

Buone pratiche per la costruzione di confronti indiretti fra tecnologie sanitarie

PREFAZIONE E GRUPPO DI LAVORO

La valutazione comparativa tra una nuova tecnologia sanitaria¹ (farmacologica o non) e le alternative terapeutiche esistenti è resa complessa dalla frequente assenza di studi clinici comparativi. Una possibile opzione è quella di evitare tale confronto e di giudicare non valutabile o quantificabile il valore terapeutico aggiunto. Rinunciare a tale valutazione significa però non fornire un supporto importante a tutte le politiche che necessitano di una valutazione comparativa (prezzo e rimborsabilità, acquisti, indirizzo al comportamento prescrittivo), nonché alle scelte terapeutiche. Un'alternativa più efficace consiste nell'impiego di confronti indiretti, i quali, se condotti con un approccio metodologicamente solido, possono fornire ai decisori strumenti affidabili per valutare l'efficacia relativa delle diverse opzioni terapeutiche. Tali analisi consentono di formulare decisioni informate, supportate da evidenze che, pur riconoscendo le incertezze esistenti, offrono un quadro più chiaro per l'allocazione delle risorse e il miglioramento della pratica clinica.

La stesura di un documento italiano di buona pratica sulla conduzione dei confronti indiretti è cruciale per vari motivi: l'Italia, come altri paesi, basa le decisioni fondamentali di prezzo e rimborsabilità dei nuovi trattamenti, inclusa l'innovatività e gli indirizzi al comportamento prescrittivo, anche sul valore terapeutico aggiunto; inoltre, ad oggi non esistono raccomandazioni specifiche sull'uso dei confronti indiretti a livello nazionale e questo genera un gap informativo rilevante in termini di trasparenza e di interlocuzione tra pubblico e privato; infine, la recente entrata in vigore del Regolamento Europeo di *Health Technology Assessment* (HTA) rappresenta una opportunità importante di innovazione e standardizzazione dei processi metodologici.

I modelli di confronti indiretti sono destinati a diventare sempre più diffusi per rispondere ai requisiti dei PICO (Popolazione, Intervento, Comparatore, Outcome) nei dossier di *Joint Clinical Assessment* (JCA). È implicito che tali dossier dovranno essere considerati con attenzione dai Paesi Membri dell'UE. Per questo, l'HTA *Coordination Group* (HTACG) ha reso disponibile sul portale della Comunità Europea le seguenti due linee guida sull'argomento, adottate l'8 marzo 2024:

1. *Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons*²
2. *Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons*³

¹ Per tecnologia sanitaria, oltre al trattamento farmacologico, in senso più ampio si includono anche i dispositivi medici, i dispositivi medici diagnostici in vitro e le procedure mediche oltre alle misure per la prevenzione, la diagnosi o il trattamento delle malattie. Le tecnologie sanitarie sono o saranno oggetto di valutazione secondo il Regolamento Europeo HTA.

² Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons Adopted on 8 March 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment Disponibile su: https://health.ec.europa.eu/document/download/1f6b8a70-5ce0-404e-9066-120dc9a8df75_en?filename=hta_practical-guideline_direct-and-indirect-comparisons_en.pdf Accesso il 27 gennaio 2025

³ Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons Adopted on 8 March 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment.

Questo documento raccoglie le principali indicazioni tecniche riportate nelle linee guida europee sopra citate, adattandole con un approccio più divulgativo al fine di contribuire alla diffusione di tematiche complesse, come i modelli di confronto indiretto, favorendo una maggiore consapevolezza e applicazione pratica da parte di professionisti coinvolti nella creazione e valutazione del valore di una tecnologia sanitaria. Allo stesso modo, il documento include i riferimenti bibliografici utili per chi desidera approfondire gli aspetti metodologici dei modelli. L'obiettivo ultimo del documento è supportare i diversi attori del sistema Paese nel raggiungimento di un linguaggio comune e nell'adozione di una metodologia uniforme nello sviluppo e valutazione dei modelli di confronti indiretti.

In coerenza con la mission che l'ISPOR Italy Rome Chapter sta portando avanti da oltre 10 anni, questo documento ha cercato di coinvolgere tutti gli stakeholder del sistema salute provando a raccogliere, integrare e riassumere le principali evidenze disponibili sul tema dei confronti indiretti. Questo, con la speranza di produrre una buona pratica che aiuti i decisori pubblici, le aziende private ed i cittadini a meglio comprendere alcuni aspetti chiave dei processi comparativi e delle decisioni che impatteranno sulla definizione di "valore" delle tecnologie sanitarie.

Un sentito ringraziamento va a tutti i membri del Gruppo di Lavoro e ai Supervisor del progetto per l'impegno profuso, le preziose interazioni e il costruttivo scambio di opinioni su questo importante tema.

Carlotta Galeone¹ & Andrea Marcellusi^{1,2}

¹Project Coordinator

² President ISPOR Italy - Rome Chapter

Coordinatori del Gruppo di Lavoro

Carlotta Galeone, Andrea Marcellusi

Membri del Gruppo di Lavoro

Martina Bianco, Giovanni Biricolti, Guido Condorelli, Paolo Di Rienzo, Daniela Ghislieri, Sara Graziadio, Antonello Lenti, Stefania Lopatriello, Barbara Muzii, Salvatore Pirri, Massimiliano Povero, Martina Rizzato, Federico Tartarelli

Supervisori del documento

Claudio Jommi, Federica Turati, Entela Xoxi

Un ringraziamento speciale a Angela Ragonese per il contributo alla gestione del progetto editoriale e a Marco Palone per il contributo nella realizzazione grafica delle figure.

ACRONIMI

AgD	Aggregate Data
ECA	External Control Arm
H2H trial	Head-to-Head trial
HTA	Health Technology Assessment
ITC	Indirect Treatment Comparison
IPD	Individual Patient Data
MAIC	Matching Adjusted Indirect Comparison
ML-NMA	Multilevel Network Meta-Regression
NMA	Network Meta-Analysis
PICO	P: Population; I: Intervention; C: Comparator; O: Outcome
PS	Propensity Score
RW	Real World
SAP	Statistical Analysis Plan
SAT	Single Arm Trial
SLR	Systematic Literature Review
STC	Simulated Treatment Comparison

Sommario

1. Introduzione.....	8
1.1 Definizione di confronti diretti e indiretti fra trattamenti.....	8
1.1.1 Utilizzo di dati individuali (patient-level) e aggregati (study-level)	10
1.2 Obiettivi del documento	11
1.3 L'attuale utilizzo degli ITC nell'HTA nei maggiori Paesi Europei	14
2. Fasi preparatorie per la conduzione di ITC	16
2.1 Punto di partenza: sintesi delle prove attraverso una revisione sistematica della letteratura. 16	
2.1.1 Selezione degli studi per la sintesi delle evidenze.....	17
2.2 Valutazione della condizione di trasferibilità (<i>exchangeability</i>)	19
2.2.1 Proprietà di similarità (similarity).....	22
2.2.2 Proprietà di omogeneità (homogeneity)	24
2.2.3 Proprietà di consistenza (consistency).....	27
2.3 Possibili azioni da intraprendere se le assunzioni di trasferibilità sono violate.....	30
2.4 Valutazione dei dati mancanti	32
3. I modelli di confronti indiretti.....	33
3.1 Introduzione e focus sui modelli naïve	33
3.1.1 Il livello di evidenza degli ITC per macro-aggregazione.....	36
3.2 Confronti indiretti con <i>network</i> di studi ancorati	37
3.2.1 Metodo di Bucher	37
3.2.2 Network Meta-Analisi	38
3.2.3 Take home message	39
3.3 Confronti indiretti basati su metodi che aggiustano per le caratteristiche della popolazione 40	
3.3.1 Matching-adjusted Indirect Comparison (MAIC)	42
3.3.2 Simulated Treatment Comparison (STC).....	42
3.3.3 Confronto fra STC e MAIC e altri modelli	43
3.3.4 Multilevel Network meta-regression (ML-NMR).....	44
3.3.5 Take home message	45
3.4 Confronti indiretti basati su evidenze non randomizzate	46
3.4.1 Considerazioni generali.....	46

3.4.2 Propensity Score	47
3.4.3 Take home message	49
3.5 Conclusioni.....	50
3.5.1 Piramide delle evidenze indirette.....	50
3.5.2 Scelta del modello di confronto indiretto più robusto in funzione all'accessibilità del dato	50
3.5.3 Check list delle fasi di costruzione o validazione di una evidenza indiretta	52
 GLOSSARIO	 53
 RIFERIMENTI BIBLIOGRAFICI	 55

1. Introduzione

1.1 Definizione di confronti diretti e indiretti fra trattamenti

Al fine di valutare l'*effectiveness*^{4,5} (intesa come l'efficacia e sicurezza relativa) di una tecnologia sanitaria di interesse⁶ (es. Farmaco A) rispetto ad un comparatore (es. Farmaco B) possono verificarsi due scenari:

A) Confronto del farmaco A vs farmaco B per via **DIRETTA**

Definizione: un confronto diretto avviene quando due trattamenti (A vs B) vengono messi a confronto all'interno dello stesso studio clinico randomizzato (RCT), detto anche Head-to-Head (H2H) Trial.

Questo metodo è considerato il *gold standard* per i confronti fra trattamenti, poiché, se l'RCT è ben disegnato con un'adeguata dimensione del campione, la randomizzazione è in grado di minimizzare il rischio di *bias* (distorsione) dovuto a differenze nelle popolazioni studiate (*selection bias*).

Qualora fossero disponibili più evidenze dirette di confronto di A vs B condotte su popolazioni simili e omogenee, l'evidenza di confronto diretto viene costruita applicando i metodi di una *pairwise* meta-analisi [1].

Figura 1 – Confronto diretto fra due trattamenti



⁴ "For simplicity, we use *effectiveness* as the common term to describe efficacy or effectiveness throughout the rest of this document. *Effectiveness* also includes safety within the context of this document." (Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons. Adopted on 8 March 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment)

⁵ *Effectiveness: describes how well a treatment works in patients; includes efficacy and safety.* (Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons Adopted on 8 March 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment)

⁶ In questo documento si farà prevalentemente riferimento al concetto di valutazione delle tecnologie sanitarie, in conformità con il Regolamento Europeo HTA. Il termine "tecnologia sanitaria" comprende sia i trattamenti farmacologici sia i trattamenti non farmacologici, come i dispositivi medici. Tuttavia, per rendere l'esposizione più chiara e fluida, il termine "trattamento" viene in alcuni casi utilizzato come sinonimo di tecnologia sanitaria.

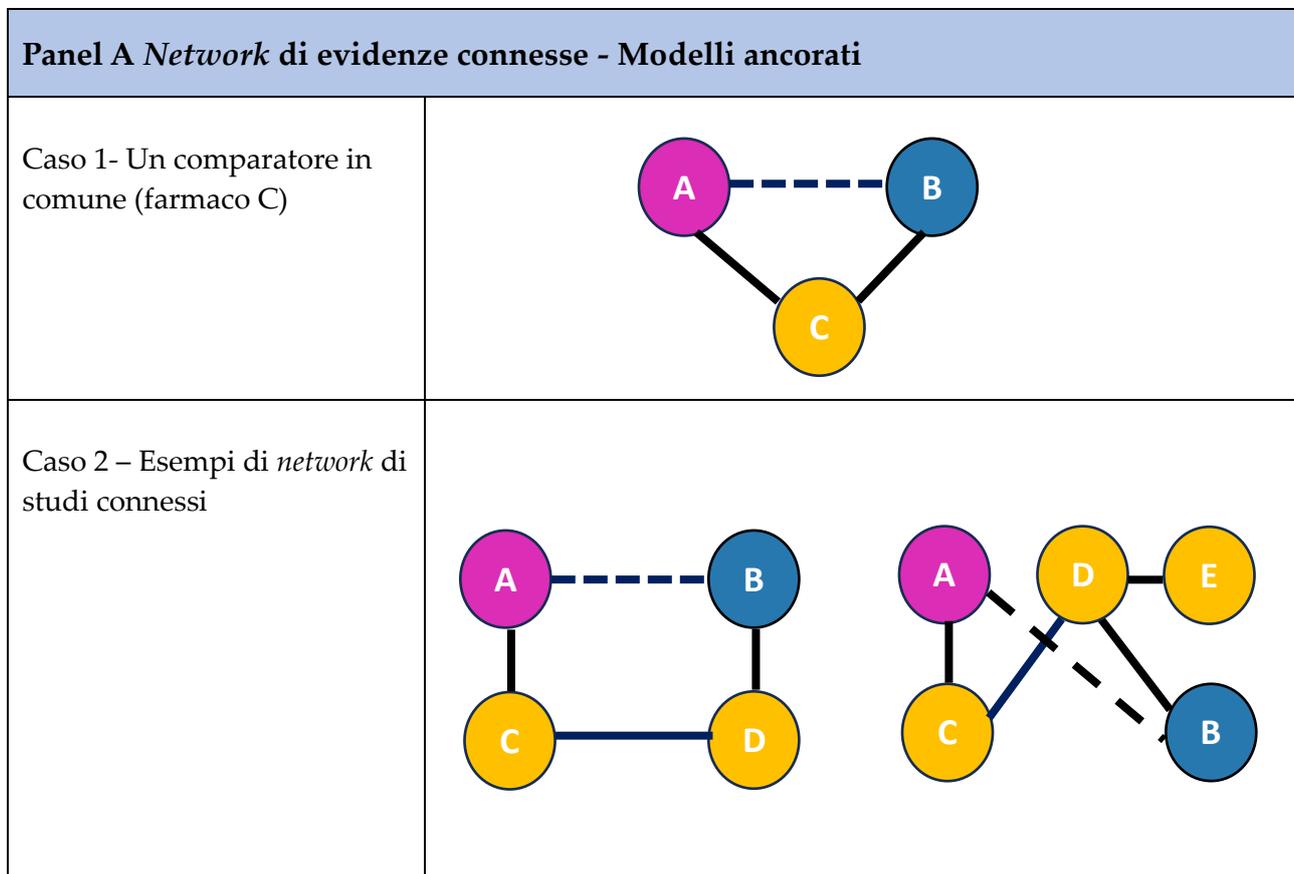
B) Confronto del farmaco A vs farmaco B per via **INDIRETTA**

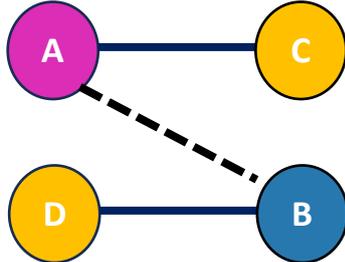
*Definizione: un confronto indiretto viene utilizzato quando **non sono disponibili studi clinici che confrontano direttamente i trattamenti tra loro**. In questo caso, si utilizzano dati provenienti da studi diversi che possono avere o meno un comparatore comune.*

I modelli di confronti indiretti (ITC) sono una famiglia di modelli, in cui la scelta del modello più appropriato dipende dal contesto specifico. L'ITC si basa su modelli statistici specifici con l'obiettivo di superare il problema delle differenze tra i gruppi di pazienti confrontati nei vari studi e di ottenere così risultati interpretabili correttamente.

A seconda del *network* delle evidenze dirette disponibili è possibile confrontare il farmaco A vs il farmaco B per via indiretta, seguendo una rigorosa metodologia statistica. Alcuni esempi di *network* di studi finalizzati al confronto indiretto di A vs B sono qui riportati in Fig. 2.

Figura 2 – Confronti indiretti fra trattamenti (linea tratteggiata) – Possibili casi con *network* di evidenze connesse (Panel A) e non connesse (Panel B)



Panel B <i>Network</i> di evidenze non connesse - Modelli non ancorati	
<p>Caso 3 - Senza un comparatore in comune Due Single Arm Trial (SAT) a confronto</p>	
<p>Caso 4 - Senza un comparatore in comune Due RCT non connessi fra loro</p>	

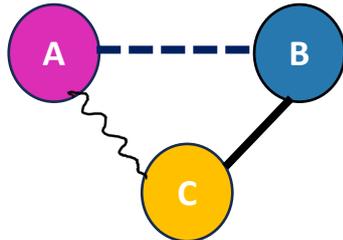
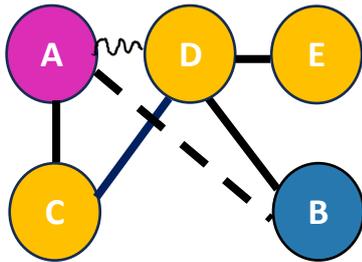
1.1.1 Utilizzo di dati individuali (*patient-level*) e aggregati (*study-level*)

In generale, nella ricerca clinica i dati individuali (*patient-level*) detti anche *individual patients data - IPD*) e i dati aggregati (*study-level; aggregate data, AgD*) rappresentano due approcci di analisi, ciascuno con specifici vantaggi e limitazioni.

Sebbene la sintesi delle evidenze combini tipicamente le stime di effetto *study-level* (ad esempio, nelle NMA, *network meta-analyses*), è anche possibile aggregare IPD provenienti da diversi studi (ad esempio, nelle MAIC, *matching adjusting indirect comparisons* e nei STC, *simulated treatment comparisons*). Queste tecniche - che utilizzano gli IPD - sono particolarmente utili quando emerge che le popolazioni degli studi oggetto di analisi siano troppo diverse per permettere una sintesi dei dati aggregati che non sia affetta da gravi distorsioni nell'interpretazione della stima finale. **Per questo le analisi statistiche che utilizzano IPD dovrebbero essere preferite rispetto a quelle che usano solo statistiche riassuntive.** La sintesi delle evidenze basata su IPD completi, cioè che includono le caratteristiche al *baseline* dei pazienti, offre migliori possibilità di modellazione rispetto alle analisi di AgD. In particolare, la disponibilità di IPD consente analisi valide di sottogruppo e aggiustamenti statistici riguardo alle caratteristiche dei pazienti. Tuttavia, gli IPD non sono frequentemente disponibili e non possono essere ricostruiti in tutti i dettagli necessari, limitando così l'uso della sintesi delle evidenze basata su IPD.

I *network* di connessione degli studi possono comprendere anche connessioni basate su IPD, come analogamente riportato nella Figura 2 quando si ha accesso ai soli dati aggregati (*study-level*), come riportato in Figura 3.

Figura 3 – Confronti indiretti fra trattamenti (linea tratteggiata) – Possibili casi con *network* di evidenze connesse (Panel A) e non connesse (Panel B) con accesso ai dati IPD (linea ondulata)

Panel A <i>Network</i> di evidenze connesse – Modelli ancorati	
Caso 1- Un comparatore in comune (farmaco C) con accesso al dato IPD per il confronto A vs C	
Caso 2 – Esempi di <i>network</i> di studi connessi con accesso al dato IPD per il confronto A vs D	
Panel B <i>Network</i> di evidenze non connesse – Modelli non ancorati	
Caso 3 - Senza un comparatore in comune Due SAT a confronto con accesso ai dati IPD di A	

1.2 Obiettivi del documento

L'obiettivo di questo documento è descrivere i metodi attualmente disponibili per la creazione di evidenze di confronti indiretti fra tecnologie sanitarie (trattamenti farmacologici e non o dispositivi medici), evidenziando i punti di forza e di debolezza con la relativa appropriatezza metodologica, cercando di contestualizzarli all'ambito italiano.

In Italia non esiste, ad oggi, una chiara linea guida sull'utilizzo delle analisi ITC ai fini delle valutazioni delle tecnologie sanitarie (*Health Technology Assessment*, HTA) [2].

La stesura di una buona pratica per la conduzione di ITC fra tecnologie sanitarie in Italia è importante per vari aspetti chiave, come ad esempio:

1. **Migliorare la valutazione comparativa delle tecnologie sanitarie:** spesso, le nuove tecnologie non sono state testate direttamente (H2H) contro quelle già esistenti. Ciò avviene soprattutto per patologie complesse con diverse opzioni terapeutiche. Gli ITC permettono di valutare l'efficacia e la sicurezza comparativa (*relative effectiveness*) di trattamenti che non sono stati oggetto di studi clinici diretti.
2. **Standardizzare i metodi di confronto:** la metodologia utilizzata nella conduzione degli ITC è fondamentale ed è necessario definire criteri chiari, riducendo così il rischio di *bias* e di errori di interpretazione. Questo garantisce che tali confronti (indiretti) siano effettuati in modo rigoroso e riproducibile, e che i risultati siano confrontabili tra diversi studi e contesti clinici.
3. **Supportare il processo decisionale di rimborsabilità e prezzo:** In Italia, l'Agenzia Italiana del Farmaco (AIFA) valuta l'efficacia comparativa dei farmaci per decidere quali farmaci rendere disponibili attraverso il Servizio Sanitario Nazionale (rimborsabilità) definendo anche il prezzo. Gli ITC possono fornire evidenze utili per giustificare il valore terapeutico di un farmaco rispetto ad altri e aiutare a negoziare prezzi che riflettano, tra gli altri elementi, tale valore terapeutico aggiunto.
4. **Supportare le scelte terapeutiche e le azioni di indirizzo a tali scelte:** La disponibilità di valutazioni comparative consente di gestire meglio le scelte terapeutiche in presenza di diverse alternative e le eventuali politiche di indirizzo a tali scelte da parte delle Regioni e Aziende Sanitarie.
5. **Ottimizzare l'uso delle risorse sanitarie:** L'Italia, come molti altri paesi, deve affrontare vincoli di *budget* nel settore sanitario. È importante utilizzare al meglio tale *budget*, per massimizzare, per date risorse esistenti, i benefici generati. Disporre di dati affidabili sulla comparabilità delle tecnologie aiuta a scegliere quelle che presentano il miglior profilo di costo-efficacia.
6. **Promuovere trasparenza e fiducia:** Una buona pratica aumenta la trasparenza del processo di valutazione dei trattamenti da parte di tutti gli *stakeholder* coinvolti. Fornire un quadro metodologico chiaro per gli ITC riduce il rischio di conflitti di interesse e aumenta la fiducia nelle decisioni di adozione di una certa tecnologia.
7. **Adattarsi al contesto regolatorio e HTA internazionale:** In molti paesi e a livello europeo, gli ITC sono una prassi consolidata nelle valutazioni comparative delle tecnologie. Dotarsi di un documento di buona pratica italiana permette di allinearsi alle *best practices*

internazionali e di favorire l'integrazione dei dati italiani nei contesti regolatori europei, soprattutto in considerazione del Regolamento HTA che è entrato in vigore a gennaio 2025.

Concludendo, una buona pratica italiana sugli ITC tra tecnologie sanitarie rappresenta uno strumento di valore per garantire scelte terapeutiche più accurate, efficaci e sostenibili, sostenendo una sanità più equa e basata sull'evidenza.

1.3 L'attuale utilizzo degli ITC nell'HTA nei maggiori Paesi Europei

L'articolo intitolato "The Acceptance of Indirect Treatment Comparison Methods in Oncology by Health Technology Assessment Agencies in England, France, Germany, Italy, and Spain" ha valutato l'accettazione dei metodi di ITC da parte delle agenzie di Health Technology Assessment (HTA) in cinque paesi europei, focalizzandosi sulle valutazioni di trattamenti oncologici [2].

I ricercatori hanno analizzato 543 rapporti di valutazione HTA pubblicati tra aprile 2018 e aprile 2021 riguardanti trattamenti oncologici per tumori solidi. Di questi, 120 rapporti (22%) includevano almeno un modello di ITC. La distribuzione dell'inclusione di ITC variava tra i paesi, con l'Inghilterra al 51% e la Francia al 6%. In Italia su 123 valutazioni di HTA, 29 presentavano almeno un risultato di un modello di ITC (24%) (Tabella 1).

Tabella 1. Numero di valutazioni HTA che includono una ITC per paese

	England	France	Germany	Italy	Spain	Total
Number of HTA evaluation reports	75	177	120	123	48	543
Number of HTA evaluation reports presenting at least one ITC	38	11	21	29	21	120
Proportion of HTA evaluation reports presenting at least one ITC	51%	6%	18%	24%	44%	22%

Abbreviations: HTA health technology assessment, ITC indirect treatment comparison

La percentuale complessiva di rifiuto dei metodi ITC dalle Agenzie di HTA è stata globalmente del 21%, con il più alto tasso osservato in Francia (91%), seguita da Germania (33%), Inghilterra (18%) e Spagna (5%), e il più basso in Italia (0%) (Tabella 2). Tuttavia, emerge chiaramente che in Spagna e Italia c'è una mancata chiarezza nell'accettazione dei metodi di ITC nelle valutazioni di HTA (rispettivamente il 90% e il 69%).

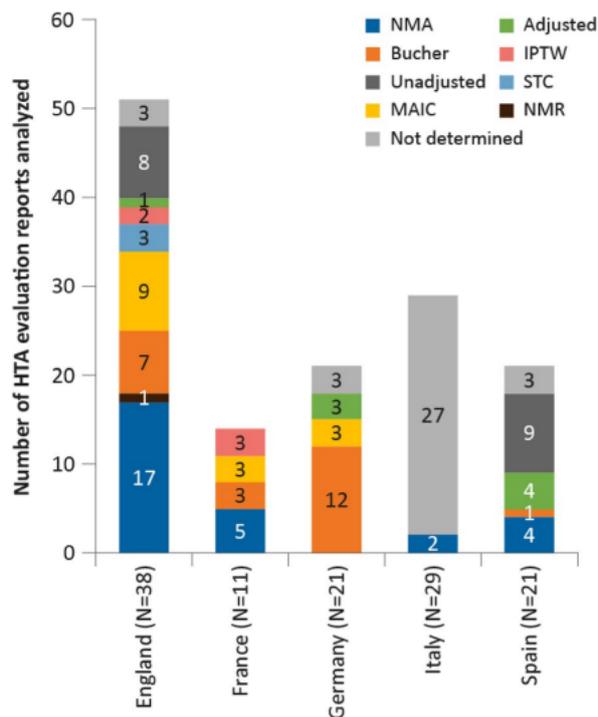
Tabella 2. Tasso complessivo di accettazione e rifiuto di ITC nelle valutazioni HTA per paese

Number of HTA evaluation reports including an ITC	England, n (%) N = 38	France, n (%) N = 11	Germany, n (%) N = 21	Italy, n (%) N = 29	Spain, n (%) N = 21	Total, n (%) N = 120
ITC methods accepted	18 (47)	0 (0)	8 (38)	9 (31)	1 (5)	36 (30)
ITC methods rejected	7 (18)	10 (91)	7 (33)	0 (0)	1 (5)	25 (21)
Unclear acceptance of ITC methods	13 (34)	1 (9)	6 (29)	20 (69)	19 (90)	59 (49)

Abbreviations: HTA health technology assessment, ITC indirect treatment comparison

L'analisi per paese ha indicato che fra i vari modelli di ITC disponibili, la NMA è il modello più comunemente utilizzato in Inghilterra, da cui provengono il 45% delle relazioni di valutazione HTA inviate al *National Institute for Health and Care Excellence* (NICE). In Italia l'utilizzo delle NMA viene riportato anche se prevale la quota di tecniche "non determinate" (Figura 4).

Figura 4. Tipi di ITC riportate nelle valutazioni HTA per paese



*In some cases, multiple ITC methods were presented for each evaluation report. *Abbreviations: HTA* health technology assessment, *IPTW* inverse probability treatment weighting, *ITC* indirect treatment comparison, *MAIC* matching adjusted indirect comparison, *NMA* network meta-analysis, *NMR* network meta-regression, *STC* simulated treatment comparison

Alla luce di questi risultati, una maggior conoscenza sulla metodologia di conduzione degli ITC da parte degli *stakeholder* nazionali, veicolata anche tramite l'ausilio di questo documento, potrebbe contribuire a una maggior consapevolezza nella valutazione della solidità di un'evidenza indiretta.

2. Fasi preparatorie per la conduzione di ITC

Prima di approfondire gli aspetti metodologici per la costruzione del modello di ITC più idoneo per la valutazione di tecnologie sanitarie di interesse, questo capitolo esaminerà le fasi preliminari essenziali per rispondere ai quesiti comparativi.

2.1 Punto di partenza: sintesi delle prove attraverso una revisione sistematica della letteratura

Prima di condurre una sintesi delle evidenze - finalizzata a valutare la *relative effectiveness* di una tecnologia - è fondamentale condurre una ricerca sistematica della letteratura (*systematic literature review*, SLR) seguendo una rigorosa metodologia come quella sviluppata dal gruppo internazionale *Cochrane*. Tale SLR [3] prevede:

- L'utilizzo del PICO Framework (*Patient Intervention Control Outcome*) per l'esplicitazione del quesito clinico di ricerca rilevante (Figura 5);
- La stesura di un protocollo di ricerca prima di procedere con l'inizio operativo della ricerca bibliografica e della sintesi quantitativa dei risultati (meglio se registrato su una banca dati come *Prospero* [4]) con pre-specificati i dettagli pratici della ricerca e della sintesi, al fine di assicurare che le decisioni non siano *data-driven*;
- L'applicazione di una strategia di ricerca della letteratura esaustiva e replicabile;
- L'esplicitazione dei criteri di inclusione/esclusione degli studi;
- La trasparenza nella valutazione del *risk of bias* degli studi.

Figura 5. Consigli per l'utilizzo del PICO Framework (*Patient Intervention Control Outcomes*)

	1	2	3	4
	Pazienti, popolazione o problema	Intervento, fattore prognostico o fattore di rischio	Confronto	Outcome di interesse
Suggerimento	Come descriveresti un gruppo di pazienti simile al mio?	Quale l'intervento, fattore prognostico o di rischio mi interessa?	Qual è la migliore alternativa all'intervento considerato?	Rispetto a quale condizione o esito mi attendo dei miglioramenti?
Esempio	Pazienti con scompenso cardiaco da cardiomiopatia dilatativa con ritmo sinusale...	... il trattamento anti-coagulante con varfarina in aggiunta alla terapia standard per lo scompenso cardiaco quando confrontato con la sola terapia standard è associato a una minore mortalità o morbilità per tromboembolismo. Questo beneficio supera il rischio di sanguinamenti dovuti al trattamento con anti-coagulante?
Tradotto da Oxford Centre for Evidence-Based Medicine http://www.cebm.net/focus_quest.asp .				

Come anche raccomandato dal gruppo di ricerca internazionale di *Enhancing the QUALity and Transparency Of health Research (EQUATOR Network [5])*, l'utilizzo di linee guida specifiche per il *reporting* e la conduzione di vari disegni di studi è fondamentale al fine di migliorare la conduzione e la pubblicazione dei risultati. In particolare, per le SLR, è necessario seguire le seguenti linee guida che forniscono un quadro dettagliato su come pianificare, eseguire e riportare in modo completo e coerente:

- Linee guida PRISMA [6] (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*) che si compone di una lista di 27 elementi che dovrebbero essere inclusi nella conduzione e nel report delle revisioni sistematiche di studi clinici;
- Linee guida STROBE [7] (*The Strengthening the Reporting of Observational Studies in Epidemiology*) che si compone di una lista di 22 elementi che dovrebbero essere inclusi nel report delle revisioni sistematiche di studi osservazionali.

2.1.1 Selezione degli studi per la sintesi delle evidenze

Al fine di combinare in modo metodologicamente corretto le stime di *relative effectiveness* delle varie tecnologie sanitarie oggetto della domanda di ricerca formulata tramite il PICO *Framework*, dopo aver condotto una SLR e aver selezionato gli articoli di interesse, è fondamentale valutare con attenzione se gli studi sono sufficientemente simili (concetto di similarità). Come per la conduzione della SLR descritta nel paragrafo precedente, ogni azione per la valutazione della somiglianza degli studi deve essere prodromica alla successiva fase di sintesi delle evidenze (mediante modelli diretti o indiretti) e pre-specificata in un protocollo di ricerca, meglio se accompagnato da un SAP (*statistical analysis plan*) e se registrato su un *repository* di protocolli *online* che certifichi la data di registrazione⁷.

Per valutare potenziali differenze importanti fra studi che, qualora aggregati, porterebbero a delle stime di confronto distorte (*biased*), è fondamentale prendere sempre in considerazione le distribuzioni delle seguenti caratteristiche dei pazienti (se applicabile al singolo caso di studio):

- Età
- Genere
- Etnia
- Durata della malattia

Inoltre, è opportuno confrontare:

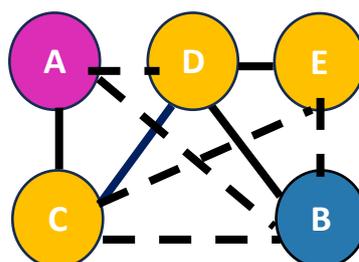
- La definizione degli *outcome* di interesse;

⁷ PROSPERO: È un registro internazionale per revisioni sistematiche e meta-analisi (inclusi i network meta-analysis. <https://www.crd.york.ac.uk/prospero/>. Non esiste ad oggi un registro nazionale e/o internazionale per i protocolli di studio di modelli di confronti indiretti.

- Le caratteristiche del disegno di studio (come, ad esempio, la fase di studio, il braccio di confronto, ...)

È tuttavia necessario considerare ulteriori aspetti, come l'anno di conduzione degli studi e la distribuzione geografica dei centri arruolatori, le modalità di somministrazione del trattamento e la durata del trattamento e del *follow-up*. Particolare attenzione deve essere posta alla valutazione della somiglianza fra gli studi rispetto alla distribuzione dei potenziali modificatori d'effetto (*effect modifiers*). Un modificatore d'effetto è un fattore che influisce sull'*effectiveness* delle tecnologie sanitarie. In altre parole, si tratta di una variabile (come, ad esempio, l'età) che "modifica" il rapporto tra l'esposizione a un trattamento e l'*outcome*, cambiando il modo in cui un'esposizione (come un medicinale o un dispositivo) produce un determinato risultato. A titolo esemplificativo, in un'analisi sull'*effectiveness* di un medicinale, l'età può risultare un effetto modificatore se l'*effectiveness* risultasse essere maggiore nei giovani rispetto agli anziani, o viceversa. Una volta determinate le tecnologie sanitarie da confrontare nell'analisi secondo il PICO *framework* e individuate le evidenze disponibili provenienti da studi simili, qualora sia necessario utilizzare dei modelli di ITC in assenza di evidenze dirette, è utile produrre una rappresentazione grafica del *network* degli studi. Tale rappresentazione grafica permette di valutare se il *network* di evidenze è connesso (ovvero esiste un percorso, *path*, che collega ogni tecnologia a ogni altra tecnologia del *network*) oppure no (come riportato nella Figura 2). L'esistenza di un *network* di studi connessi rappresenta uno degli elementi che guiderà la scelta del modello di ITC più appropriato. Si riporta qui un esempio di un *network* fra studi, di grande dimensione e complessità che, qualora fosse soddisfatta la condizione di similarità fra studi, consentirebbe di costruire molteplici evidenze indirette fra trattamenti.

Figura 6 – Esempio di un *network* fra studi, di grande dimensione



2.2 Valutazione della condizione di trasferibilità (*exchangeability*)

Dopo aver selezionato gli studi rilevanti sulla base del PICO, avendo eventualmente escluso quelli troppo dissimili in termini di popolazione, trattamento e disegno di studio (come discusso nel paragrafo precedente), prima di procedere con la sintesi delle evidenze per via indiretta (in particolare per il *network* di studi connessi) è fondamentale valutare l'assunzione di trasferibilità (*exchangeability*). Tale valutazione si realizza mediante un'attenta analisi delle seguenti proprietà (Tabella 3):

- **similarità** (*similarity*)
- **omogeneità** (*homogeneity*)
- **consistenza** (*consistency*)

Queste tre proprietà non sono assunzioni distinte ma sono interconnesse fra loro, poiché un'assenza di omogeneità o consistenza è spesso il risultato di uno squilibrio nei modificatori di effetto tra gli studi (cioè una violazione della similarità).

Tabella 3 – Sintesi della valutazione della condizione di trasferibilità (*exchangeability*) e delle sue 3 proprietà

L'assunzione di **trasferibilità** è fondamentale nella sintesi delle evidenze. Essa implica che, se i partecipanti di uno studio fossero sostituiti in un altro, l'effetto della tecnologia sanitaria rimarrebbe lo stesso. Per verificare questa assunzione, si analizzano le proprietà di **similarità** e **omogeneità**, e nel caso di comparazioni indirette, di **consistenza** fra le stime ottenute tramite confronti diretti e indiretti.

È fondamentale valutare attentamente l'assunzione di trasferibilità prima di procedere con una sintesi delle evidenze, poiché una violazione di questa assunzione può portare a risultati distorti.

In caso di violazione di questa assunzione è necessario procedere con delle azioni correttive prima di procedere con la sintesi delle evidenze, come la suddivisione in sottogruppi, l'utilizzo di tecniche di meta-regressione, l'esclusione di studi, l'analisi di sensitività. Nel caso l'applicazione di tali azioni correttive non fosse possibile e/o non portasse all'accettazione dell'assunzione di trasferibilità è fondamentale non procedere con la sintesi delle evidenze aggregate (tramite, ad esempio, la NMA) e procedere con l'applicazione di modelli aggiustati per la popolazione (come tramite MAIC e STC qualora si abbia accesso ai dati a livello individuali di pazienti per almeno una tecnologia sanitaria in esame).

<p>Similarità (<i>similarity</i>)</p>	<p>Per validare l'assunzione di trasferibilità è necessaria una sufficiente similarità tra gli studi riguardo ai modificatori di effetto. Se ci sono differenze sostanziali nel disegno dello studio o nelle caratteristiche dei pazienti, l'affidabilità dei risultati può essere compromessa.</p>	<p><i>Nota: le ipotesi non sono necessariamente completamente distinte; ad esempio, differenze nella similarità possono causare eterogeneità e inconsistenza</i></p>
<p>Omogeneità (<i>homogeneity</i>)</p>	<p>L'assunzione di trasferibilità richiede anche che l'efficacia relativa tra i trattamenti sia sufficientemente omogenea tra gli studi. Se i risultati sono molto diversi si osserva eterogeneità, e combinare i</p>	

risultati potrebbe non essere appropriato. L'eterogeneità può essere:

- Statistica: esistono dei test, anche se poco potenti, per valutare se i risultati degli studi differiscono significativamente;
- Clinica: viene valutata rispetto alla variabilità dei criteri di inclusione degli studi (ad esempio genere, linea di terapia), intervento (ad esempio dosaggio e posologia) e *outcome* (ad esempio i *time point* a cui viene valutato);
- Metodologica: è relativa al disegno dello studio (es. fase di studio)

Le analisi dei sottogruppi e la meta-regressione possono aiutare a esplorare ulteriormente l'eterogeneità e a identificare i fattori che la influenzano.

Consistenza
(*consistency*)

Le evidenze indirette e dirette dovrebbero essere concordanti nella direzione e intensità (tale condizione è verificabile solo in caso di network di studi connessi).

Proprietà da valutare per sintesi che prevedono sia confronti diretti che indiretti per una coppia di tecnologie sanitarie

2.2.1 Proprietà di similarità (similarity)

L'ipotesi di similarità afferma che tutti gli studi considerati sono confrontabili per quanto riguarda i possibili modificatori d'effetto per tutte le diverse tecnologie sanitarie oggetto di interesse. Per testare tale ipotesi è necessario partire dall'approccio del PICO *framework* e utilizzare le raccomandazioni delle linee guida europee sulle valutazioni degli effetti dei trattamenti [8]. Per valutare la similarità dovrebbero essere valutate le distribuzioni delle seguenti variabili al fine di identificare possibili modificatori di effetto:

1. Caratteristiche dello studio e del paziente (compresa la durata del follow-up): dovrebbe essere pre-specificata in un protocollo una lista di possibili modificatori d'effetto (si rimanda al paragrafo 1.5). La scelta di queste variabili deve partire non solo da considerazioni cliniche, ma anche dai risultati di altri studi sulla medesima indicazione terapeutica. Solitamente le variabili importanti sono: età, genere, gravità della malattia, regione di provenienza/etnia e durata dello studio. Solo i fattori identificati come possibili modificatori d'effetto dovrebbero essere inclusi in questa lista. Tali variabili dovrebbero essere identificate (a-priori) attraverso una ricerca bibliografica, il contributo di professionisti sanitari esperti e altri metodi (se necessario);

2. Caratteristiche dell'intervento e del comparatore: esempi tipici sono la posologia, il modo di somministrazione e i trattamenti concomitanti.

La valutazione della similarità dovrebbe considerare anche fattori metodologici che non dovrebbero differire sostanzialmente tra gli studi (come, ad esempio, la fase dello studio). Anche la valutazione dei valori osservati degli *outcome* di interesse si è dimostrata utile per valutare tale proprietà:

3. Caratteristiche degli outcome (ad esempio, definizione degli outcome): di solito è difficile definire a priori ciò che viene considerato sufficientemente simile per ciascuna definizione di *outcome*. Dipenderà spesso anche da ciò che è presente negli studi inclusi; tuttavia, anche questo aspetto non deve essere sottovalutato.

4. Valori osservati degli outcome di interesse al baseline: un esame dei valori osservati degli *outcome* di interesse al baseline può fornire informazioni sulla similarità delle popolazioni negli studi, in particolare dei bracci di studio in cui viene utilizzato il comparatore. Tuttavia, per determinare la similarità non è necessario che i valori osservati degli *outcome* siano identici, poiché la distribuzione dei fattori prognostici potrebbe differire tra gli studi. In ogni caso, non dovrebbero esistere differenze molto marcate nei valori degli *outcome* tra gli studi.

È importante notare le seguenti questioni relative alla modifica dell'effetto:

- Per una data misura dell'effetto della tecnologia sanitaria, non tutte le variabili prognostiche sono modificatori di effetto, quindi uno squilibrio nelle variabili prognostiche tra gli studi non indica necessariamente dissimilarità;

- La modifica dell'effetto è una proprietà dell'effetto relativo tra una coppia di tecnologie sanitarie. Pertanto, è possibile che una variabile modifichi l'effetto relativo della tecnologia A rispetto alla tecnologia B, ma non l'effetto della tecnologia A rispetto alla tecnologia C.

Questo potrebbe verificarsi, ad esempio, quando A è un placebo, B è una terapia mirata a una particolare mutazione genetica [ad esempio, un inibitore tirosin-chinasi del recettore del fattore di crescita epidermico (EGFR TKI)] e C è un altro trattamento attivo non specifico per questa mutazione (ad esempio, chemioterapia): in questo caso, la presenza della mutazione genetica in un individuo potrebbe essere un modificatore dell'effetto per il confronto A vs B ma non per il confronto A vs C. In generale, una variabile potrebbe essere un modificatore d'effetto sia per A vs B che per A vs C, ma l'entità e anche la direzione di questo effetto potrebbero differire tra i confronti (ad esempio, se i pazienti rispondessero meno bene alla chemioterapia in presenza di questa mutazione genetica).

Nella conduzione di una NMA è importante che la valutazione dei potenziali modificatori d'effetto sia effettuata per ciascun confronto a coppie nel network degli studi, e che l'ipotesi di similarità sia verificata per tutte le variabili. L'assunzione di similarità richiede che qualsiasi variabile che funge da modificatore dell'effetto per uno o più confronti a coppie nel network di studi sia distribuita in modo simile in tutti gli studi della rete. Infatti, per garantire validità dei confronti indiretti di una NMA è fondamentale verificare che i modificatori dell'effetto siano distribuiti in modo uniforme tra gli studi della rete, così che le differenze tra studi non influenzino i risultati.

La valutazione di similarità dovrebbe includere un'analisi quantitativa dell'impatto di tutte le variabili dei pazienti osservate. Tuttavia, i test statistici per la valutazione della modifica dell'effetto, basandosi sui risultati dell'analisi di sottogruppo degli studi clinici (ad esempio, il test per la significatività dell'interazione), spesso sono poco potenti e hanno problemi legati alla molteplicità (*multiplicity*). Dato che il rischio di errore di tipo 1 (errore alpha - di prima specie - cioè di osservare un risultato di efficacia di un trattamento che sia però un falso positivo) e di tipo 2 (errore beta - di seconda specie - cioè di non osservare un risultato di efficacia di un trattamento che sia però un falso negativo) è tipicamente alto, i test statistici per la modifica dell'effetto non dovrebbero essere gli unici utilizzati per giustificare la selezione dei modificatori d'effetto [9-12]. In fase di stesura del protocollo, è auspicabile sempre coinvolgere clinici esperti di patologia per validare la lista dei potenziali modificatori d'effetto.

Dopo la valutazione di tutti questi aspetti, bisogna prendere una decisione se tutti gli studi considerati nella sintesi delle evidenze sono comparabili per quanto riguarda i modificatori dell'effetto tra tutti i trattamenti (similitudine sufficiente) o no (similitudine insufficiente). Se i dati su un modificatore dell'effetto rilevante non sono disponibili da uno o più studi, allora tale valutazione non può essere effettuata e ciò dovrebbe essere chiaramente riportato come una limitazione dell'analisi. È possibile considerare delle *proxy* per il modificatore d'effetto quando mancante, ma prima di applicare tale decisione è necessario fornire prove sufficienti per dimostrare la validità di tale *proxy*. Necessariamente, l'introduzione di *proxy* inserisce un maggior grado di incertezza nell'interpretazione dei risultati finali.

In sintesi, per valutare la similarità fra gli studi è necessario seguire i seguenti step durante tutto il processo di studio (dal protocollo al giudizio finale sulla proprietà in esame)

<p>1 – Fase iniziale Pre-specificato nel protocollo</p>	<p>Descrizione della metodologia utilizzata per identificare i potenziali modificatori d'effetto e indicazione del fatto che la metodologia è adatta a rilevare tutti i possibili modificatori d'effetto.</p>
<p>2 – Fase iniziale Fase di validazione Da esplicitare nel protocollo</p>	<p>Valutazione dell'elenco di tutti i potenziali modificatori d'effetto identificati e probabilità che questo elenco sia completo; se possibile, stime dell'entità e della direzione degli effetti di interazione.</p>
<p>3 – Fase iniziale – se applicabile Da esplicitare nel protocollo</p>	<p>Descrizione di eventuali possibili modificatori d'effetto mancanti e della direzione del potenziale <i>bias</i> dovuto alla modifica dell'effetto, se questo può essere determinato.</p>
<p>4 – Fase conclusiva</p>	<p>Giudizio finale sul fatto che la condizione di sufficiente similarità sia valida o meno, fornendo una motivazione. In caso positivo, continuare a valutare le proprietà di trasferibilità; altrimenti valutare approcci alternativi di sintesi delle evidenze (come l'aggiustamento per popolazione) e, se non fattibili, non proseguire con le analisi di sintesi.</p>

2.2.2 Proprietà di omogeneità (homogeneity)

L'ipotesi di omogeneità afferma che non ci sia eterogeneità significativa tra le stime degli effetti dei singoli studi per ciascun confronto diretto. Anche se gli studi sono sufficientemente simili, è comunque possibile che i dati mostrino un'eterogeneità significativa. L'eterogeneità può essere causata da modificatori d'effetto sconosciuti e/o da fattori inizialmente giudicati sufficientemente simili o non considerati potenziali modificatori d'effetto. Per testare l'ipotesi di omogeneità per un confronto a coppie di trattamenti, in linea di principio devono essere disponibili almeno due studi con confronto diretto, anche se di solito sono necessari almeno cinque studi per una valutazione affidabile [13]. Se è disponibile un solo studio per ciascun confronto a coppie, l'ipotesi di omogeneità non può essere testata. Tuttavia, ciò non impedisce l'esecuzione di un confronto indiretto.

Due approcci statistici ampiamente utilizzati per valutare l'eterogeneità sono rappresentati dal test statistico basato sulla statistica Q (*Q-test*) [14, 15] e dalla misura di eterogeneità I^2 [14, 16] (inserire due referenze) che misura la proporzione di varianza nella meta-analisi spiegata dall'eterogeneità, al netto del caso. Come riportato nelle linee guida del gruppo *Cochrane* per la conduzione di SLR e meta-analisi, le soglie indicative per l'interpretazione di I^2 sono le seguenti [13]:

- 0-40%: potrebbe non essere importante;
- 30-60%: potrebbe rappresentare un'eterogeneità moderata;
- 50-90%: potrebbe rappresentare un'eterogeneità sostanziale;
- 75-100%: notevole eterogeneità.

Tuttavia, l'importanza dei valori I^2 osservati dipende dall'entità e dalla direzione degli effetti del trattamento e dalla forza dell'evidenza circa la presenza di eterogeneità (*p-value* del *Q-test*, incertezza di I^2 o numero di studi).

Un criterio oggettivo comunemente usato per decidere se gli studi non dovrebbero essere combinati è dato dalla significatività statistica del *Q-test* ($p < 0.05$). Tuttavia, è noto che basare le decisioni solo su questo test non è una buona pratica poiché il test statistico è noto per essere poco potente, specialmente nella situazione in cui il network sia basato su pochi studi [14]. Il problema nell'interpretazione del *Q-test* è che il fatto che non risulti significativo non indica necessariamente che non ci sia eterogeneità rilevante. D'altra parte, nel caso di un grande numero di studi, il *Q-test* potrebbe risultare statisticamente significativo anche se il *forest plot* (la rappresentazione grafica di tutti gli studi inclusi, riportata a titolo esemplificativo nella figura 7) mostra un'eterogeneità ridotta. In tali casi, la misura di I^2 può aiutare a quantificare l'entità dell'eterogeneità. Ad esempio, con $I^2 < 50\%$, potrebbe essere appropriato ritenere che non ci sia un'eterogeneità sostanziale anche se il *Q-test* è statisticamente significativo. In sintesi, per la valutazione dell'eterogeneità è consigliabile un'ispezione grafica del *forest plot*, oltre all'uso delle due statistiche precedentemente descritte (*Q-test*, I^2).

Figura 7 – Valutazione visiva dell'eterogeneità delle stime di un network di studi tramite *forest plot*. Il caso 2 riporta un livello di eterogeneità molto più evidente rispetto al caso 1. Oltre alla valutazione visiva, che rappresenta il primo step, è auspicabile procedere con il calcolo delle misure di eterogeneità (*Q-test* e *I²*)



Dopo aver valutato l'eterogeneità, è necessario determinare se esiste un'eterogeneità sostanziale tra le stime degli effetti dei singoli studi per ciascun possibile confronto diretto (omogeneità insufficiente) o meno (omogeneità sufficiente). Se si può stabilire che esiste un'omogeneità sufficiente ed ha senso combinare gli studi inclusi è necessario determinare se utilizzare un modello a effetti fissi (*fixed-effect*) o a effetti casuali (*random-effect*) per la sintesi quantitativa delle evidenze. Un modello a effetti fissi assume un effetto di trattamento comune in tutti gli studi, il che potrebbe essere improbabile in molte situazioni e richiede una rigorosa giustificazione prima dell'analisi dei dati. La giustificazione richiede forte evidenza di similarità degli studi inclusi in termini di modificatori d'effetto *insieme* a un'eterogeneità statistica, clinica e metodologica limitata. Pertanto, l'approccio standard in molti contesti pratici è di utilizzare il modello a effetti casuali, come anche suggerito dal manuale *Cochrane* di conduzione delle meta-analisi.

In sintesi, per valutare l'omogeneità fra studi è necessario seguire i seguenti *step* durante tutto il processo di studio (dal protocollo al giudizio finale sulla proprietà in esame)

<p>1 – Fase iniziale</p> <p>Pre-specificato nel protocollo</p>	<p>Descrizione dei metodi che si intendono utilizzare per valutare l'omogeneità e delle strategie da implementare in caso di presenza di eterogeneità, includendo anche una regola sulla scelta del modello da utilizzare (effetti-fissi o effetti-casuali)</p>
<p>2 – Fase di analisi</p>	<p>Valutazione se le analisi fornite a sostegno dell'ipotesi di omogeneità (compresi i <i>forest plot</i>, i <i>p-value</i> per il <i>Q-test</i> e i valori di I^2) per tutti i confronti a coppie siano sufficienti per dimostrare che è probabile che sia valida tale assunzione.</p>
<p>3 – Fase conclusiva</p>	<p>Giudizio finale sul fatto che:</p> <ul style="list-style-type: none"> - sia valida o meno l'ipotesi di sufficiente omogeneità, comprensivo di motivazione (includendo anche le analisi di sensibilità) - sia adeguato combinare gli studi inclusi, comprensivo di motivazione <p>In caso positivo, continuare a valutare le proprietà di trasferibilità; altrimenti valutare approcci alternativi di sintesi delle evidenze (come l'aggiustamento per popolazione) e, se non fattibili, non proseguire con le analisi di sintesi.</p>

2.2.3 Proprietà di consistenza (*consistency*)

È necessario inoltre valutare la proprietà di consistenza (*consistency*). Tale condizione è verificabile solo nel caso di *network* di studi connessi. La consistenza è la concordanza tra le stime degli effetti delle tecnologie sanitarie ottenute tramite comparazioni dirette (cioè, confronti tra tecnologie sanitarie che sono state valutate nello stesso studio) e tramite comparazioni indirette (cioè, confronti derivati combinando studi che confrontano tecnologie sanitarie diverse). La NMA integra entrambe queste fonti di informazioni per stimare gli effetti relativi tra tutti i trattamenti. La scelta dei metodi per valutare la consistenza dipende dalla struttura del *network* degli studi e non tutti i metodi sono adatti per *network* di qualsiasi complessità. I metodi utilizzati per testare la consistenza dovrebbero essere chiaramente identificati, pre-specificati nel protocollo e giustificati fornendo una spiegazione della scelta fatta, facendo riferimento alla struttura del *network* di studi. La presenza di discrepanza

significativa tra le stime dirette e indirette può compromettere la validità delle conclusioni. Può capitare che il *network* di studi contenga un numero insufficiente di studi e dati scarsi per valutare adeguatamente la consistenza. Si pensi ad esempio a *network* di studi non ancorati (Figura 2). Il fatto di non poter rilevare l'inconsistenza non implica che le evidenze siano consistenti. La rilevazione statistica dell'inconsistenza richiede maggiori dati rispetto a quelli necessari per stabilire un effetto del trattamento. L'inconsistenza può essere causata da uno sbilanciamento nella distribuzione dei modificatori dell'effetto tra le prove dirette e indirette, come ad esempio età, gravità di malattia e linea di trattamento, che potrebbero essere confusi fra loro. Pertanto, la valutazione dell'inconsistenza dovrebbe essere effettuata insieme alla valutazione della similarità e dell'eterogeneità del *network* di studi.

Esistono in letteratura vari metodi per testare la consistenza delle stime [17]:

- **Test di incoerenza globale:** valuta la coerenza complessiva del *network* di studi considerando tutte le possibili discrepanze tra evidenze dirette e indirette.
- **Test di incoerenza locale:** valuta la coerenza nei singoli cicli di tecnologie sanitarie nel *network*. Ad esempio, in un ciclo A-B-C-A, si confronta l'effetto A vs C tramite il percorso diretto (se disponibile) con quello indiretto (A-B + B-C).
- **Modelli di inconsistenza specifici:** alcuni modelli statistici, come il modello di inconsistenza di Bucher o il metodo di inconsistenza di Lu e Ades, sono utilizzati per esaminare se esistono differenze significative [18].

In sintesi, per valutare la consistenza delle stime è necessario seguire i seguenti *step* durante tutto il processo di studio (dal protocollo al giudizio finale sulla proprietà in esame)

<p>1 – Fase iniziale</p> <p>Pre-specificato nel protocollo</p>	<p>Descrizione dei metodi che si intendono utilizzare per valutare la consistenza delle stime e giustificazione di questa scelta con riferimento alla struttura del <i>network</i> degli studi.</p> <p>Pre-specificazione dei criteri che si utilizzeranno per determinare se è stata rilevata una violazione significativa della consistenza.</p>
<p>2 – Fase di analisi</p>	<p>Riepilogo dei risultati dei test statistici e/o dei modelli utilizzati per indagare la consistenza, indicando se questi indicano la presenza di inconsistenza e descrivendo l'entità dell'inconsistenza e l'incertezza risultante in questi risultati.</p> <p>Nei casi in cui viene rilevata inconsistenza del modello è necessario descrivere le possibili fonti di inconsistenza in termini di modificatori dell'effetto e, se possibile, le stime dell'entità dell'effetto modificatore.</p>
<p>3 – Fase conclusiva</p>	<p>Giudizio finale sul fatto che il requisito di sufficiente consistenza sia rispettato, con relative motivazioni. Se non è possibile effettuare una valutazione formale della consistenza (ovvero il <i>network</i> di studi non è connesso), questo dovrebbe essere esplicitamente indicato.</p> <p>In caso positivo, continuare a valutare la possibilità di continuare il progetto; altrimenti valutare approcci alternativi di sintesi delle evidenze (come l'aggiustamento per popolazione) e, se non fattibili, non proseguire con le analisi di sintesi.</p>

2.3 Possibili azioni da intraprendere se le assunzioni di trasferibilità sono violate

Se almeno una delle tre proprietà di trasferibilità sopra descritte non risultasse valida (rispetto ai criteri pre-specificati nel protocollo), si dovrebbero prendere in considerazione azioni alternative per rispondere alle domande di ricerca di interesse esplicitate mediante il PICO *framework*, come ad esempio:

1. **Suddivisione in sottogruppi:** Se viene mostrata una dissimilarità per una potenziale variabile modificatore dell'effetto o viene mostrata dell'eterogeneità che può essere spiegata dal modificatore di effetto, potrebbe essere utile dividere gli studi in diversi sottogruppi e trarre conclusioni separate (ad esempio, per uomini e donne). Dovrebbero essere presi in considerazione i limiti delle analisi di sottogruppo basate su dati aggregati [19];
2. **Utilizzo di tecniche di meta-regressione:** I potenziali modificatori di effetto possono essere inclusi come covariate in un modello di meta-regressione (con riferimento specifico alla NMA). Ciò richiede un numero sufficiente di punti dato (cioè di studi) in modo che tutti i parametri possano essere stimati nel modello. Dovrebbero essere presi in considerazione i limiti e le assunzioni della meta-regressione basata su dati aggregati e pre-specificare le possibili strategie da adottare [20, 21];
3. **Esclusione degli studi:** Nel caso in cui solo pochi studi siano la causa della dissimilarità o dell'eterogeneità, si potrebbero fare una serie di analisi di sensitività, includendo ed escludendo tali studi, al fine di avere una chiara panoramica sull'assunzione di similarità;
4. **Analisi di sensitività:** Se non è possibile implementare un'azione correttiva chiaramente utile, dovrebbero comunque essere eseguite una serie di analisi di sensitività che consentano di valutare l'impatto della violazione delle ipotesi. Dato che esistono molte aree di incertezza riguardo ai metodi "corretti" per condurre meta-analisi (come le tecniche di meta-regressione), le analisi di sensitività sono anche un aiuto importante per tutte le decisioni successive nel processo per stimare il loro impatto sui risultati;
5. **Confronti indiretti aggiustati per la popolazione:** Quando c'è una presunta violazione dell'ipotesi di similarità per uno o più variabili modificatori di effetto (a livello di paziente) che rende non corretto metodologicamente procedere con una sintesi delle stime di *effectiveness*, si può pensare di procedere con un approccio diverso applicando metodi che aggiustano per la popolazione. Infatti, qualora si avesse accesso ai dati dei singoli pazienti di almeno una tecnologia sanitaria in esame, è possibile applicare metodi di confronti indiretti aggiustati come ad esempio MAIC, STC, e *multilevel network meta-regression* (ML-NMR). Tuttavia, come verrà discusso successivamente nel documento, questi metodi presentano numerose limitazioni e potrebbero non generare risultati pertinenti alla domanda di ricerca.

L'applicazione delle azioni da intraprendere qualora le assunzioni di trasferibilità fossero violate, sopra descritte, potrebbero portare alla creazione di nuovi *network* di studi. Per questo motivo è fondamentale testare tutte le assunzioni di trasferibilità nei nuovi *network* di studi prima di procedere con la sintesi delle evidenze.

In sintesi, per valutare come procedere in caso di violazione dell'assunzione di trasferibilità delle evidenze è necessario procedere come segue:

<p>1 – Fase iniziale</p> <p>Pre-specificato nel protocollo</p>	<p>Pre-specificare le strategie che saranno attuate in caso di violazione della trasferibilità</p>
<p>2 - Fase di analisi</p>	<p>Riportare i risultati completi della valutazione delle variabili modificatori d'effetto e delle analisi di sensitività. Qualora si ritenesse più sensato tenere separati diversi sottogruppi, al posto di analizzare la popolazione intera, è necessario riportare:</p> <ul style="list-style-type: none"> ○ Una descrizione completa dei sottogruppi; ○ I risultati completi delle analisi sulle ipotesi di similarità e omogeneità di tutti i confronti a coppie all'interno di tutti i sottogruppi (al fine di testare la condizione di trasferibilità); ○ I risultati della sintesi delle evidenze nell'intero gruppo di studio e un confronto di questi con i risultati ottenuti da ciascun sottogruppo.
<p>3 – Fase conclusiva</p>	<p>Giudizio finale sul fatto che le azioni intraprese abbiano portato alla trasferibilità. In particolare, esplicitare se si ha accesso a dati di singoli pazienti di almeno una tecnologia sanitaria per procedere con la sintesi delle evidenze mediante modelli che aggiustano per la popolazione (es., MAIC o STC). Se le azioni intraprese non hanno condotto alla trasferibilità degli studi e/o non si ha accesso ai dati di singoli pazienti per almeno una tecnologia sanitaria di interesse, non proseguire con la sintesi delle evidenze.</p>

2.4 Valutazione dei dati mancanti

Come spesso accade nell'analisi dei dati di studi clinici, una grande quantità di dati mancanti (*missing data*) nei singoli studi potrebbe portare a gravi distorsioni nelle stime dei dati aggregati. Per questo motivo, fin dalla fase di pianificazione delle analisi, è necessario pre-specificare quali metodi saranno presi in considerazione e quali analisi di sensitività saranno condotte per analizzare questo aspetto. In letteratura ci sono vari modi per gestire i dati mancanti. Per una trattazione completa si rimanda al Manuale *Cochrane* [13].

In sintesi, per valutare l'impatto dei dati mancanti sulla stima dell'effetto finale è necessario procedere come segue:

1 – Fase iniziale	Pre-specificare i metodi che si adotteranno in caso di dati mancanti
Pre-specificato nel protocollo	
2- Fase di analisi	Riportare: <ul style="list-style-type: none"> - Le informazioni sulla quantità e sulle ragioni dei dati mancanti (se noto); - I risultati completi di tutte le analisi di sensitività
3 – Fase conclusiva	Giudizio finale sul potenziale impatto dei dati mancanti sull'accuratezza dei risultati complessivi (se ciò può essere determinato). Qualora tale impatto fosse valutato rilevante, valutare approcci alternativi di sintesi delle evidenze e/o di non procedere con le analisi di sintesi.

4. I modelli di confronti indiretti

3.1 Introduzione e focus sui modelli naïve

I confronti indiretti sono una famiglia di modelli che vengono utilizzati quando non ci sono studi H2H che confrontino direttamente le tecnologie sanitarie di interesse. Attraverso vari modelli a seconda del *network* di studi disponibili (ancorato o non ancorato – Figura 2, con accesso ai dati IPD) e dopo aver testato le assunzioni di trasferibilità, è possibile costruire delle nuove evidenze di confronto relativo fra le tecnologie sanitarie di interesse. La forza dell'evidenza di un confronto relativo mediante un confronto indiretto non potrà mai sostituire l'evidenza di un confronto diretto proveniente da uno studio randomizzato. Come già anticipato nel paragrafo precedente, i modelli di ITC possono essere basati su un accesso del dato degli studi di tipo *study-level* (come, ad esempio, le NMA) o *patient-level* (come, ad esempio, le MAIC e STC).

Tuttavia, esiste un ulteriore tipo di ITC: i confronti indiretti *naïve*. Tali confronti sono una forma di confronto semplice tra tecnologie sanitarie che si basano su dati provenienti da studi clinici separati, ma senza l'uso di modelli statistici avanzati per tenere conto delle differenze tra gli studi. In pratica, i confronti *naïve* combinano i dati sugli esiti dei gruppi di trattamento di studi diversi come se provenissero da un unico grande trial, rompendo così la randomizzazione [22, 23]. A titolo esemplificativo si riporta un caso di ITC *naïve*:

- Studio 1 confronta A con C e mostra che A è più efficace di C del 30%
- Studio 2 confronta B con C e mostra che B è più efficace di C del 20%
- Un confronto indiretto *naïve* potrebbe concludere semplicemente che A è più efficace di B del 10% (30% - 20%) senza considerare altri fattori

Tali modelli forniscono una stima semplice e immediata ma spesso inaccurata della differenza tra trattamenti, perché non considerano i fattori che variano tra gli studi. Per questo nelle linee guida sui confronti diretti e indiretti del *Member State Coordination Group on Health Technology Assessment* (HTACG) [24, 25] viene esplicitato che i *naïve* ITC sono da evitare poiché affetti da vari *bias* (come, ad esempio, il *bias* di confondimento) che riducono drasticamente l'affidabilità delle stime, preferendo modelli di ITC più complessi che tentano di bilanciare per le differenze tra studi.

Quando possibile, per stimare l'efficacia relativa dei trattamenti oggetto di interesse sono da preferirsi i modelli di ITC ancorati (cioè con *network* di studi collegati con uno o più comparatori in comune – Figura 2 – Panel A) rispetto ai non ancorati. Qualora la condizione di trasferibilità degli effetti fra studi inclusi nel *network* (descritta nel paragrafo precedente) non fosse verificata è necessario non proseguire con le analisi e attuare delle strategie alternative. Nel caso in cui si avesse accesso agli IPD, si possono integrare le maggiori informazioni sui fattori confondenti nell'analisi, rendendo i confronti più accurati. I metodi basati sull'aggiustamento per le caratteristiche della popolazione (come, ad esempio, MAIC e STC) vengono applicati su una combinazione di AgD e IPD, aggiustando i confronti tra trattamenti per le caratteristiche delle popolazioni arruolate negli studi inclusi. L'obiettivo è quello di ridurre il *bias* e migliorare la validità delle inferenze, assicurando

che i confronti tra i trattamenti, tenendo in considerazione le differenze demografiche e cliniche tra le popolazioni, siano più accurati.

Nel caso in cui i *network* di studi fossero non ancorati (Figura 2 – Panel B) o se i confronti includessero studi non randomizzati (SAT) l'accesso agli IPD è solitamente raccomandato per costruire modelli di confronto indiretto più solidi. I confronti indiretti non ancorati tentano di compensare la mancanza di randomizzazione, ma richiedono l'assunzione di "consistenza condizionale degli effetti assoluti"⁸, che è molto improbabile che venga soddisfatta.

Al contrario, i confronti *naïve* combinano i dati sugli esiti dei gruppi di trattamento di studi diversi come se provenissero da un unico grande trial, rompendo così la randomizzazione [22, 23]. I confronti *naïve* non tengono conto delle differenze tra studi, come le caratteristiche della popolazione, le dosi, o i metodi utilizzati per cui sono considerati meno robusti di altri metodi indiretti non-ancorati come il MAIC, che permettono l'aggiustamento per questi fattori.

Take home message – ITC Naïve

1	I confronti indiretti <i>naïve</i> , seppur facili da implementare, presentano notevoli limitazioni che compromettono la validità dei confronti, rendendo preferibili approcci più sofisticati come le NMA o altre tecniche di analisi statistica avanzata. Tali metodi statistici più avanzati sono decisamente più robusti per affrontare le problematiche derivanti dai dati raccolti da studi diversi, come la differenza nei metodi di raccolta dei dati o la variabilità nelle popolazioni di pazienti.
2	Il livello di evidenza dei confronti indiretti <i>naïve</i> è bassissimo e il loro utilizzo non è raccomandabile per prendere decisioni sulle tecnologie sanitarie oggetto di interesse.
3	<p>L'utilizzo di confronti <i>naïve</i> può portare a conclusioni errate nel confronto indiretto fra tecnologie a causa:</p> <p>A. Assunzione di omogeneità non è giustificabile I modelli <i>naïve</i> assumono implicitamente che gli studi siano simili sotto ogni aspetto, come la popolazione, i criteri di inclusione, le metodologie, ecc. Questa assunzione non è quasi mai vera nella pratica, portando a un rischio di eterogeneità (differenze tra gli studi) che riduce la validità dei risultati.</p> <p>B. Assenza di controllo per i fattori confondenti Senza un adeguato controllo per variabili confondenti, i confronti <i>naïve</i> non sono in grado di isolare l'effetto del trattamento dall'influenza di altri fattori. Questo può portare a stime inaccurate dell'efficacia o della sicurezza dei trattamenti introducendo fonti importanti di <i>Risk of Bias</i>.</p> <p>C. Comparazioni non basate su dati comparabili</p>

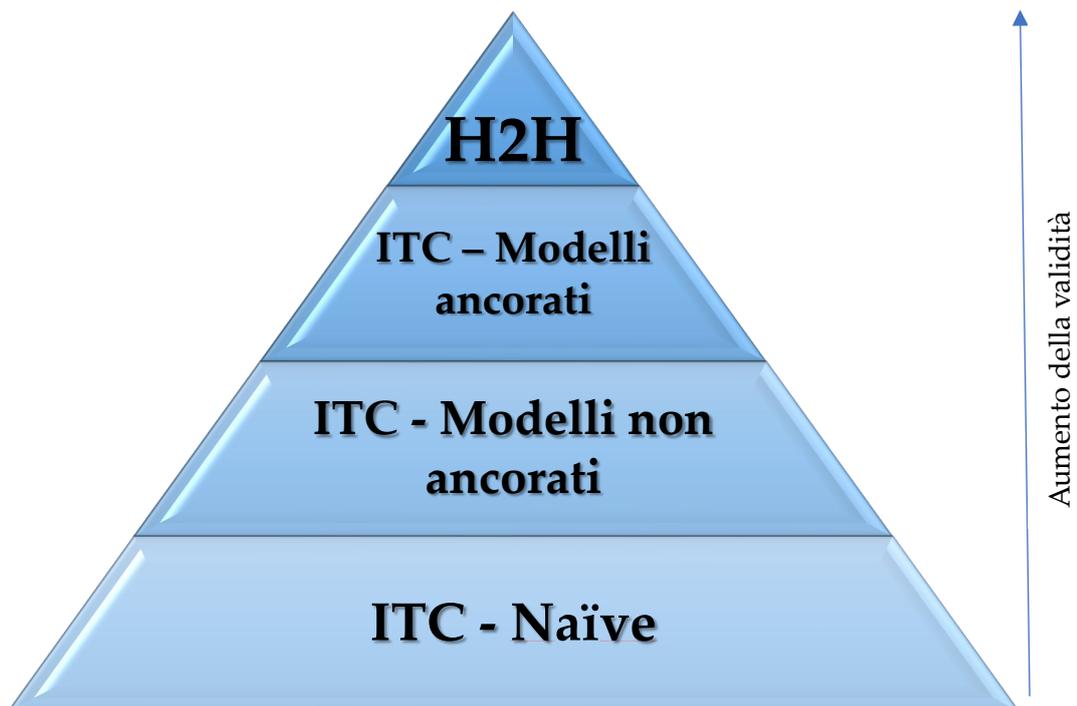
⁸ Questa assunzione si riferisce all'ipotesi che l'effetto di un trattamento rispetto a un altro sia costante attraverso diverse popolazioni e studi, una volta che si tengono sotto controllo le variabili rilevanti.

Nei confronti *naïve*, i dati dei trattamenti confrontati provengono da studi diversi, che potrebbero utilizzare popolazioni, trattamenti, dosaggi o tempi di follow-up differenti.

3.1.1 Il livello di evidenza degli ITC per macro-aggregazione

In conclusione, la validità delle evidenze dei confronti indiretti provenienti da vari metodi disponibili può essere rappresentata nella seguente piramide delle evidenze (Figura 8).

Figura 8 – Piramide delle evidenze dei confronti relativi fra trattamenti

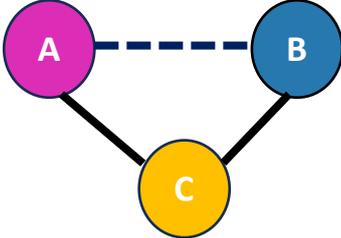
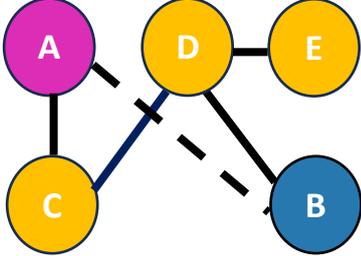


H2H: studi randomizzati di confronto di tecnologie sanitarie *head-to-head*

3.2 Confronti indiretti con *network* di studi ancorati

Qualora fossero dimostrate le assunzioni di trasferibilità, a seconda del tipo di *network* ancorati di studi disponibili è possibile procedere con una sintesi indiretta delle evidenze, come segue

Figura 8 – Tipi di modelli di ITC a seconda del *network* di studi ancorati a disposizione

<p>Panel A – Più facile Network di 3 studi ancorato tramite un comparatore comune</p> 	<p>Metodo di Bucher</p>
<p>Panel B – Più complesso Network di più di 3 studi ancorato tramite una rete di collegamenti</p> 	<p>Network Meta-Analisi (NMA)</p>

3.2.1 Metodo di Bucher

Alla fine degli anni '90, Bucher e colleghi hanno sviluppato il primo metodo per confronti indiretti ancorati nei casi in cui siano disponibili evidenze dirette dell'effetto delle tecnologie sanitarie A e B rispetto al comparatore comune C (Figura 8 – Panel A) [26]. Questo metodo è ampiamente utilizzato in quanto compatibile con numerose misure di effetto, inclusi *odds ratio*, *risk ratio*, differenze di rischio, differenze medie standardizzate e *hazard ratio*.

Non essendo disponibili confronti diretti tra le tecnologie sanitarie A e B, l'assunzione di consistenza non può essere valutata. In questo caso, una valutazione approfondita della similarità è ancora più importante. Nei casi in cui ci siano più studi per un confronto a coppie, questi devono essere combinati per ottenere una stima dell'effetto totale (ad esempio meta-analizzando le stime di *effectiveness* tramite una *pairwise* meta-analisi) prima di applicare il metodo di Bucher. Alcuni *network* di studi più complessi, inclusi i cicli chiusi derivanti da più trial, possono essere analizzati, ma solo

sotto forma di confronti a coppie multipli. Tuttavia, questo metodo assume l'indipendenza tra i confronti a coppie e quindi non può essere facilmente applicato a cicli chiusi derivanti da trial multi-braccio, per i quali questa assunzione non è soddisfatta.

3.2.2 Network Meta-Analisi

Una NMA combina confronti diretti e indiretti per determinare l'efficacia relativa di una tecnologia rispetto a due o più tecnologie di interesse. Per chiarire il linguaggio spesso utilizzato in letteratura, ci si riferisce a confronti misti di trattamenti (*mixed treatment comparisons*, MTC) quando nello stesso *network* di studi si trovano sia confronti diretti che indiretti; quindi, con un insieme di cicli chiusi e aperti.

Un metodo per visualizzare la rete degli studi da includere nella NMA è basato sulla teoria dei grafi (*graph theory*), originariamente sviluppata per l'analisi delle reti elettriche [27]. Questo approccio permette di identificare facilmente il numero di trattamenti inclusi, la connessione tra i vari elementi, se le reti sono disconnesse e quanto è indiretto il confronto tra due nodi. Se due nodi sono più lontani, come ad esempio A e B (Figura 8 – Panel B) rispetto ad A e C, il livello di confidenza sulla loro efficacia relativa sarà più basso. Come già esplicitato nel documento, nel *network* di studi devono essere inserite contemporaneamente tutte le tecnologie identificate nelle domande di ricerca esplicitate tramite il PICO *framework*.

Prima di procedere con l'applicazione della NMA è assolutamente necessario verificare la stessa assunzione di trasferibilità di tutti i confronti indiretti, che richiede sufficiente similarità, omogeneità e consistenza, come discusso nel Capitolo 1. Una differenza statisticamente significativa nelle stime dell'efficacia relativa tra confronti diretti e indiretti indica inconsistenza del modello. Una differenza nella direzione dell'efficacia relativa, anche se non statisticamente significativa, solleverebbe comunque preoccupazioni sulla consistenza. L'assunzione di consistenza può essere valutata solo se sono disponibili confronti diretti e indiretti corrispondenti, cioè se la stessa coppia di trattamenti è valutata sia con un confronto diretto che indiretto (*network* di studi ancorato). Se la consistenza non può essere testata, una valutazione approfondita della similarità è ancora più importante prima di procedere con le analisi statistiche. Le cause dell'inconsistenza in un *network* di studi complesso possono essere difficili da identificare. Infatti, la probabilità di una violazione delle assunzioni di trasferibilità per NMA può aumentare con l'aumentare della complessità del *network* e del numero di tecnologie. Uno dei motivi più comuni dell'inconsistenza è l'eterogeneità clinica tra le popolazioni degli studi (strettamente associata con la proprietà di similarità). Un altro motivo può essere legato al rapporto tra il numero di confronti a coppie e il numero di studi inclusi nell'analisi: se ci sono troppi confronti con pochi studi l'analisi potrebbe essere sottopotenziata, cioè non avere abbastanza potenza statistica per rilevare un effetto [28]. Pertanto, l'inclusione di comparatori e studi aggiuntivi, oltre a quelli necessari per collegare il *network* dovrebbe essere evitata, a meno che non ci siano forti indicazioni che la loro inclusione migliori la robustezza delle evidenze. Un metodo per identificare la sorgente dell'eterogeneità nella NMA è equivalente alla meta-regressione nella *pairwise* meta-analisi [29]. Questo approccio permette di aggiustare per i fattori confondenti o i modificatori

d'effetto che possano influenzare l'esito, come ad esempio le caratteristiche dei pazienti, della malattia o i dettagli della tecnologia sanitaria. La meta-regressione può essere una tecnica utile nel caso in cui l'assunzione di similarità non sia accettata o per comprendere l'effetto delle variabili confondenti. È invece sconsigliata se il numero di studi incluso nella NMA è limitato perché la potenza statistica diminuisce, come accade anche per le *pairwise* meta-analisi o le analisi di regressione lineare.

Storicamente le NMA sono state sviluppate con un approccio frequentista [30] ma con il diffondersi della statistica bayesiana si sono diffuse NMA basate su tale approccio [31-33]. Per la sintesi di evidenze di dati di sopravvivenza mediante NMA è necessario testare l'ipotesi dei rischi proporzionali. Qualora tale assunzione non fosse verificata è possibile procedere con un NMA basata su "*restricted mean survival time*" oppure con "*flexible survival models*". Per approfondimenti metodologici si rimanda alla letteratura scientifica e in particolare alle raccomandazioni del NICE al sito <http://www.nicedsu.org.uk/>.

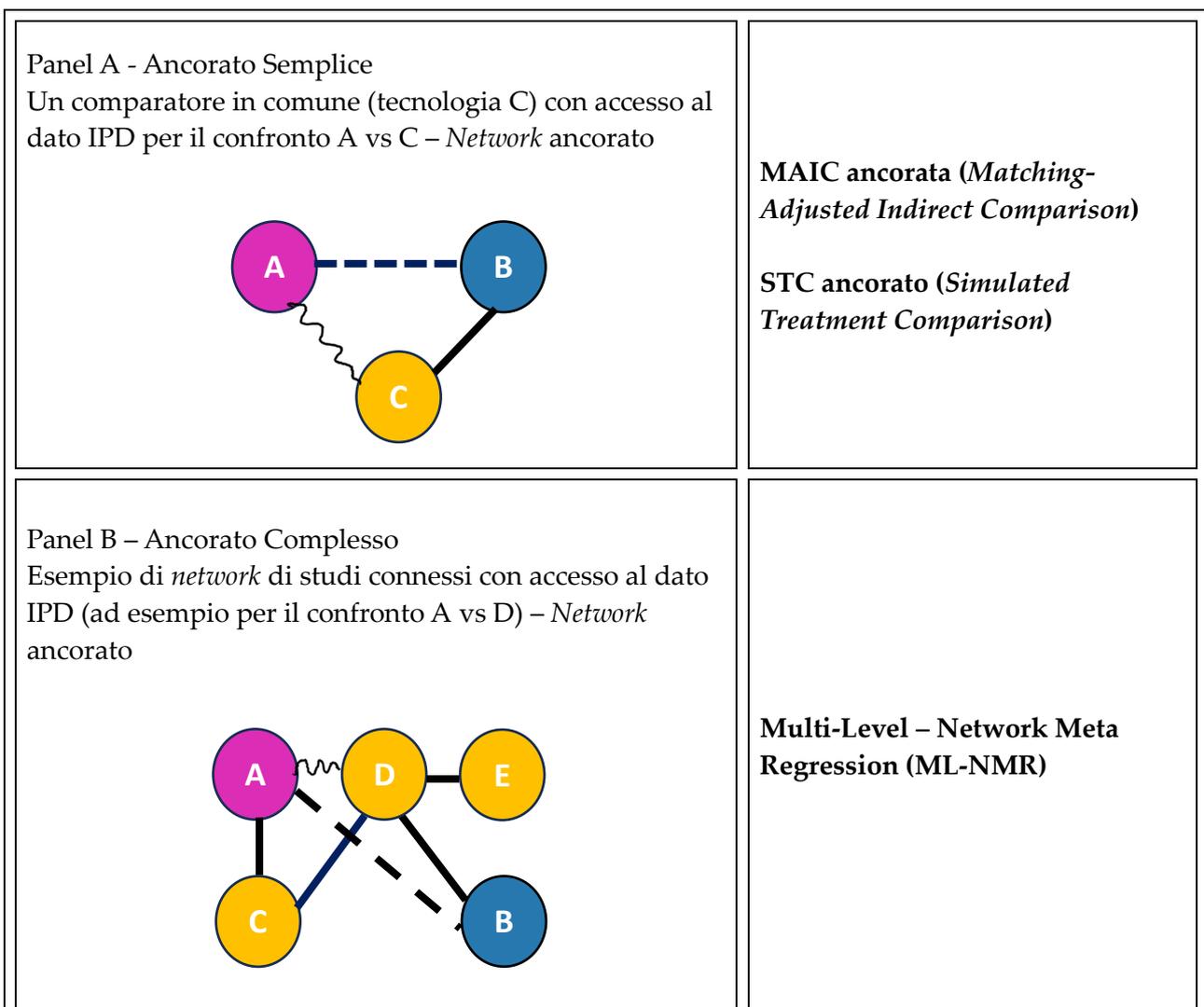
3.2.3 Take home message

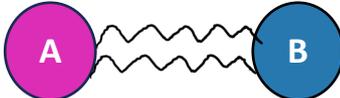
1	I confronti indiretti si basano su più assunzioni rispetto ai confronti diretti provenienti da RCT per stimare un effetto relativo della tecnologia sanitaria di interesse e quindi hanno solitamente risultati più incerti.
2	Quando si effettuano confronti indiretti sono maggiormente appropriati e metodologicamente più solidi i confronti indiretti sviluppati partendo da un <i>network</i> ancorato di studi, poiché preservano la randomizzazione all'interno dello studio.
3	I confronti indiretti ancorati di dati aggregati richiedono l'assunzione di trasferibilità tra studi. Le proprietà di similarità, omogeneità e consistenza dovrebbero essere ben valutate prima di iniziare le analisi (<i>protocol-driven</i>). Se una qualsiasi di queste proprietà non risultasse valida, è opportuno non procedere con la sintesi delle evidenze poiché tali risultati sarebbero inaccurati.
4	Gli approcci utili per i confronti indiretti di <i>network</i> ancorati di studi includono il metodo di Bucher e i modelli NMA frequentisti e bayesiani.

3.3 Confronti indiretti basati su metodi che aggiustano per le caratteristiche della popolazione

Come precedentemente trattato, qualora non ci fosse un *network* ancorato fra studi o qualora ci fosse un *network* ancorato ma non fossero dimostrate le assunzioni di trasferibilità dei risultati, non è indicato procedere con una sintesi indiretta delle evidenze tramite una NMA. Se fossero disponibili IPD di almeno uno studio incluso nel *network* di studi è possibile utilizzare dei metodi che aggiustano per le caratteristiche della popolazione. In Figura 10 si riportano i *network* di studi possibili con la relativa tipologia di analisi di confronti indiretti più appropriata.

Figura 10 – Tipologia di modelli di ITC a seconda del *network* di studi a disposizione, avendo a disposizione i dati *patient-level* (IPD) di almeno uno studio



<p>Panel C - Senza un comparatore in comune (non ancorato) Due studi SAT a confronto con accesso ai dati IPD di almeno uno studio (ad esempio A)</p> 	<p>MAIC non ancorata</p> <p>STC non ancorata</p>
<p>Panel D – Studi separati con accesso ai dati IPD di tutti gli studi</p> 	<p>Pooled analysis –</p> <p>Caso più frequente: External control Arm (ECA)</p>

La validità dei metodi di aggiustamento della popolazione dipende dall'inclusione di tutti i modificatori di effetto rilevanti come covariate nel modello. La selezione delle covariate deve essere basata su una revisione comprensiva della letteratura e sulla consulenza di professionisti sanitari. È importante che:

- Le covariate inizialmente bilanciate non siano omesse per evitare la creazione di nuovi squilibri
- Le variabili prognostiche che non sono anche modificatori di effetto siano incluse per migliorare l'accuratezza
- La selezione delle covariate sia prespecificata e non *data-driven*

In caso non fossero inclusi dei modificatori di effetto rilevanti, le stime aggregate che si otterranno saranno inevitabilmente distorte e non interpretabili.

Nel caso si avesse accesso agli IPD di tutti gli studi a supporto delle tecnologie sanitarie oggetto di indagine, si può procedere a confrontare le tecnologie utilizzando un unico database contenente i dati di ogni singolo paziente incluso negli studi utilizzando la tecnica di *pooled analysis* (Figura 10 – Panel D). Avendo a disposizione tutti i dati IPD degli studi, rispetto ad altri modelli, questo disegno di studio ha dei vantaggi importanti, quali ad esempio un maggiore controllo sui fattori di confondimento, contenere l'eterogeneità delle analisi (ad esempio standardizzando le analisi e i criteri di inclusione/esclusione). Una delle applicazioni più frequenti di questo tipo di modello di confronto indiretto è l'utilizzo dell'*External Control Arm (ECA)*. Gli ECA sono gruppi di confronto in uno studio clinico che non provengono da un assegnamento randomizzato all'interno dello stesso studio, ma da fonti esterne. Le fonti esterne dei dati possono essere di diversa natura come, ad esempio, le cartelle cliniche dei pazienti trattati nella pratica clinica, i registri di patologia o bracci di altri studi clinici già condotti. Tali coorti di dati possono essere retrospettive oppure prospettiche (da

preferire, se disponibili, per l'accuratezza generalmente superiore delle informazioni raccolte rispetto a quelle nelle coorti retrospettive). Questo disegno di studio è particolarmente utile quando è necessario quantificare per via indiretta il valore terapeutico di una nuova tecnologia rispetto agli attuali standard di cura quando le evidenze a supporto derivano da un SAT. Questa condizione è sempre più frequente, ad esempio, perché non è etico assegnare pazienti a un placebo o a un trattamento meno efficace, oppure perché il reclutamento di pazienti è difficile a causa della rarità della patologia. Sebbene questi modelli di *pooled analysis/ECA* siano generalmente preferibili rispetto ad altri ITC, l'accesso completo agli IPD delle evidenze dirette è poco frequente e spesso non fattibile. Per un maggiore approfondimento metodologico di questi disegni di studio si rimanda alle più recenti pubblicazioni in letteratura [34-36].

3.3.1 Matching-adjusted Indirect Comparison (MAIC)

Il metodo di confronto indiretto MAIC ancorato pesa i dati IPD di uno studio (es A vs C) in modo che le caratteristiche medie risultino simili a quelle dell'altro studio (es B vs C) (Figura 10 - Panel A) [37, 38]. I pesi sono stimati generalmente mediante un modello di regressione logistica in maniera analoga al metodo del *propensity score* utilizzato per gli studi osservazionali. L'effetto della tecnologia sanitaria A rispetto a B si ottiene dai dati IPD dello studio 1 (i.e., A vs C) ripesati applicando il *propensity score weighting* usando le distribuzioni medie delle variabili al basale dello studio 2 (i.e., B vs C). In questo modo il confronto indiretto della *relative effectiveness* della tecnologia sanitaria A rispetto a quella B sarà meno distorto.

Lo stesso metodo MAIC può essere applicato nei casi in cui le tecnologie sanitarie A e B siano state valutate in trial a braccio singolo (SAT) (Figura 10 – Panel C). È importante notare che il metodo MAIC, ancorato o non ancorato, è valido solo se esiste una sufficiente sovrapposizione tra le popolazioni di pazienti in entrambi gli studi, cioè se la dimensione effettiva del campione pesato non è troppo inferiore a quella originale non pesata (il cosiddetto *effective sample size* definito come la grandezza del campione che effettivamente contribuisce alla stima dell'effetto).

3.3.2 Simulated Treatment Comparison (STC)

Il metodo STC è un metodo alternativo alla MAIC per costruire evidenze indirette fra tecnologie sanitarie di interesse. Come per la MAIC, il modello STC è utilizzabile quando si dispone di IPD per una tecnologia sanitaria e dei soli AgD per il trattamento di confronto (ad esempio medie, deviazioni standard e proporzioni per le caratteristiche dei pazienti, derivate dalle pubblicazioni scientifiche). Si costruisce un modello statistico (regressione lineare, logistica o altro) basato sui dati IPD (A vs C) che descrive come le caratteristiche dei pazienti influenzano l'*outcome*. Questo modello viene poi utilizzato per predire l'effetto relativo medio di A vs C in una popolazione che simula (*simulated treatment comparison*), in termini di caratteristiche dei pazienti, quella AgD (C vs B), e dunque poi di ottenere un effetto relativo medio di interesse (A vs B) aggiustato per le popolazioni. Questo approccio è particolarmente utile per scenari complessi dove si sospetta una relazione non lineare

tra le covariate e l'*outcome*. I limiti principali sono che si assume che tutte le differenze tra le popolazioni possano essere modellizzate e che non tiene conto di eventuali fattori confondenti non misurati. Come per la MAIC, anche il modello STC è applicabile a *network* non ancorati di studi (Figura 10 – Panel C).

3.3.3 Confronto fra STC e MAIC e altri modelli

Le differenze principali tra STC e MAIC sono qui riassunte:

Caratteristica	STC	MAIC
Obiettivo principale	Simulare una popolazione <i>target</i> per confrontare i trattamenti.	Bilanciare le differenze di popolazione tra studi utilizzando un metodo di <i>weighting</i> .
Approccio	Utilizza un modello per rappresentare la popolazione <i>target</i> dello studio comparativo.	Applica un peso statistico ai pazienti dello studio con IPD.
Dati utilizzati	Utilizza dati individuali (IPD) per generare una popolazione simulata, con caratteristiche comparabili alla popolazione per cui sono disponibili solo dati aggregati (AgD).	Utilizza IPD per calcolare pesi che bilanciano le caratteristiche basali rispetto ai AgD dell'altro studio.
Metodo statistico	Regressione: stima i parametri di un modello utilizzando IPD e li applica alle medie delle caratteristiche della popolazione AgD.	<i>Propensity score weighting</i> : assegna un peso ai pazienti IPD in modo che la popolazione pesata risulti simile alla popolazione AgD.
Output principale	Stima dei parametri del modello (es. rischio relativo, <i>odds ratio</i>) per la popolazione <i>target</i> simulata.	Un confronto "aggiustato" tra i trattamenti, bilanciando le differenze di popolazione tra gli studi.
Flessibilità	Adatto per modelli complessi anche non lineari.	Più semplice da applicare per bilanciare differenze nelle caratteristiche basali.
Uso pratico	Utile quando è necessaria una modellizzazione dettagliata delle relazioni tra variabili.	Adatto per situazioni in cui le caratteristiche basali possono essere bilanciate con pesi.

In sintesi, i vantaggi e limiti di ciascun metodo sono qui riportati:

Metodo	Vantaggi	Limiti
STC	Flessibile per modelli complessi Consente simulazioni specifiche per la popolazione target	Richiede forti assunzioni sul modello Dipende dalla qualità del modello scelto
MAIC	Relativamente semplice da implementare Non richiede modellizzazione delle relazioni tra covariate e <i>outcome</i>	Funziona bene solo con covariate bilanciabili Non cattura la complessità delle relazioni come le interazioni non lineari

In generale, la scelta tra STC e MAIC dipende dalla complessità del problema, dalla disponibilità dei dati e dagli obiettivi dell'analisi. In ogni caso, la scelta del modello deve essere pre-specificata e non *data-driven*.

A titolo informativo si riporta che recentemente è stato proposto un metodo alternativo alle MAIC e STC che permette di ovviare alle limitazioni note di questi approcci e produrre stime meno distorte dell'effetto relativo di interesse, noto come il metodo *Predictive-adjusted indirect comparison* [39]. Tale metodo, data l'elevata complessità, ad oggi non sembra essere stato adottato in modo diffuso, ma è utile monitorare questo filone di ricerca per capire la potenziale applicazione di nuovi metodi che dimostrino una minor distorsione nelle stime.

3.3.4 Multilevel Network meta-regression (ML-NMR)

L'approccio della ML-NMR per i confronti indiretti aggiustati per la popolazione è stato proposto per estendere il metodo della NMA nei contesti in cui non valga l'assunzione di trasferibilità [40, 41]. L'approccio ML-NMR permette di includere sia i dati IPD sia i dati AgD in un *network* complesso di studi connessi (Figura 10 – Panel B). I metodi ancorati di MAIC e STC possono essere visti come suoi casi particolari di ML-NMR in cui si confrontano solo due tecnologie sanitarie (Figura 10 – Panel A). La ML-NMR è una tecnica avanzata che combina la NMA con la meta-regressione multilivello, consentendo di modellare in modo esplicito le variazioni tra gli studi e le interazioni tra i trattamenti e le caratteristiche della popolazione. L'accesso ai dati IPD consente di correggere le differenze di popolazione tra gli studi, introducendo covariate (es. età, genere, comorbidità) e di modellare l'eterogeneità tra studi (es. differenze nei protocolli, nei metodi o nelle popolazioni) attraverso un approccio gerarchico/multilivello. In questo modo, gli studi sono modellati come "livelli" in una struttura gerarchica, che consente di separare le variazioni dovute agli studi dalle variazioni intrinseche dei trattamenti; successivamente viene stimata la relazione tra le covariate e l'efficacia dei trattamenti. La ML-NMR permette di stimare tutti gli effetti relativi aggiustandoli per una qualunque delle popolazioni incluse nel network inserendo nel modello di regressione specifico per

il confronto di interesse le caratteristiche medie della popolazione target oggetto di studio [42]. Recentemente i metodi ML-NMR sono stati estesi con successo anche all'analisi dei dati time-to-event ampliando gli ambiti clinici di applicazione di tale metodologia [43].

3.3.5 Take home message

1	Quando la distribuzione dei modificatori d'effetto non è bilanciata tra gli studi selezionati, per eseguire un confronto indiretto non è consigliato applicare i metodi di confronto indiretto classici (Bucher o NMA) poiché la violazione dell'assunzione di trasferibilità potrebbe portare ad una stima distorta degli effetti relativi.
2	I metodi basati sull'aggiustamento per le caratteristiche della popolazione richiedono che siano disponibili i dati IPD per almeno una delle tecnologie sanitarie in analisi. A seconda del tipo di <i>network</i> di evidenze disponibili (semplice o complessa, ancorata o non) e del numero di studi a cui si ha accesso ai dati IPD si possono applicare modelli diversi (Figura 10).
3	Le validità delle stime ottenute con questi metodi aggiustati per le caratteristiche della popolazione si basa sul presupposto che siano state incluse nel modello tutti i modificatori d'effetto rilevanti per i trattamenti in analisi; la scelta di tali variabili deve essere pre-specificata in un protocollo di studio e giustificata dal punto di vista clinico e non statistico. Inoltre, dichiarare la semplice mancanza del dato come giustificazione per l'esclusione di un modificatore di effetto non è sufficiente.

3.4 Confronti indiretti basati su evidenze non randomizzate

3.4.1 Considerazioni generali

Nel caso di confronti indiretti tra due trattamenti osservati in studi differenti senza un comparatore comune (ossia, *network* di studi non ancorati – Figura 2 Panel B), l'utilizzo dei dati è simile a quella di uno studio non randomizzato. In altre parole, i dati invece di essere osservati in uno studio osservazionale comparativo senza randomizzazione provengono da studi diversi. Quando le evidenze non randomizzate sono disponibili solo a livello di AgD, non esistono metodi che garantiscano di ottenere una stima affidabile di *relative effectiveness* della nuova tecnologia. I confronti *naïve* tra i bracci di trattamento in tali situazioni sono soggetti a *bias* a causa dei fattori confondenti e non dovrebbero essere eseguiti. Quando le evidenze non randomizzate sono disponibili solo a livello di AgD, non esistono metodi che garantiscono di ottenere una stima affidabile dell'efficacia del trattamento. Se sono disponibili IPD da SAT e statistiche aggregate da un'altra fonte, l'uso di STC e MAIC è stato proposto e applicato per correggere il *bias* da confondimento (Figura 10 Panel C). Tuttavia, questi metodi, quando non esiste un comparatore comune, si basano su un'assunzione molto forte di "consistenza condizionale degli effetti assoluti". Tale assunzione implica che l'esito nei bracci di trattamento sia costante a ogni livello delle variabili prognostiche e dei modificatori di effetto [44]. Tuttavia, in quasi tutte le applicazioni pratiche, questa assunzione è non giustificabile. Pertanto, l'uso di STC e MAIC senza un comparatore comune è fortemente problematico. Quando gli effetti del trattamento vengono stimati da *network* di studi di evidenze non ancorate, dovrebbero essere preferibilmente utilizzati metodi per l'analisi di dati non randomizzati con accesso a dati IPD completi per tutti gli studi (*Pooled Analysis/ECA* - Figura 10 Panel D). Nel caso in cui fossero disponibili IPD per tutte le covariate rilevanti (fattori confondenti, variabili prognostiche o modificatori dell'effetto), è possibile eseguire analisi con aggiustamento per tali fattori in maniera molto efficace. Per quanto riguarda i metodi di aggiustamento per la popolazione (discussi nel paragrafo precedente), è importante che le covariate rilevanti incluse siano clinicamente giustificate e predefinite in un piano di analisi statistica prima dell'analisi dei dati [45]. Sono disponibili diversi approcci per l'aggiustamento dei fattori confondenti utilizzando IPD, come la regressione multipla, la *g-computation* e i *propensity scores* – PS [45, 46]. Nel contesto della stima dell'efficacia relativa dei trattamenti, sono comunemente utilizzati i metodi basati sul PS, incluso [47] il *matching* e il *weighting* attraverso il peso inverso. Nella sezione successiva saranno forniti maggiori dettagli sull'applicazione dei metodi basati sulla tecnica del PS.

Analogamente ai metodi di aggiustamento della popolazione con un comparatore comune basati su IPD, i metodi per i confronti indiretti basati su IPD richiedono che non vi siano confondenti non misurati. In altre parole, tutti i confondenti rilevanti e i modificatori dell'effetto devono essere stati raccolti e inclusi nel modello. Anche in questo caso, è importante che i confondenti e i modificatori dell'effetto rilevanti siano clinicamente giustificati e predefiniti in un piano di analisi statistica prima dell'analisi dei dati. Tuttavia, rimane sempre la possibilità che un confondente o un modificatore dell'effetto rilevante non sia incluso o non sia stato identificato. Pertanto, raccomandazioni chiare e inferenza valida sugli effetti del trattamento basate su confronti indiretti con aggiustamento per

confondenti su IPD sono possibili solo se l'entità dell'effetto stimato del trattamento è così grande da non poter essere spiegato dal *bias* dovuto alla mancanza del controllo per alcuni confondenti o modificatori dell'effetto. Ciò può essere formalmente valutato testando ipotesi nulle modificate (*shifted null hypothesis*). Questo significa che una conclusione su un effetto del trattamento può essere tratta solo se l'intervallo di confidenza si trova al di sopra o al di sotto di una certa soglia diversa da quella standard (es. 1 nel caso di *risk ratio*). Questo approccio tiene conto dell'incertezza legata alla possibile non considerazione di alcuni confondenti o modificatori dell'effetto rilevanti [48].

Esistono anche altri metodi di misurazione dell'incertezza (*sensitivity analysis*) per esplorare l'impatto potenziale dei confondenti non misurati [49]. Un esempio è l'uso del valore E (*E-value*), definito come la forza minima di associazione, tramite una scala di rischio, che un confondente non misurato dovrebbe avere (sia con il trattamento che senza) per spiegare una specifica associazione al trattamento-esito. Maggiore è il valore E, maggiore sarebbe il confondimento residuo necessario per spiegare l'effetto del trattamento [50].

3.4.2 Propensity Score

Il PS viene utilizzato principalmente quando si dispone di IPD da studi non randomizzati, ovvero in presenza di un *network* di evidenze non connesso (Figura 2 – Panel B; Figura 10 – Panel C e D). Attraverso una procedura di appaiamento (*matching* o di *weighting*), il PS consente di comparare i soggetti trattati con le varie tecnologie sanitarie mimando la randomizzazione. Un punteggio di propensione rappresenta la probabilità condizionata di assegnazione a un trattamento specifico, data una serie di variabili osservabili. Si possono utilizzare una vasta gamma di modelli statistici, dalla regressione logistica ai modelli di machine learning. Le stime dei parametri del modello vengono poi utilizzate per calcolare il punteggio di propensione per ciascun paziente.

Quando si usa il PS devono essere soddisfatte alcune assunzioni dette di scambiabilità condizionata [51]. Queste assunzioni possono essere valutate esaminando le seguenti proprietà:

- **Positività:** i pazienti di entrambi i gruppi devono essere teoricamente idonei a ricevere entrambi i trattamenti di interesse.
- **Sovrapposizione:** deve esserci una sufficiente sovrapposizione nei dati disponibili, misurata attraverso il punteggio di propensione, tra le popolazioni che ricevono i trattamenti di interesse.
- **Bilanciamento:** le popolazioni nei gruppi a confronto devono risultare sufficientemente bilanciate dopo il *matching* o il *weighting*.

Il grado di sovrapposizione e bilanciamento tra i gruppi dipende in larga misura dal modello scelto. Se mediante il PS non si ottiene un grado sufficiente di sovrapposizione o bilanciamento non è possibile effettuare un adeguato aggiustamento per un set appropriato di covariate e, di conseguenza, non si possono ottenere confronti robusti sui trattamenti [51]. In questo caso, optare per la regressione multipla non rappresenta una soluzione poiché richiederebbe solo di attuare

estrapolazioni inappropriate in aree prive di dati osservati [51]. Per questo motivo è necessario interrompere le analisi e non proseguire con le stime delle evidenze poiché il risultato ottenuto non sarebbe accurato.

Il grado di sovrapposizione e bilanciamento può essere influenzato anche dal *trimming*, cioè dalla procedura di esclusione dei soggetti con punteggi (di propensione) al di sotto di una determinata soglia [52]. Se il *trimming* permette di ottenere una sufficiente sovrapposizione e bilanciamento, la popolazione finale di pazienti diventa la popolazione target alla quale si riferiscono gli effetti stimati. Pertanto, se si utilizza l'approccio del PS, la popolazione target finale deve essere descritta in dettaglio, specialmente se vengono applicati metodi di *trimming*.

3.4.3 Take home message

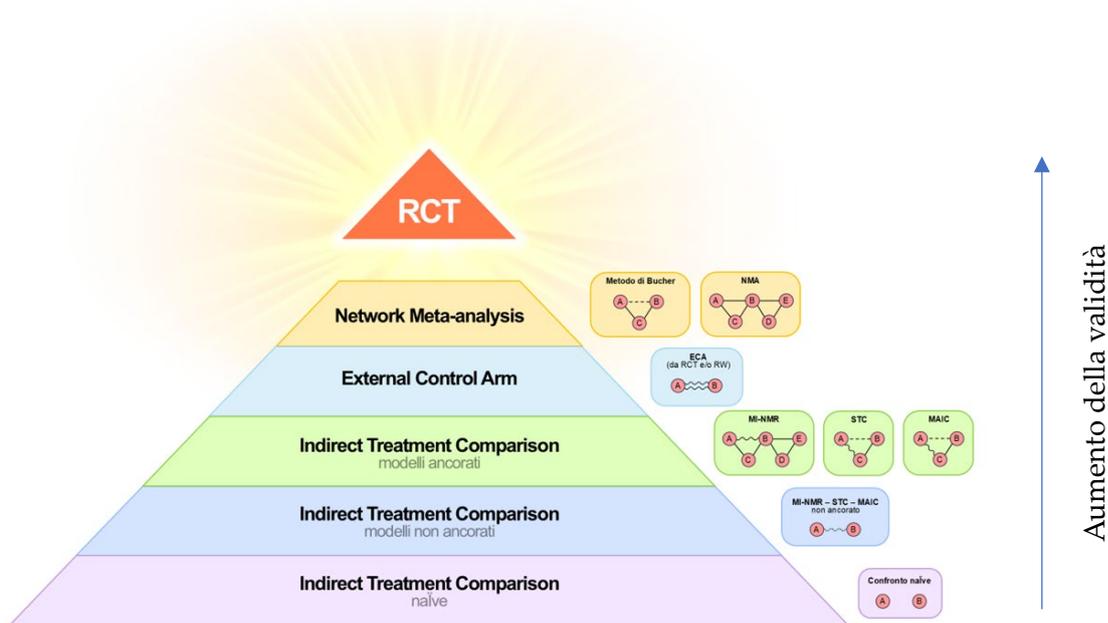
1	Tutte le scelte metodologiche sulle tecniche di analisi dei dati e dei modelli (es. PS) devono essere pre-specificate nel protocollo di ricerca prima di iniziare le analisi
2	I metodi di aggiustamento per i confondenti si basano sull'accesso completo ai IPD
3	I metodi di aggiustamento della popolazione, come MAIC e STC, quando applicati a <i>network</i> di studi non ancorati richiedono assunzioni più forti rispetto ai metodi basati su IPD completi (<i>pooled analysis/ECA</i>) e generalmente non sono sufficienti per correggere i confondenti
4	Le strategie per la selezione del modello e delle covariate per l'aggiustamento dei confondenti devono essere predefinite e basate su criteri trasparenti
5	Le applicazioni del PS richiede che siano soddisfatte alcune assunzioni dette di scambiabilità condizionata. Se tali assunzioni non sono soddisfatte, un adeguato aggiustamento per i confondenti non è possibile, e i risultati dell'analisi corrispondente non forniranno una stima accurata dell'effetto del trattamento
6	Se viene applicato l'approccio del PS, la popolazione target finale deve essere descritta in dettaglio, soprattutto se vengono utilizzati metodi di <i>trimming</i> . Infatti, la popolazione target a cui le stime risultanti dai metodi STC/MAIC/ECA fanno riferimento può non coincidere con la popolazione di interesse per cui si sta svolgendo l'analisi. Quindi è necessario discutere e interpretare a posteriori la plausibilità e la validità dei risultati di <i>relative effectiveness</i> ottenuti.
7	Gli effetti del trattamento stimati da dati non randomizzati sono associati a un'incertezza aggiuntiva derivante da varie fonti causata dalla maggior variabilità degli studi inclusi (che spesso non sono stati condotti in contemporanea o in contesti geografici di cura diversi). A causa di questa maggiore incertezza, è necessario un effetto stimato più ampio per fare inferenza. In particolare, è consigliabile testare ipotesi statistiche più restrittive valutando la superiorità di una nuova tecnologia rispetto ad un'altra solamente se le differenze superano un valore soglia (beneficio clinico) pre-specificato. Inoltre, le analisi di <i>sensitivity</i> (pre-specificate) sono una buona strategia per valutare la robustezza delle stime di <i>relative effectiveness</i> , al fine di quantificare l'incertezza.

4.5 Conclusioni

3.5.1 Piramide delle evidenze indirette

In accordo con le linee guida sui confronti indiretti del Regolamento europeo di HTA, la solidità delle evidenze dei vari modelli di confronti indiretti fra tecnologie sanitarie è sintetizzata in Figura 11. I modelli con solidità dell'evidenza più bassa sono i confronti naïve mentre quelli con solidità più elevata sono le NMA. Si rammenta che in generale la solidità delle evidenze indirette sono sempre molto più basse rispetto alle evidenze dirette provenienti da RCT.

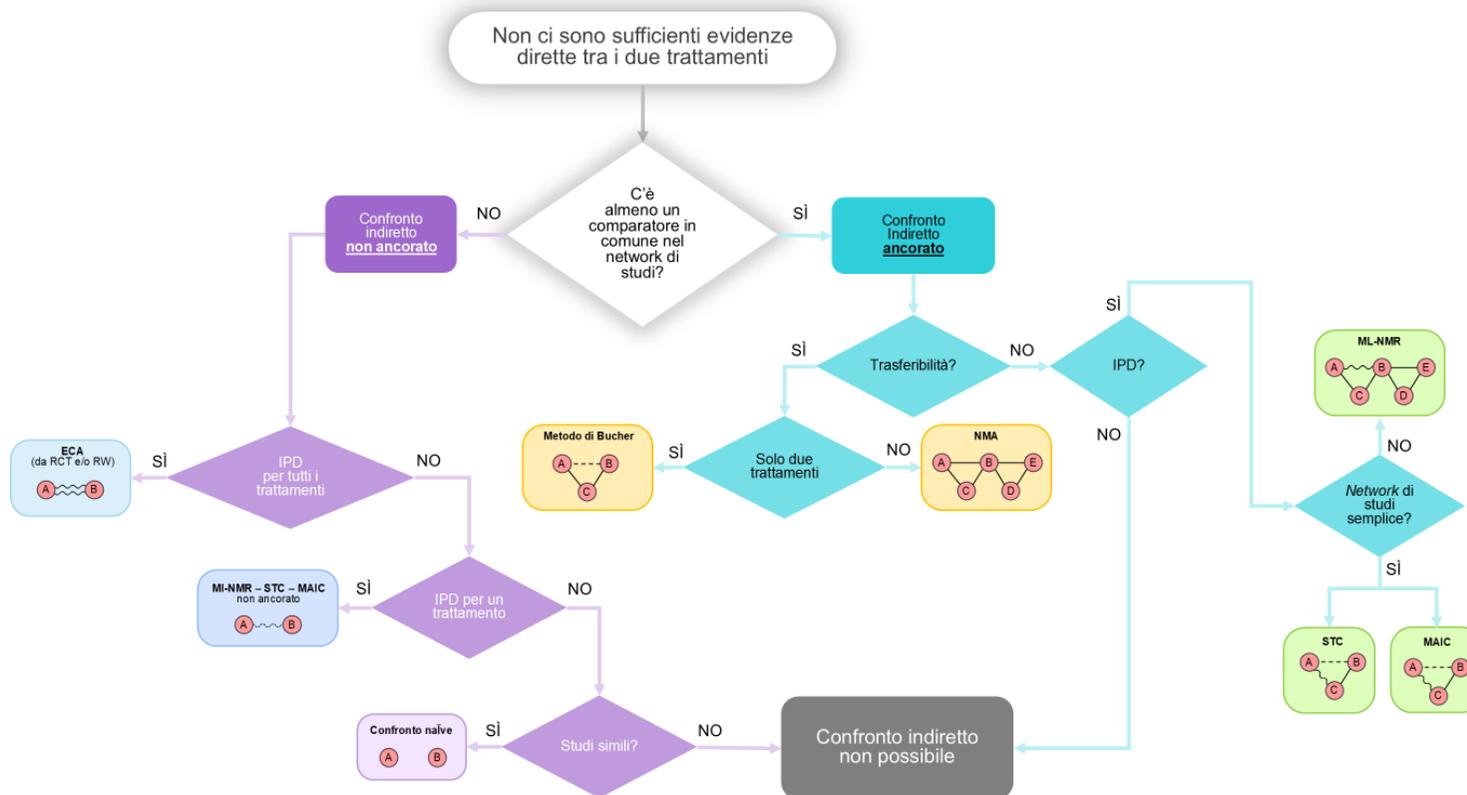
Figura 11 – Piramide delle evidenze indirette di confronto fra tecnologie sanitarie



3.5.2 Scelta del modello di confronto indiretto più robusto in funzione all'accessibilità del dato

Viene qui presentato un diagramma di flusso che guida nella scelta del modello di ITC più adeguato, basandosi sulla metodologia più solida e appropriata in relazione alla complessità del *network* di evidenze disponibili (in relazione al numero di trattamenti e/o alla possibilità di un ancoraggio) e alla disponibilità dei dati (AgD e/o IPD) (Figura 12).

Figura 12 – Diagramma di flusso per la scelta dell’approccio metodologico più appropriato e solido a partire dai dati in possesso



3.5.3 Check list delle fasi di costruzione o validazione di una evidenza indiretta

A conclusione di tutti gli argomenti trattati nel documento, si propone una check list da seguire per la conduzione e/o valutazione di un'evidenza indiretta a supporto di una tecnologia sanitaria di interesse.

Step	Descrizione	Riferimento nel documento
1	Definizione chiara e precisa della domanda di ricerca mediante il PICO <i>framework</i>	Figura 5
2	Ricerca sistematica della letteratura con selezione trasparente degli studi potenzialmente utili, seguendo un protocollo di ricerca	Paragrafo 2.1
3	Valutazione della condizione di trasferibilità (<i>exchangeability</i>) mediante un'attenta analisi delle seguenti proprietà: <ul style="list-style-type: none"> - Similarità (<i>similarity</i>) - Omogeneità (<i>homegeneity</i>) - Consistenza (<i>consistency</i>) 	Paragrafo 2.2
4	Disegno del <i>network</i> di evidenze disponibili	Figura 2, Figura 3
5	Scelta del modello di ITC più appropriato sulla base dei dati a disposizione	Figura 12
6	Interpretazione delle stime di <i>relative effectiveness</i> delle tecnologie sanitarie di interesse alla luce della solidità del modello di ITC utilizzato	Figura 11

GLOSSARIO

Confondente: una caratteristica che influenza sia la probabilità di ricevere la tecnologia sanitaria sia l'esito. Un confondente è quindi necessariamente una variabile prognostica (ma non è vero il contrario) e può essere un modificatore di effetto. Un esempio potrebbe essere lo stadio di una malattia: è una variabile prognostica perché influenza l'*outcome*, ma può anche influenzare la scelta del trattamento per il paziente.

Confronto diretto: confronto di tecnologie sanitarie tramite uno studio comparativo singolo o una *pairwise* meta-analisi o altro metodo per la sintesi di studi comparativi senza confronti indiretti.

Confronto indiretto: termine generico che si riferisce a qualsiasi sintesi di evidenze in cui vengono confrontate tecnologie sanitarie utilizzando i dati da studi diversi. Ciò include la sintesi delle evidenze in cui viene effettuata un'inferenza sull'efficacia relativa di due trattamenti senza l'uso di studi che confrontano i trattamenti direttamente; vengono anche effettuati confronti indiretti quando vengono applicati metodi più generali di *network meta-analysis*, anche se sono disponibili studi diretti per il confronto di interesse.

Metodo per confronti indiretti che aggiustano per la popolazione: metodo per confronti indiretti in cui vengono utilizzati dati individuali di pazienti (*individual patient data* - IPD) provenienti da uno o più studi e dati aggregati provenienti da altri studi per tenere in considerazione le differenze nelle caratteristiche rilevanti delle popolazioni a confronto al fine di stimare un effetto del trattamento aggiustato.

Modificatori dell'effetto: caratteristiche che influenzano l'efficacia relativa tra due tecnologie sanitarie. In termini statistici, le covariate che modificano l'effetto sono quelle variabili per cui esiste un effetto di interazione tra il trattamento, le caratteristiche a livello di studio o di paziente, e l'esito. È possibile che una particolare caratteristica rappresenti sia una variabile prognostica che un modificatore di effetti, sebbene in generale non tutte le variabili prognostiche saranno modificatori di effetto. Un esempio di caratteristica che può essere sia una variabile prognostica che un modificatore di effetto potrebbe essere lo stadio di malattia: un paziente in una fase precoce può avere una prognosi migliore rispetto a un paziente in una fase successiva, indipendentemente dal trattamento che riceve.

Network meta analysis (NMA): generalizzazione della meta-analisi alle reti di evidenze che includono più di due trattamenti, che possono includere sia evidenze dirette che indirette. La NMA incorpora altri termini utilizzati in letteratura per descrivere la sintesi di evidenze dirette e indirette, come comparazioni di trattamenti misti e comparazioni indirette di trattamenti.

Pairwise meta-analisi: la sintesi di due o più studi comparativi provenienti da evidenze dirette con un trattamento comune e un comparatore, per produrre una stima combinata dell'effetto relativo della tecnologia sanitaria di interesse.

Relative Effectiveness: è la misura in cui una tecnologia sanitaria di interesse migliora gli esiti di salute (efficacia e sicurezza) rispetto alle tecnologie sanitarie alternative identificate.

Risk of Bias: Il rischio di *bias* si riferisce alla probabilità che errori sistematici (dette distorsioni, da non confondersi con gli errori casuali) influenzino i risultati o le interpretazioni di uno studio. Può derivare da problemi nella progettazione, conduzione, analisi o interpretazione dello studio. Un elevato rischio di *bias* può compromettere la validità dei risultati di uno studio, rendendo difficile trarre conclusioni affidabili.

Single Arm Trial (SAT): è uno studio clinico in cui tutti i pazienti ricevono lo stesso trattamento, senza un gruppo di controllo randomizzato. In certi ambiti terapeutici (es malattie rare) i SAT sono spesso fase 1 e/o 2 dello sviluppo della tecnologia sanitaria condotti per fini regolatori.

Trasferibilità: implica che, se pazienti di un gruppo di trattamento fossero sostituiti in un altro, l'effetto del trattamento rimarrebbe lo stesso; comprende i componenti di similarità, omogeneità e, nel caso di confronti indiretti, consistenza.

Tecnologia sanitaria: le tecnologie sanitarie comprendono prodotti medicinali, dispositivi medici, dispositivi medici per diagnosi in vitro e procedure mediche, nonché misure per la prevenzione, diagnosi o trattamento delle malattie.

Variabili prognostiche: sono caratteristiche (es. caratteristiche del paziente) che influenzano l'esito di interesse indipendentemente dal trattamento ricevuto.

RIFERIMENTI BIBLIOGRAFICI

- [1] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, "Introduction to Meta-Analysis", ed: John Wiley & Sons, 2009, Print ISBN:9780470057247, Online ISBN:9780470743386, DOI:10.1002/9780470743386.
- [2] B. Macabeo, T. Rotrou, A. Millier, C. François, and P. Laramée, "The Acceptance of Indirect Treatment Comparison Methods in Oncology by Health Technology Assessment Agencies in England, France, Germany, Italy, and Spain," *Pharmacoecon Open*, vol. 8, no. 1, pp. 5-18, Jan 2024, doi: 10.1007/s41669-023-00455-6.
- [3] V. Sala, L. Moja, I. Moschetti, S. Bidoli, V. Pistotti, and A. Liberati, "Revisioni sistematiche - Breve guida all'uso", Centro Cochrane Italiano, 2006.
- [4] PROSPERO International prospective register of systematic reviews, <https://www.crd.york.ac.uk/prospero/>. Accessed 9 November 2024.
- [5] EQUATOR Enhancing the QUALity and Transparency Of health Research, <https://www.equator-network.org/about-us/>. Accessed 9 November 2024.
- [6] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", *Bmj*, vol. 372, p. n71, Mar 29 2021, doi: 10.1136/bmj.n71.
- [7] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandembroucke, "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies", *Ann Intern Med*, vol. 147, no. 8, pp. 573-7, Oct 16 2007, doi: 10.7326/0003-4819-147-8-200710160-00010.
- [8] Committee for Medicinal Products for Human Use (CHMP). ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (Step 5) (EMA/CHMP/ICH/436221/2017) [online]. 2020. https://www.ema.europa.eu/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf.2020. Accessed 9 November 2024.
- [9] D. J. Fisher, J. R. Carpenter, T. P. Morris, S. C. Freeman, and J. F. Tierney, "Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach?", *Bmj*, vol. 356, p. j573, Mar 3 2017, doi: 10.1136/bmj.j573.
- [10] S. C. Freeman and J. R. Carpenter, "Bayesian one-step IPD network meta-analysis of time-to-event data using Royston-Parmar models", *Res Synth Methods*, vol. 8, no. 4, pp. 451-464, Dec 2017, doi: 10.1002/jrsm.1253.
- [11] S. C. Freeman, N. J. Cooper, A. J. Sutton, M. J. Crowther, J. R. Carpenter, and N. Hawkins, "Challenges of modelling approaches for network meta-analysis of time-to-event outcomes in the presence of non-proportional hazards to aid decision making: Application to a melanoma network", *Stat Methods Med Res*, vol. 31, no. 5, pp. 839-861, May 2022, doi: 10.1177/09622802211070253.

- [12] S. Greenland, "Tests for interaction in epidemiologic studies: a review and a study of power", *Stat Med*, vol. 2, no. 2, pp. 243-51, Apr-Jun 1983, doi: 10.1002/sim.4780020219.
- [13] Deeks JJ, Higgins JPT, Altman DG, McKenzie JE, Veroniki AA (editors). Chapter 10: Chapter 10: Analysing data and undertaking meta-analyses [last updated November 2024]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook. Accessed 9 November 2024.
- [14] C. P. Cordero and A. L. Dans, "Key concepts in clinical epidemiology: detecting and dealing with heterogeneity in meta-analyses", *J Clin Epidemiol*, vol. 130, pp. 149-151, Feb 2021, doi: 10.1016/j.jclinepi.2020.09.045.
- [15] Sutton AJ, Abrams KR, Jones DR et al., "Methods for Meta-Analysis in Medical Research." Chichester, UK: Wiley, 2000, ISBN: 978-0-471-49066-1.
- [16] J. P. Higgins and S. G. Thompson, "Quantifying heterogeneity in a meta-analysis", *Stat Med*, vol. 21, no. 11, pp. 1539-58, Jun 15 2002, doi: 10.1002/sim.1186.
- [17] S. Dias, N. J. Welton, A. J. Sutton, D. M. Caldwell, G. Lu, and A. E. Ades, "NICE Decision Support Unit Technical Support Documents" in *NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials*. London: National Institute for Health and Care Excellence (NICE), Copyright © 2014 National Institute for Health and Clinical Excellence, unless otherwise stated. All rights reserved., 2014.
- [18] S. Dias, A. E. Ades, N. J. Welton, J. P. Jansen, and A. J. Sutton, "Network Meta-Analysis for Decision Making" ed: John Wiley & Sons Ltd., 2018.
- [19] S. Dias, A. J. Sutton, N. J. Welton, and A. E. Ades, "Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment", *Med Decis Making*, vol. 33, no. 5, pp. 618-40, Jul 2013, doi: 10.1177/0272989x13485157.
- [20] J. A. Berlin, J. Santanna, C. H. Schmid, L. A. Szczech, and H. I. Feldman, "Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head", *Stat Med*, vol. 21, no. 3, pp. 371-87, Feb 15 2002, doi: 10.1002/sim.1023.
- [21] J. P. T. Higgins and S. G. Thompson, "Controlling the risk of spurious findings from meta-regression" *Statistics in Medicine*, vol. 23, no. 11, pp. 1663-1682, 2004, doi: <https://doi.org/10.1002/sim.1752>.
- [22] S. J. Edwards, M. J. Clarke, S. Wordsworth, and J. Borrill, "Indirect comparisons of treatments based on systematic reviews of randomised controlled trials" *International Journal of Clinical Practice*, vol. 63, no. 6, pp. 841-854, 2009, doi: <https://doi.org/10.1111/j.1742-1241.2009.02072.x>.
- [23] G. Gartlehner and C. G. Moore, "Direct versus indirect comparisons: a summary of the evidence," *Int J Technol Assess Health Care*, vol. 24, no. 2, pp. 170-7, Spring 2008, doi: 10.1017/s0266462308080240.
- [24] HTA CG, "Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons", <https://health.ec.europa.eu/document/download/1f6b8a70-5ce0-404e-9066->

[120dc9a8df75_en?filename=hta_practical-guideline_direct-and-indirect-comparisons_en.pdf](#). Accessed 9 November 2024.

- [25] HTA CG, "Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons", https://health.ec.europa.eu/document/download/4ec8288e-6d15-49c5-a490-d8ad7748578f_en?filename=hta_methodological-guideline_direct-indirect-comparisons_en.pdf. Accessed 9 November 2024.
- [26] H. C. Bucher, G. H. Guyatt, L. E. Griffith, and S. D. Walter, "The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials", *J Clin Epidemiol*, vol. 50, no. 6, pp. 683-91, Jun 1997, doi: 10.1016/s0895-4356(97)00049-8.
- [27] G. Rücker, "Network meta-analysis, electrical networks and graph theory", *Res Synth Methods*, vol. 3, no. 4, pp. 312-24, Dec 2012, doi: 10.1002/jrsm.1058.
- [28] A. A. Veroniki, D. Mavridis, J. P. Higgins, and G. Salanti, "Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study", *BMC Med Res Methodol*, vol. 14, p. 106, Sep 19 2014, doi: 10.1186/1471-2288-14-106.
- [29] I. R. White, "Network Meta-analysis" *The Stata Journal*, vol. 15, no. 4, pp. 951-985, 2015, doi: 10.1177/1536867x1501500403.
- [30] T. Lumley, "Network meta-analysis for indirect treatment comparisons", *Stat Med*, vol. 21, no. 16, pp. 2313-24, Aug 30 2002, doi: 10.1002/sim.1201.
- [31] G. Lu and A. E. Ades, "Combination of direct and indirect evidence in mixed treatment comparisons", *Stat Med*, vol. 23, no. 20, pp. 3105-24, Oct 30 2004, doi: 10.1002/sim.1875.
- [32] G. Salanti, J. P. Higgins, A. E. Ades, and J. P. Ioannidis, "Evaluation of networks of randomized trials", *Stat Methods Med Res*, vol. 17, no. 3, pp. 279-301, Jun 2008, doi: 10.1177/0962280207080643.
- [33] A. Sutton, A. E. Ades, N. Cooper, and K. Abrams, "Use of indirect and mixed treatment comparisons for technology assessment", *Pharmacoeconomics*, vol. 26, no. 9, pp. 753-67, 2008, doi: 10.2165/00019053-200826090-00006.
- [34] J. Lambert, E. Lengliné, R. Porcher, R. Thiébaud, S. Zohar, and S. Chevret, "Enriching single-arm clinical trials with external controls: possibilities and pitfalls", *Blood Adv*, vol. 7, no. 19, pp. 5680-5690, Oct 10 2023, doi: 10.1182/bloodadvances.2022009167.
- [35] P. S. Mishra-Kalyani *et al.*, "External control arms in oncology: current use and future directions", *Ann Oncol*, vol. 33, no. 4, pp. 376-383, Apr 2022, doi: 10.1016/j.annonc.2021.12.015.
- [36] M. Cucherat *et al.*, "From single-arm studies to externally controlled studies. Methodological considerations and guidelines", *Therapie*, vol. 75, no. 1, pp. 21-27, Jan-Feb 2020, doi: 10.1016/j.therap.2019.11.007.
- [37] J. E. Signorovitch *et al.*, "Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research", *Value Health*, vol. 15, no. 6, pp. 940-7, Sep-Oct 2012, doi: 10.1016/j.jval.2012.05.004.

- [38] K. J. Ishak, I. Proskorovsky, and A. Benedict, "Simulation and Matching-Based Approaches for Indirect Comparison of Treatments" *PharmacoEconomics*, vol. 33, no. 6, pp. 537-549, 2015/06/01 2015, doi: 10.1007/s40273-015-0271-1.
- [39] A. Remiro-Azócar, A. Heath, and G. Baio, "Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data", *Res Synth Methods*, vol. 13, no. 6, pp. 716-744, Nov 2022, doi: 10.1002/jrsm.1565.
- [40] D. M. Phillippo *et al.*, "Multilevel Network Meta-Regression for Population-Adjusted Treatment Comparisons", *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 183, no. 3, pp. 1189-1210, 2020, doi: 10.1111/rssa.12579.
- [41] D. M. Phillippo *et al.*, "Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal." *Med Decis Making*. 2018 Feb;38(2):200-211. doi: 10.1177/0272989X17725740.
- [42] D. M. Phillippo, S. Dias, A. E. Ades, and N. J. Welton, "Target estimands for efficient decision making: Response to comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study", " *Statistics in Medicine*, vol. 40, no. 11, pp. 2759-2763, 2021, doi: <https://doi.org/10.1002/sim.8965>.
- [43] D. Maciel *et al.*, "Implementing Multilevel Network Meta-Regression for Time-To-Event Outcomes: A Case Study in Relapsed Refractory Multiple Myeloma" *Value in Health*, vol. 27, no. 8, pp. 1012-1020, 2024, doi: 10.1016/j.jval.2024.04.017.
- [44] D. M. Phillippo, A. E. Ades, S. Dias, S. Palmer, K. R. Abrams, and N. J. Welton, "Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal" *Medical Decision Making*, vol. 38, no. 2, pp. 200-211, 2018, doi: 10.1177/0272989x17725740.
- [45] M. A. Hernán and J. M. Robins, "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available", *Am J Epidemiol*, vol. 183, no. 8, pp. 758-64, Apr 15 2016, doi: 10.1093/aje/kwv254.
- [46] T. Agoritsas, A. Merglen, N. D. Shah, M. O'Donnell, and G. H. Guyatt, "Adjusted Analyses in Studies Addressing Therapy and Harm: Users' Guides to the Medical Literature", *Jama*, vol. 317, no. 7, pp. 748-759, Feb 21 2017, doi: 10.1001/jama.2016.20029.
- [47] M. Webster-Clark *et al.*, "Using propensity scores to estimate effects of treatment initiation decisions: State of the science", *Stat Med*, vol. 40, no. 7, pp. 1718-1735, Mar 30 2021, doi: 10.1002/sim.8866.
- [48] IQWiG, Concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a Social Code Book V (SGB V), Version 1.0 [online]. 2020; https://www.iqwig.de/download/a19-43_routine-practice-data-for-the-benefit-assessment-of-drugs_rapid-report_v1-0.pdf. Accessed 9 November 2024.
- [49] X. Zhang, D. E. Faries, H. Li, J. D. Stamey, and G. W. Imbens, "Addressing unmeasured confounding in comparative observational research", *Pharmacoepidemiol Drug Saf*, vol. 27, no. 4, pp. 373-382, Apr 2018, doi: 10.1002/pds.4394.

- [50] T. J. VanderWeele and P. Ding, "Sensitivity Analysis in Observational Research: Introducing the E-Value", *Ann Intern Med*, vol. 167, no. 4, pp. 268-274, Aug 15 2017, doi: 10.7326/m16-2607.
- [51] E. Williamson, R. Morley, A. Lucas, and J. Carpenter, "Propensity scores: from naive enthusiasm to intuitive understanding", *Stat Methods Med Res*, vol. 21, no. 3, pp. 273-93, Jun 2012, doi: 10.1177/0962280210394483.
- [52] R. J. Glynn, M. Lunt, K. J. Rothman, C. Poole, S. Schneeweiss, and T. Stürmer, "Comparison of alternative approaches to trim subjects in the tails of the propensity score distribution" *Pharmacoepidemiology and Drug Safety*, vol. 28, no. 10, pp. 1290-1298, 2019, doi: <https://doi.org/10.1002/pds.4846>.