para realizar la comparación indirecta? Rechazar comparaciones "falsas" realizadas con estimaciones puntuales derivadas de diferentes ensayos controlados o de los brazos activos de diferentes ensayos controlados		
-¿Se establece claramente cómo se realizó la búsqueda y la selección de los ensayos para la inclusión?		
-¿Se hace una descripción de los métodos de análisis/síntesis de la evidencia? Manejo de sesgos		
-¿Se analiza la homogeneidad de los ensayos y la estabilidad de los efectos? Todo los ensayos responden a la misma pregunta clínica en relación con el tipo de pacientes estudiado y los tratamientos comparados		
-¿Existe concordancia de los resultados? El resultado es el mismo independientemente de la cadena de comparaciones usadas para obtenerlo		
-¿Si existe una comparación directa ¿se observa consistencia en los resultados?		
-Interpretación de los resultados. ¿Permiten obtener conclusiones claras? Analizar diferencias significativas en los riesgos basales y las respuestas en el grupo placebo		
-Otros sesgos o limitaciones encontradas en el estudio		

## Análisis de Aplicabilidad de la comparación indirecta

Tabla 5.3.b.1.3 CUESTIONARIO SOBRE LA APLICABILIDAD DE UNA COMPARACIÓN INDIRECTA		
	SÍ/NO	JUSTIFICAR
¿Considera adecuado el/los comparadores? ¿Es el tratamiento del comparador común adecuado en nuestro medio?		
¿Son importantes clínicamente los resultados?		
¿Considera adecuada la variable de medida utilizada?		
¿Considera adecuados los criterios de inclusión y/o exclusión de los pacientes? Generalización de los hallazgos (Población de pacientes en los ensayos y entre los ensayos)		
¿Cree que los resultados pueden ser aplicados directamente a la práctica clínica?		
Otros sesgos o limitaciones encontradas en el estudio		

# Transparencia

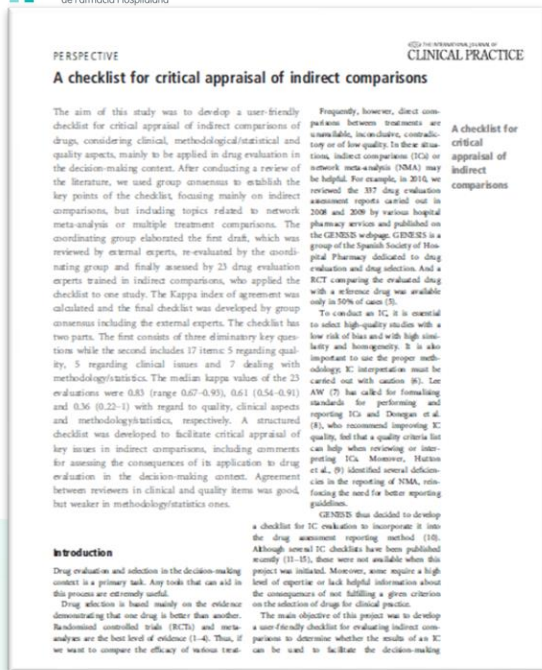


Tabla 5.3.b.1 RESULTADOS DE LAS COMPARACIONES INDIRECTAS PUBLICADAS				
Variable principal estudiada	Intervención A/ control	Intervención B/ control	RR/OR/HR/DM/DR (IC 95%)	P
RESULTADOS DE LA COMPARACIÓN DIRECTA (en caso de que existan comparaciones directas)				
Variable principal estudiada	Intervención A/ control	Intervención B/ control	RR/OR/HR/DM/DR (IC 95%)	p
INTERPRETACIÓN DE LOS RESULTADOS DE LA COMPARACIÓN INDIRECTA				
Se interpretan correctamente <input type="checkbox"/> SÍ <input type="checkbox"/> NO <input type="checkbox"/> Dudoso <input type="checkbox"/>				
Cómo se interpretaron: Se discutió la heterogeneidad <input type="checkbox"/> SÍ <input type="checkbox"/> NO <input type="checkbox"/> Dudoso <input type="checkbox"/> Se realizó análisis de sensibilidad (reanálisis cambiando datos de partida de la comparación) <input type="checkbox"/> SÍ <input type="checkbox"/> NO <input type="checkbox"/>				
Otros comentarios:				
Está justificada la comparación indirecta: <input type="checkbox"/> SÍ <input type="checkbox"/> NO <input type="checkbox"/> Dudoso <input type="checkbox"/>				
Realizada por: Laboratorio fabricante <input type="checkbox"/> Patrocinada por laboratorio <input type="checkbox"/> Organismo independiente <input type="checkbox"/>				
Nombre:				

Tabla 5.3.b.1.1. Modelo general de tabla de resultados de comparaciones indirectas:						
Referencia: COMPARACIONES DE ECAs						
CI NO AJUSTADA <input type="checkbox"/>		METANÁLISIS DE REDES DE TRATAMIENTOS <input type="checkbox"/>				
CI AJUSTADA <input type="checkbox"/>						
Medicamento A:						
Tipo de estudio, diseño, ref.	Intervención A (N) (dosis, duración del tto)	Control(N) (dosis, duración del tto)	Características de los pacientes (edad, % M:V, ...)		Duración del estudio (seguimiento)	
ECa (ese III, DC, R, otros)						
Medicamento B:						
Tipo de estudio, diseño, ref.	Intervención B (N) (dosis, duración del tto)	Control(N) (dosis, duración del tto)	Características de los pacientes (edad, % M:V, ...)		Duración del estudio (seguimiento)	
Medicamentos	Medicamento A	Medicamento B	Medicamento C			
Nº de ensayos						
Nº de comparaciones						
COMPARACIÓN DIRECTA: <input type="checkbox"/> SÍ <input type="checkbox"/> NO <input type="checkbox"/>						
Tipo de estudio, diseño, ref.	Medicamento A (N) (dosis, duración del tto)	Medicamento B (N) (dosis, duración del tto)	Características de los pacientes (edad, % M:V, ...)		Duración del estudio (seguimiento)	
MÉTODOS USADOS PARA REALIZAR LA COMPARACIÓN INDIRECTA						
Buñol:	<input type="checkbox"/> No especificado	<input type="checkbox"/> Bayesiano	<input type="checkbox"/> Frequentista			
		One				
¿Se especifica si se ajusta el sesgo? <input type="checkbox"/> SÍ <input type="checkbox"/> NO <input type="checkbox"/> SE <input type="checkbox"/> NO						

# Puntos clave en la valoración crítica

Lista de comprobación del grupo GENESIS:



Ortega A, Fraga MD, Alegre-del-Rey EJ, et al. Int J Clin Pract. 2014 Oct;68(10):1181-9.



**PARTE A. Tres preguntas de eliminación**



**PARTE B. 17 puntos: 5 de aspectos clínicos, 7 de metodología/estadística, 5 de calidad**



Esta lista de comprobación permite evaluar la **calidad metodológica y estadística** de las comparaciones indirectas, la **presencia de sesgos** y su **influencia en el resultado**, además de **ofrecer ayudas para una mejor interpretación de los resultados**. Está pensada para su **aplicación en la toma de decisiones clínicas**

2 partes:

A. 3 preguntas de eliminación

B. 17 items: 5 de calidad, 5 de aspectos clínicos, 7 de metodología/estadística

**PARTE A/ ¿SON VÁLIDOS LOS RESULTADOS DE LA COMPARACIÓN INDIRECTA?**

<p>- <b>¿Está justificado realizar una comparación indirecta (CI)?</b></p> <p><i>¿Existe una comparación directa? En general, se considera que las comparaciones directas tienen menos sesgos que las indirectas. Si hay una comparación directa y una indirecta puede ser interesante valorar si llegan a la misma conclusión. Si no hay diferencias aumentará nuestra fiabilidad en el resultado y si hay diferencias debemos cuestionarnos a qué pueden deberse esas diferencias.</i></p> <p style="text-align: center;"><b>¿HAY COMPARACIÓN DIRECTA?</b></p> <p><i>Para valorar si se puede hacer una comparación indirecta se debe cumplir:</i></p> <ul style="list-style-type: none"> <li>- <i>¿Están claros el objetivo, la población, los métodos (incluidas asunciones) y los resultados? (si no hay transparencia en todos estos aspectos no se puede o dificulta la evaluación de la CI)</i></li> <li>- <i>¿Los estudios comparados son comparables clínicamente? (Características de la población estudiada, condiciones coexistentes, cointervenciones, dosificación, duración del tratamiento y resultados evaluados que influyen o modifican la magnitud del efecto de la intervención)</i></li> <li>- <i>¿La medida de resultados es comparable en cuanto a su definición y su presentación?</i></li> <li>- <i>¿La medida de resultados es clínicamente relevante?</i></li> </ul>	<p><input type="checkbox"/> SÍ    <input type="checkbox"/> NO <input type="checkbox"/> NO SÉ</p>
<p>- <b>¿El método de CI es apropiado?</b></p> <p><i>Las comparaciones indirectas no ( <b>¿MÉTODO AJUSTADO?</b> ) lativo, sino que cogen un brazo de un estudio sin considerar el comparador) se c</i></p> <p>- <i>Métodos de comparaciones indirectas: Bucher, Metanálisis de redes de tratamiento, mezclas de comparaciones indirectas, metanálisis de múltiples tratamientos.</i></p>	<p><input type="checkbox"/> SÍ    <input type="checkbox"/> NO <input type="checkbox"/> NO SÉ</p>
<p>- <b>¿Se realizó una búsqueda adecuada de todos los estudios pertinentes?</b></p> <p><i>¿Se dirige adecuadamente ( <b>¿BÚSQUEDA TIPO REVISIÓN SISTEMÁTICA?</b> ) ntion; Comparison; Outcome)</i></p> <p><i>¿Están incluidos todos los estudios relevantes?</i></p> <p><i>¿El comparador es relevante?</i></p>	<p><input type="checkbox"/> SÍ    <input type="checkbox"/> NO <input type="checkbox"/> NO SÉ</p>



# Gracias por su atención

[mdfraga@jccm.com](mailto:mdfraga@jccm.com)