

## AMSTAR 2: strumento per la valutazione critica di revisioni sistematiche di trial randomizzati e/o di studi non randomizzati sull'efficacia degli interventi sanitari

Beverley J Shea<sup>1,2,3\*</sup>, Barnaby C Reeves<sup>4</sup>, George Wells<sup>3,5</sup>, Micere Thuku<sup>1,2</sup>, Candyce Hamel<sup>1</sup>, Julian Moran<sup>6</sup>, David Moher<sup>1,3</sup>, Peter Tugwell<sup>1,2,3,7</sup>, Vivian Welch<sup>2,3</sup>, Elizabeth Kristjansson<sup>8</sup>, David A Henry<sup>9,10,11</sup>

<sup>1</sup>Ottawa Hospital Research Institute, Clinical Epidemiology Program, Ottawa, Canada, <sup>2</sup>Bruyère Research Institute, Ottawa, Canada, <sup>3</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada, <sup>4</sup>School of Clinical Sciences, University of Bristol, Bristol, UK, <sup>5</sup>University of Ottawa Heart Institute, Ottawa, Canada, <sup>6</sup>The Hospital for Sick Children, the Genetics and Genome Biology Program, Toronto, Canada, <sup>7</sup>Department of Medicine, The Ottawa Hospital, Ottawa, Canada, <sup>8</sup>Centre for Research in Educational and Community, Services, School of Psychology, Faculty of Social Sciences, University of Ottawa, Canada, <sup>9</sup>Centre for Research in Evidence-Based Practice, Bond University, Gold Coast, Australia, <sup>10</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, Canada, <sup>11</sup>Institute for Clinical Evaluative Sciences, Toronto, Canada

### ABSTRACT

Le revisioni sistematiche di studi che valutano l'efficacia degli interventi sanitari sono sempre più numerose ed ampiamente utilizzate per informare decisioni cliniche e politiche sanitarie. Tali revisioni, oltre ad essere influenzate da numerosi bias, sempre più frequentemente includono studi non randomizzati. Considerata l'importanza di identificare le revisioni sistematiche di elevata qualità sono stati proposti numerosi strumenti, ma solo pochi permettono un'adeguata valutazione critica. AMSTAR è stato sviluppato per valutare le revisioni sistematiche di trial controllati randomizzati. Il presente articolo presenta l'aggiornamento e l'adattamento di AMSTAR, che permette una valutazione critica più analitica delle revisioni sistematiche che includono sia trial controllati randomizzati, sia studi non randomizzati per valutare l'efficacia degli interventi sanitari. Considerato che un numero cre-

sciente di decisioni vengono prese sulla base di evidenze di natura osservazionale dal *real world*, AMSTAR 2 può aiutare i decisori ad identificare revisioni sistematiche di elevata qualità, incluse quelle basate su studi non randomizzati.

**Citazione.** Shea BJ, Reeves BC, Wells G et al. AMSTAR 2: strumento per la valutazione critica di revisioni sistematiche di trial randomizzati e/o di studi non randomizzati sull'efficacia degli interventi sanitari. Evidence 2020;12(2): e1000206.

**Publicato** 18 febbraio 2020

**Copyright.** © 2017 Shea et al. Questo è un articolo open-access, distribuito con licenza *Creative Commons Attribution*, che ne consente l'utilizzo, la distribuzione e la riproduzione su qualsiasi supporto esclusivamente per fini non commerciali, a condizione di riportare sempre autore e citazione originale.

**Fonti di finanziamento.** Il presente lavoro è stato sostenuto da un grant del Canadian Institutes for Health Research (MOP-130470). BCR è stato in parte sostenuto dalla UK National Institute for Health Research Bristol Cardiovascular Biomedical Research Unit.

**Conflitti d'interesse.** Tutti gli autori hanno compilato il ICMJE uniform disclosure a [http://www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) e dichiarano: nessun sostegno istituzionale per il lavoro presentato; nessuna relazione finanziaria con organizzazioni che potrebbero avere un interesse per il lavoro presentato nei tre anni precedenti, nessun'altra relazione o attività che potrebbe influenzare il lavoro presentato.

**Provenienza.** Tradotto da: Shea Beverley J, Reeves Barnaby C, Wells George et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017; 358:j4008.

\* E-mail: bevshea@uottawa.ca

Con il rapido incremento del numero di pubblicazioni biomediche, rimanere aggiornati utilizzando gli studi primari è pressoché impossibile sia per i clinici che per i policy maker<sup>1</sup>. Di conseguenza, chi deve prendere decisioni in sanità utilizza principalmente le revisioni sistematiche (RS) per garantire un'assistenza sanitaria basata sulle evidenze scientifiche<sup>2</sup>. Le RS consentono di basare le decisioni cliniche su sintesi accurate, affidabili e complete delle migliori evidenze su specifici quesiti<sup>2</sup>, ma accettarne i risultati in maniera acritica è rischioso. Tuttavia, anche se sono disponibili linee guida per il reporting delle RS sia di trial controllati randomizzati – *randomized controlled trial* (RCT), come il PRISMA statement<sup>3</sup> (Preferred Reporting Items for Systematic Review and Meta-Analyses), sia di studi osservazionali, come il MOOSE (*Meta-analysis of Observational Studies in Epidemiology*)<sup>4</sup>, la qualità del reporting di una RS può riflettere più la capacità degli autori di scrivere in modo comprensibile che non la metodologia di conduzione della RS stessa, enfatizzando la necessità di strumenti per valutare le modalità di pianificazione e conduzione delle RS<sup>5,6</sup>.

L'*Handbook* della Cochrane Collaboration fornisce una guida completa per gli autori di RS, ma non prevede uno strumento per la valutazione critica di quelle già completate<sup>5</sup>. Sono stati proposti numerosi strumenti per valutare criticamente gli studi inclusi nelle RS e per la metodologia di conduzione di step specifici (es. meta-analisi, test per il bias di pubblicazione)<sup>7-15</sup>, ma solo pochi strumenti valutano tutti gli step rilevanti nella conduzione di una RS<sup>16-21</sup>. AMSTAR (*A Measurement Tool to Assess systematic Reviews*) è uno degli strumenti più utilizzati<sup>22-24</sup>: pubblicato nel 2007, nasce per consentire a professionisti sanitari e policy maker, non necessariamente esperti di epidemiologia, di effettuare una valutazione critica rapida e riproducibile della qualità metodologica delle RS di RCT sull'efficacia degli interventi sanitari. Tuttavia, sin dalla sua pubblicazione, nello strumento sono state rilevate numerose criticità<sup>25-31</sup> che, insieme ai feedback ricevuti durante i workshop e all'evoluzione della scienza delle RS, ne hanno reso necessaria la revisione e l'aggiornamento.

### Inclusione di studi non randomizzati nelle revisioni sistematiche

Circa la metà delle RS oggi include studi non randomizzati sull'efficacia degli interventi sanitari – *non-randomized studies of intervention effects* (NRSI)<sup>4,32-34</sup> che comprendono vari disegni di studio che valutano l'efficacia degli interventi sanitari senza assegnare in maniera randomizzata i partecipanti. Tuttavia, esistono numerose problematiche nella conduzione e reporting di RS di NRSI<sup>32,35,36</sup>, poiché questi studi sono esposti a numerosi bias, assenti o minimizzati nei RCT, che rendono necessaria una diversa valutazione del rischio di bias (*risk of bias* – RoB). In

### Key points

- Le revisioni sistematiche di trial che valutano l'efficacia degli interventi sanitari spesso includono studi non randomizzati.
- AMSTAR è uno strumento sviluppato per la valutazione critica di revisioni sistematiche di trial controllati e randomizzati.
- AMSTAR è stato ulteriormente elaborato per consentire la valutazione critica di revisioni sistematiche che includono sia trial randomizzati che studi non randomizzati per valutare l'efficacia degli interventi sanitari.
- La nuova versione dello strumento (AMSTAR 2) include 16 item (di cui 10 degli 11 dello strumento originale), prevede categorie di risposta più semplici, fornisce una guida più completa per i valutatori e prevede una valutazione complessiva basata su eventuali punti di debolezza su specifici item "critici".
- AMSTAR 2 non è finalizzato a generare un punteggio complessivo.
- Considerato che un numero crescente di decisioni vengono prese sulla base di evidenze di natura osservazionale dal *real world*, AMSTAR 2 dovrebbe contribuire ad identificare le revisioni sistematiche di elevata qualità.

particolare, gli studi osservazionali (una tipologia di NRSI) vengono sempre più spesso condotti all'interno di grandi database di popolazione, a volte con centinaia di migliaia o addirittura milioni di soggetti ai quali vengono erogati interventi sanitari, generando stime precise sull'efficacia degli interventi sanitari, ma che potrebbero essere inaccurate per i bias presenti. Se queste stime vengono combinate con quelle provenienti da RCT (generalmente condotti su un numero minore di partecipanti), le meta-stime saranno maggiormente influenzate dalle stime cumulative degli studi osservazionali. Lo strumento AMSTAR originale non prevede nessuna analisi del RoB dei NRSI inclusi in una RS, un problema rilevante considerata la potenziale variabilità dei disegni di tali studi e dei bias che potrebbero influenzarne la validità interna.

### Sviluppo di AMSTAR 2

Lo sviluppo e la validazione dello strumento AMSTAR originale pubblicato nel 2007 è stato già dettagliatamente descritto<sup>22-24</sup>. In sintesi, l'elenco degli item era stato creato sulla base dei risultati di una *scoping review* nell'ambito degli strumenti di valutazione disponibili. Questa revisione aveva identificato numerosi item sovrapposti, principalmente da due report ampiamenti citati<sup>16,17</sup>. L'elenco degli item estratto da questi report è stato combinato e ridotto attraverso un'analisi fattoriale. Dopo un test pilota, gli item sono stati riformulati secondo necessità ed è stato effettuato un test di affidabilità ed usabilità dello strumento; quindi la versione modificata è stata

validata con successo da un panel esterno di esperti<sup>23</sup>. Le pubblicazioni che descrivono lo strumento AMSTAR originale sono ampiamente citate e lo strumento stesso è stato molto utilizzato e criticato<sup>22-31</sup>.

È stato convocato un panel, che include gli autori dello strumento originale, esperti nella conduzione di NRSI, nello sviluppo di strumenti di valutazione critica, in biostatistica e nei disegni di studio. Il panel si è riunito per un giorno ad Ottawa, in Canada, e ai membri sono stati presentati i risultati di revisioni aggiornate della letteratura sugli strumenti di valutazione rilevanti, quelli di una survey condotta sugli utenti di AMSTAR, le esperienze dei partecipanti ai workshop AMSTAR, i feedback dal sito web ([www.amstar.ca](http://www.amstar.ca)) e le critiche pubblicate allo strumento originale<sup>16-26</sup>.

L'obiettivo del panel era di aumentare il valore di AMSTAR come strumento completo di valutazione critica sviluppato principalmente per RS di studi sull'efficacia degli interventi sanitari. Il panel ha ritenuto che le revisioni dello strumento dovevano riguardare tutti gli aspetti della conduzione di una RS con l'obiettivo di includere anche i NRSI. Inoltre, il gruppo ha suggerito che la nuova revisione dello strumento dovrebbe essere utilizzata per l'insegnamento e come checklist di controllo per gli autori di RS. La revisione dello strumento non mirava ad affrontare i requisiti speciali di revisioni di test diagnostici, meta-analisi di dati di pazienti individuali o network meta-analisi, *scoping reviews* e *realist reviews*<sup>37-41</sup>.

È stato utilizzato il metodo del gruppo nominale per proporre e definire le priorità per modificare lo strumento e concordare la bozza degli item. Sulla base della loro esperienza sullo strumento e delle presentazioni condivise durante la riunione, è stato chiesto ai partecipanti di registrare in maniera indipendente e riservata le proprie idee, per poi renderle note secondo un formato a *round-robin*. Inizialmente, un facilitatore ha presentato al gruppo tutte le idee raccolte da ciascuno, a turno; quindi i membri hanno registrato individualmente i loro giudizi e voti che sono stati aggregati statisticamente per ottenere un giudizio di gruppo. Sono state così concordate le modifiche, di seguito riportate senza alcun ordine di priorità in quanto ritenute tutte rilevanti:

- Semplificare le categorie di risposta.
- Allineare la definizione dei quesiti di ricerca al framework PICO: popolazione, intervento, controllo, outcome.
- Ricercare la motivazione che ha spinto gli autori ad includere nelle RS differenti disegni di studio (RCT e NRSI).
- Ricercare maggiori dettagli sulle motivazioni per l'esclusione di studi dalla RS.
- Determinare se gli autori hanno effettuato una valutazione sufficientemente dettagliata del RoB negli studi inclusi, sia RCT che NRSI.
- Determinare se il RoB negli studi inclusi è stato adeguatamente considerato nella sintesi statistica dei risultati, se eseguita.

- Determinare se il RoB negli studi inclusi è stato adeguatamente considerato nell'interpretazione e nella discussione dei risultati.

Per ogni bozza degli item è stata formulata una descrizione e un sottogruppo ristretto ha perfezionato la formulazione degli item, assemblando lo strumento per il test pilota, che inizialmente è stato effettuato dai membri del gruppo. Le bozze sono state presentate ai workshop in occasione dei Cochrane Colloquium del 2015 e 2016 dove, grazie ai numerosi feedback, sono state apportate ulteriori modifiche ed è stata elaborata una nuova bozza. La versione qui presentata è stata sottoposta a test di affidabilità tra esaminatori e test di usabilità.

### Confronto con lo strumento originale

La checklist allegata fornisce i dettagli del nuovo strumento (AMSTAR 2). Dello strumento originale sono stati mantenuti 10 item, con modifiche al testo basate sui feedback relativi allo strumento originale e sui test del nuovo strumento.

La selezione in doppio degli studi e l'estrazione dei dati, che nello strumento originale facevano parte dello stesso item, in AMSTAR 2 costituiscono due item separati. La possibile influenza delle fonti di finanziamento viene considerata separatamente per i singoli studi inclusi nella revisione e per la revisione stessa, mentre in precedenza erano combinate nello stesso item. Sono state inoltre aggiunte considerazioni più dettagliate e separate sul RoB per RCT e per NRSI. Entrambi i sotto-item sono basati sul contenuto degli strumenti per il RoB di RCT e NRSI della Cochrane (ROBINS-I)<sup>42,43</sup>. È stato eliminato l'item sulla ricerca della letteratura grigia, in precedenza autonomo e adesso incluso in quello sulla ricerca bibliografica. Complessivamente sono stati aggiunti 4 item. Due di questi - elaborazione del PICO, modalità con cui il RoB è stato gestito nella sintesi delle evidenze - derivano direttamente dallo strumento ROBINS-I<sup>43</sup>; la discussione delle possibili cause e il significato di eterogeneità è un'elaborazione di quanto contenuto nello strumento AMSTAR originale; la motivazione della selezione di determinati disegni di studio è parte dell'adattamento di AMSTAR finalizzato ad includere NRSI.

Le domande specifiche per ogni item in AMSTAR 2 sono progettate in modo che la risposta "Sì" indichi un risultato positivo. Rispetto allo strumento AMSTAR originale sono state eliminate le opzioni "non applicabile" e "non so" perché tutti gli item sono stati ritenuti rilevanti nelle attuali RS sull'efficacia degli interventi sanitari. In caso di assenza di informazioni che rende impossibile esprimere il giudizio su un item, agli autori non è concesso il beneficio del dubbio e la risposta deve essere "NO". Ove ritenuto utile è stata lasciata la possibilità di rispondere "Sì, in parte", per segnalare l'aderenza parziale allo standard di quell'item.

### Razionale per la scelta degli item

Viene di seguito riportato il razionale degli item di AMSTAR 2, numerati come nello strumento (cfr. checklist). Per una guida più dettagliata si rimanda all'appendice 1.

#### 1. I quesiti di ricerca e i criteri di inclusione della revisione comprendono gli elementi del PICO?

Il PICO (popolazione, intervento, controllo, outcome) viene frequentemente utilizzato come framework pratico e facilmente memorizzabile per i quesiti di ricerca. Talora è necessario prevedere anche un arco temporale per essere certi che lo studio rilevi outcome clinici rilevanti (es. quando l'effetto dell'intervento è prevedibile solo dopo diversi anni).

#### 2. La revisione sistematica dichiara esplicitamente che i metodi sono stati definiti prima della sua conduzione, motivando tutte le violazioni significative del protocollo?

Considerato che le RS sono una forma di ricerca osservazionale, i metodi per la loro conduzione dovrebbero essere sempre definiti prima del loro avvio. Gli autori dovrebbero dimostrare di aver condotto la RS sulla base di un protocollo scritto e verificato da un ente indipendente, in quanto l'aderenza a un protocollo riduce il rischio di errori nella conduzione della RS.

#### 3. Gli autori motivano la scelta del disegno degli studi inclusi nella revisione?

Per alcuni quesiti (es. effetti delle modifiche a normative e politiche sanitarie) o per ragioni etiche i NRSI possono essere gli unici in grado di rispondere al quesito della revisione. Con l'estensione di AMSTAR 2 alla valutazione di revisioni che includono anche NSRI è importante che gli autori giustifichino l'inclusione dei diversi disegni di studio nelle RS, indicando inoltre quale strategia hanno seguito per farlo. Nei casi in cui sia RCT sia NSRI rispondono allo stesso quesito sull'efficacia di un intervento sanitario, gli autori dovrebbero considerare se limitare la revisione ai RCT può fornire una sintesi incompleta degli effetti rilevanti di un trattamento.

#### 4. Gli autori hanno effettuato una ricerca sistematica della letteratura?

L'importanza di un'adeguata ricerca bibliografica nelle RS è ben nota<sup>5</sup> e questo item ha subito piccole modifiche rispetto allo strumento originale. In particolare, in AMSTAR 2 sono state rese più chiare le opzioni di risposta e fornite indicazioni più dettagliate sull'item, in particolare in relazione all'identificazione di NRSI (appendice 1).

#### 5. La selezione degli studi è stata effettuata da almeno due autori in maniera indipendente?

La best practice prevede l'impiego di due revisori per determinare l'inclusione degli studi nelle RS<sup>5</sup>, al fine di verificare le caratteristiche di uno studio rispetto alle componenti del PICO. Nello strumento originale questo item

riguardava sia l'eleggibilità dello studio, sia l'estrazione dei dati. AMSTAR 2 prevede due item separati (5 e 6), in quanto il panel ha ritenuto che sono due processi distinti.

#### 6. L'estrazione dei dati è stata effettuata da almeno due autori in maniera indipendente?

Il panel ha riconosciuto che l'estrazione dei dati potrebbe essere più complessa per i NRSI, in quanto di solito comporta l'estrazione di misure di efficacia del trattamento e altre misure di associazione che vengono aggiustate per potenziali confondenti, piuttosto che i dati grezzi sugli outcome dei gruppi trattati e di controllo. Considerato che uno studio può presentare molteplici effetti di un trattamento, è necessario giudicare quale si adatta meglio al PICO ed è a minor rischio di confondimento.

#### 7. Gli autori forniscono l'elenco degli studi esclusi giustificando le motivazioni?

In AMSTAR 2 gli studi esclusi vengono considerati separatamente da quelli inclusi e devono essere considerati integralmente dagli autori delle RS, per evitare che restino invisibili e che l'impatto della loro esclusione rimanga sconosciuto.

#### 8. Gli autori descrivono con sufficiente livello di dettaglio gli studi inclusi?

AMSTAR 2 richiede agli autori delle RS di fornire dettagli su disegni di studio, popolazioni degli studi, interventi, controlli e outcome. Il livello di dettaglio dovrebbe essere tale da permettere ai valutatori di giudicare se gli studi sono stati selezionati in modo appropriato (in relazione al PICO) e se le popolazioni e gli interventi dello studio erano coerenti con i quesiti di ricerca. Queste informazioni permettono di determinare in quale misura i risultati di diversi studi dovrebbero essere combinati, per contribuire a spiegare l'eterogeneità e supportare chi applica i risultati della RS.

#### 9. Gli autori hanno utilizzato un metodo adeguato per analizzare il rischio di bias dei singoli studi inclusi nella revisione?

I bias possono essere introdotti in diverse fasi del disegno, pianificazione, conduzione e analisi di uno studio. Questo item ne sostituisce uno più generico relativo alla "qualità scientifica", specificando i domini di bias che dovrebbero essere considerati dai revisori per RCT e NRSI, in base agli strumenti della Cochrane<sup>42,43</sup>. AMSTAR 2 valuta se gli autori della RS hanno effettuato un'adeguata valutazione dei tentativi fatti in ciascuno studio per evitare, controllare o aggiustare i risultati per le variabili confondenti alla baseline, i bias di selezione, di misurazione di esposizioni ed outcome, il reporting selettivo di analisi e/o outcome. Maggiori dettagli sono disponibili nell'appendice 1 e nel ROBINS-I<sup>43</sup>. Si è deciso di non includere la valutazione dei confondenti variabili nel tempo, i bias di performance e quelli dovuti a dati mancanti, sebbene attualmente siano inclusi in ROBINS-I<sup>43</sup>. Questo è dovuto

alla natura complessa delle tecniche utilizzate per correggere queste potenziali fonti di bias e alla frequente mancanza di dati negli studi primari che ne consentano la valutazione. La versione 2.0 dello strumento della Cochrane sul RoB nei RCT, attualmente disponibile in forma di bozza<sup>44</sup>, sarà presa in considerazione da AMSTAR 2 per un aggiornamento futuro.

*10. Gli autori riportano le fonti di finanziamento degli studi inclusi nella revisione?*

È stato aggiunto un item sulle fonti di finanziamento dei singoli studi in quanto le evidenze suggeriscono che i risultati degli studi finanziati dall'industria a volte favoriscono i prodotti sponsorizzati e hanno meno probabilità di essere pubblicati rispetto agli studi indipendenti<sup>45-47</sup>. Tali influenze potrebbero non essere rilevate come criticità nel disegno o nei metodi dello studio (item 9).

*11. Se è stata condotta una meta-analisi, gli autori hanno utilizzato metodi appropriati per la combinazione statistica dei risultati?*

Questo item è stato modificato rispetto allo strumento originale e deve essere valutato separatamente per RCT e NRSI. Gli autori delle RS dovrebbero dichiarare esplicitamente nel protocollo della RS i principi per decidere se eseguire, o meno, la meta-analisi dei dati estratti dagli studi inclusi, valutando la comparabilità degli studi (in termini di pazienti, controlli e interventi) e il valore di un singolo effetto aggregato (es. da diversi studi comparabili, ma con insufficiente dimensione del campione). Se i revisori ritengono appropriato condurre una meta-analisi, l'inclusione di NRSI aumenta la complessità delle analisi e può incrementare l'eterogeneità (appendice 1).

*12. Se è stata condotta una meta-analisi, gli autori analizzano il potenziale impatto del rischio di bias dei singoli studi nei risultati della meta-analisi o nelle altre sintesi delle evidenze?*

Questo nuovo item richiede agli autori di esaminare come i risultati variano con l'inclusione/esclusione di studi primari giudicati ad elevato RoB. Nei casi in cui gli autori abbiano scelto di includere solo RCT di elevata qualità, il potenziale impatto dei bias sui risultati potrebbe essere poco rilevante. Al contrario, se hanno incluso RCT di qualità variabile e/o NSRI dovrebbero valutare l'impatto del RoB sui risultati della revisione<sup>48</sup>.

*13. Gli autori tengono in considerazione il rischio di bias nei singoli studi quando interpretano/discutono i risultati della revisione?*

Questo item è stato modificato rispetto allo strumento originale. Con una maggiore enfasi sulla valutazione del RoB gli autori dovrebbero fare riferimento esplicito ai potenziali impatti del RoB nell'interpretazione e nella discussione dei risultati della RS, nel trarre conclusioni e nel formulare raccomandazioni.

*14. Gli autori spiegano e discutono in maniera soddisfacente ogni eterogeneità osservata nei risultati della revisione?*

Questo item è formulato in maniera diversa rispetto allo strumento originale. È importante che gli autori analizzino le possibili cause di eterogeneità, compresa la variazione degli elementi inclusi nel PICO (item 1) e quelli derivanti da considerazioni sui disegni di studio o metodologiche (item 9). Con l'inclusione di NRSI le variazioni nel disegno e nell'analisi possono aumentare l'eterogeneità.

*15. Se è stata effettuata una meta-analisi, gli autori hanno esplorato adeguatamente il bias di pubblicazione e discusso il potenziale impatto sui risultati della revisione?*

Questo item è formulato in maniera diversa rispetto allo strumento originale. Il bias di pubblicazione è un problema rilevante, ma per gli autori può essere difficile annullarne completamente l'impatto. Infatti, i test statistici o le rappresentazioni grafiche non hanno una elevata sensibilità, per cui se positivi indicano la presenza del bias di pubblicazione, ma se negativi non ne garantiscono l'assenza. Sono richiesti almeno 10 studi per mostrare un'asimmetria nel *funnel plot*<sup>5</sup>. La tendenza a pubblicare selettivamente studi positivi di piccole dimensioni può essere ulteriormente aggravata da vari fattori: bassa qualità metodologica, maggiore tendenza a riportare selettivamente i risultati, maggiore eterogeneità clinica se condotti in sottogruppi di pazienti<sup>49</sup>.

*16. Gli autori hanno riportato ogni fonte potenziale di conflitto di interessi, includendo anche eventuali finanziamenti ricevuti per condurre la revisione?*

Questo item è formulato in maniera diversa rispetto allo strumento originale, in quanto ora è distinto da quello relativo ai finanziamenti degli studi primari inclusi nella revisione (item 10). Di conseguenza, gli autori delle RS dovrebbero segnalare le loro fonti di finanziamento<sup>50,51</sup>.

### Identificazione di item critici

Tutti gli step nella conduzione di una RS e di una meta-analisi sono importanti, ma 7 item possono avere un impatto rilevante sulla sua validità e sulle conclusioni (box 1). Due di questi riguardano il RoB, se è stato valutato adeguatamente e in che modo può influenzare i risultati della RS. L'importanza attribuita al RoB è motivata dal fatto che AMSTAR 2 sarà utilizzato per valutare molte RS che includono NRSI.

Gli item elencati nel box 1 non sempre sono critici; ad esempio, quelli correlati al RoB possono essere considerati meno importanti quando una RS è limitata a RCT di elevata qualità, oppure quando viene eseguita la meta-analisi per sintetizzare evidenze già consolidate in letteratura (es. il risultato di uno o più gruppi collaborativi di trial clinici). In questa circostanza, l'adeguatezza della ricerca bibliografica (item 4), l'elenco degli studi esclusi

(item 7) e la possibilità di bias di pubblicazione (item 15) potrebbero non essere considerati critici. Se non è stata eseguita una meta-analisi, l'item 11 relativo all'adeguatezza dei metodi meta-analitici non si applica. Tuttavia, in questa circostanza è importante che i valutatori siano attenti al possibile impatto del RoB quando i singoli studi selezionati dagli autori vengono descritti in maniera narrativa.

Risposte negative relative agli item identificati come critici potrebbero non essere rilevanti quando ulteriori informazioni (es. direttamente dagli autori) indicano che la risposta originale era sbagliata. Ciò può assicurare riguardo i risultati della RS o consentire una modifica della stessa tramite ulteriori analisi. In ogni caso l'elenco riportato è puramente orientativo visto che gli esperti possono aggiungere o sostituire altri item critici. Ad esempio, la mancata inclusione di NRSI (item 3) in una RS che valuta gli effetti avversi di un trattamento può essere una lacuna rilevante, così come l'incapacità di spiegare ampie variazioni degli effetti del trattamento tra gli studi (item 14).

### Applicare AMSTAR 2 alle revisioni sistematiche

Quando una o più RS devono fornire la base scientifica per rilevanti pratiche cliniche o politiche sanitarie, si raccomanda al gruppo di valutazione di concordare le modalità per applicare gli item di AMSTAR 2. Questo include la pratica clinica o il contesto sanitario e i quesiti che dovrebbero essere affrontati, secondo le componenti rilevanti del PICO. Ad esempio, le RS disponibili dovrebbero aver incluso studi con confronti o tempi di follow-up diversi e la loro rilevanza per quesiti di ricerca importanti per le politiche sanitarie deve ancora essere indagata. Dovrebbe esserci accordo anche sulle possibili fonti di bias. Ad esempio, negli studi osservazionali sull'efficacia degli interventi sanitari, gli effetti confondenti sulle indicazioni terapeutiche (o sulla severità della malattia) possono essere problematici quando gli interventi sono destinati a specifici sottogruppi di pazienti<sup>52</sup>. È buona prassi arruolare nuovi utilizzatori di una tecnologia o di un farmaco negli studi, per evitare il bias di prevalenza<sup>53</sup>. Se l'inizio di un intervento tende a essere ritardato, la scelta del confronto può introdurre un *immortal time bias*<sup>54</sup>. Gli errori di misurazione possono determinare un'errata classificazione di esposizioni ed outcome, con sbilanciamenti tra i gruppi di confronto. Il reporting selettivo tra più analisi e outcome può fornire una misura non accurata degli effetti di un intervento sanitario.

L'appendice 1 fornisce una guida alle sezioni di AMSTAR 2. Alcuni giudizi (in particolare se gli autori delle RS hanno adeguatamente valutato il RoB nei NRSI) sono complessi e potrebbero essere necessari ulteriori approfondimenti sia metodologici che di contenuto. Talvolta è necessario conoscere lo specifico argomento per determinare se gli autori della RS hanno effettuato un'adegua-

### Box 1. Item critici di AMSTAR 2

- Registrazione del protocollo prima di avviare la RS (item 2)
- Conduzione di una ricerca bibliografica esaustiva (item 4)
- Descrizione delle motivazioni per l'esclusione dei singoli studi (item 7)
- Inclusione nella RS del RoB dei singoli studi (item 9)
- Utilizzo di metodi appropriati di meta-analisi (item 11)
- Considerazione del RoB nell'interpretazione dei risultati della RS (item 13)
- Valutazione della presenza e del potenziale impatto del bias di pubblicazione (item 15)

ta valutazione degli elementi rilevanti per il PICO (item 1) e per identificare potenziali fattori confondenti.

Le singole valutazioni degli item non devono essere combinate per creare un punteggio complessivo<sup>55,56</sup>: piuttosto, gli utenti di AMSTAR 2 dovrebbero considerare il potenziale impatto di una valutazione inadeguata per ciascun item.

Il box 2 riporta uno schema per interpretare i punti di debolezza rilevati negli item critici e non critici: la sua funzione è meramente informativa perché spetta agli esperti decidere quali item sono più rilevanti per le RS in esame.

### Box 2. Valutare l'affidabilità complessiva dei risultati della revisione sistematica

#### Elevata

- *0-1 punto di debolezza non critico*. La RS fornisce una sintesi accurata e completa dei risultati degli studi disponibili che rispondono al quesito di interesse

#### Moderata

- *>1 punto di debolezza non critico\**. La RS presenta oltre un punto di debolezza, ma non reali criticità; può fornire una sintesi accurata dei risultati degli studi disponibili inclusi nella revisione

#### Bassa

- *1 punto critico ± punti di debolezza non critici*. La RS presenta una criticità e potrebbe non fornire una sintesi accurata e completa degli studi disponibili che rispondono al quesito di interesse

#### Molto bassa

- *≥1 punto critico ± punti di debolezza non critici*. La RS presenta più di una criticità e non fornisce una sintesi accurata e completa degli studi disponibili

\*Considerato che la presenza di diversi punti di debolezza non critici può ridurre l'affidabilità complessiva dei risultati della RS, in questi casi potrebbe essere opportuno spostare la valutazione complessiva da moderata a bassa

### Accordo tra valutatori nell'uso di AMSTAR 2

È stato misurato con tre coppie di valutatori e tre set di RS. La prima coppia era coinvolta nello sviluppo di AMSTAR 2: MT e CH hanno valutato individualmente 20 RS derivate da una ricerca rapida in Google Scholar (condotta nel 2015 con i termini "revisione sistematica" e "meta-analisi" nel titolo); dalle prime 200 sono state selezionate 20 RS su qualsiasi intervento sanitario. Le altre due coppie di valutatori avevano esperienza pregressa nella valutazione critica di RS e non erano coinvolti nello sviluppo di AMSTAR o AMSTAR 2: hanno applicato AMSTAR 2 durante il loro lavoro quotidiano, valutando RS su interventi per ridurre gli errori farmacologici (14 revisioni) e su terapie non farmacologiche per il morbo di Parkinson (20 revisioni). In entrambi i casi le RS sono state identificate tramite ricerche bibliografiche esaustive (dettagli disponibili su richiesta). Tutti i valutatori avevano a disposizione la guida per l'utente (appendice 1) ed hanno applicato individualmente lo strumento senza cercare di raggiungere un consenso. Complessivamente, 6 valutatori hanno applicato lo strumento a 54 RS, di cui 20 includevano solo RCT, 18 includevano solo NSRI e 16 includevano una combinazione di entrambi i disegni di studio.

L'accordo tra le tre coppie di valutatori sui tre set di RS variava in maniera sostanziale sia tra gli item, sia tra le coppie di valutatori. La maggior parte dei valori (46/50) di k si collocavano in un range accettabile (accordo tra moderato ed eccellente) e 39/50 tra buono ed eccellente.

Non c'erano grandi differenze tra i valutatori, e tra quelli coinvolti nello sviluppo di AMSTAR 2 sono stati riscontrati livelli di accordo più elevati rispetto ai valutatori non coinvolti. Gli item 9, 12 e 13 riguardano la misurazione del RoB e il modo in cui questo viene gestito durante la discussione della meta-analisi e dell'interpretazione dei risultati. I range di valori k per questi item erano simili a quelli osservati per altri item dello strumento. Per gli item 9 e 11, i valori k per la valutazione del RoB dei RCT erano simili a quelli dei NSRI.

### Usabilità di AMSTAR 2

Escludendo il tempo dedicato alla lettura delle RS, i revisori 1 e 2 hanno impiegato per la valutazione delle 20 RS da 15 a 32 minuti, un tempo quasi doppio rispetto a quello necessario a completare lo strumento AMSTAR originale (range 10-15 minuti) su RS di RCT<sup>57</sup>. I commenti dei revisori sono stati:

- la rimozione delle opzioni di risposta "impossibile rispondere" e "non applicabile" dello strumento originale impone di esprimere un giudizio;
- la valutazione delle revisioni di NSRI e misti richiede più tempo, ma costringe ad affrontare rilevanti questioni metodologiche;
- spesso gli autori delle RS menzionavano la presenza o l'assenza di bias di pubblicazione senza tuttavia fornire alcuna prova tangibile;

- gli autori delle RS hanno dichiarato i loro potenziali conflitti di interesse, senza riportare le modalità con cui li hanno gestiti.

### Discussione

AMSTAR 2 rappresenta una revisione radicale dello strumento originale, progettato per valutare le RS che includevano solo RCT<sup>22-24</sup>. Le modifiche principali riguardano la semplificazione delle categorie di risposta; una considerazione più dettagliata del RoB degli studi inclusi e le modalità con cui gli autori hanno gestito il RoB nel sintetizzare e interpretare i risultati della RS; un migliore allineamento con il framework PICO per i quesiti di ricerca; motivazioni più dettagliate per la selezione dei disegni di studio da includere nella RS; ulteriori informazioni sugli studi esclusi dalla RS. Inoltre, AMSTAR 2 consiglia di definire gli item critici prima di iniziare la valutazione critica di una RS, perché l'eventuale identificazione di punti di debolezza su questi item dovrebbe compromettere l'affidabilità dei risultati di una RS. I giudizi sugli item di AMSTAR 2 non devono essere utilizzati per ricavare un punteggio complessivo<sup>55,56</sup>. Anche se lo strumento AMSTAR originale è stato spesso utilizzato per questo scopo - anche grazie al sito web ([www.amstar.ca](http://www.amstar.ca)) - il punteggio complessivo rischia di non far emergere i punti critici di debolezza che riducono l'affidabilità di una RS. Si raccomanda pertanto di adottare un processo di valutazione basato sull'identificazione degli item critici (box 2), o che poco si discosti da questi principi<sup>56</sup>.

Auspichiamo che AMSTAR 2 venga utilizzato come supporto didattico e come checklist dagli autori di RS. Tuttavia, lo strumento non spiega in dettaglio logica e metodi di conduzione delle RS, per i quali si rimanda all'*Handbook Cochrane*<sup>5</sup>.

La considerazione del RoB nei singoli studi è importante sia per i RCT che per i NSRI, anche se generalmente è meglio conosciuta per i primi. NSRI di grandi dimensioni, spesso condotti tramite grandi database amministrativi, sono sempre più utilizzati per valutare l'impatto nel real world di numerosi interventi sanitari. Sebbene tali studi spesso utilizzino metodi sofisticati, fattori confondenti residui o altre fonti di bias possono generare stime imprecise di efficacia. Di conseguenza la loro inclusione nelle meta-analisi può generare stime precise ma parziali degli effetti di un intervento sanitario<sup>32</sup>.

Gli item di AMSTAR 2 che riguardano il RoB identificano i domini specificati negli strumenti di valutazione del RoB della Cochrane per RCT e NSRI<sup>42,43</sup>, definiti tramite consenso da oltre 30 metodologi. Tuttavia, AMSTAR 2 attualmente non specifica quali strumenti dovrebbero essere utilizzati per valutare il RoB dei NSRI inclusi in una RS. ROBINS-I, lo strumento più completo per NSRI, è stato pubblicato nel 2016 e non è realistico che gli autori di RS avviate prima della sua pubblicazione possano averlo

utilizzato<sup>43</sup>. Attualmente AMSTAR 2 lascia agli autori della RS e a coloro che ne valutano la qualità la libertà di accertarsi che lo strumento di analisi del RoB utilizzato sia in grado di discriminare i relativi domini. Una revisione di Sanderson et al. identifica 86 strumenti per valutare la qualità degli studi osservazionali, senza esprimere una netta preferenza tra i vari strumenti<sup>58</sup>. Gli autori hanno sottolineato la necessità di definire elementi critici per valutare la presenza di bias nell'epidemiologia osservazionale. In parte questa revisione ha contribuito allo sviluppo dello strumento ROBINS-I<sup>43</sup>. Gli strumenti comunemente utilizzati per la valutazione di singoli studi, come la *Newcastle Ottawa Scale* e la checklist dello *Scottish Intercollegiate Guidelines Network* (SIGN), possono non concentrarsi solo sulla validità<sup>59,60</sup>. La Newcastle Ottawa scale sembra non avere sufficiente sensibilità e talora viene utilizzata per generare uno score complessivo, non raccomandato perché potrebbe nascondere eventuali punti critici di debolezza in una RS<sup>56,61</sup>.

Come strumento di valutazione critica per le RS, AMSTAR 2 integra altri strumenti progettati a fini analoghi<sup>3,4,16,17,19,20,25,62</sup> quali le linee guida per il reporting e gli standard per la metodologia di conduzione delle RS<sup>3,4</sup>. Due strumenti molto citati sono stati utilizzati come base per lo sviluppo dello strumento AMSTAR originale<sup>16,17,22</sup>, da cui sono poi derivati due strumenti pubblicati<sup>19,25</sup>. Un'altra pubblicazione include una checklist utilizzata per valutare le RS inserite in una *umbrella review*<sup>20</sup>, che si sovrappone ampiamente allo strumento AMSTAR originale<sup>22</sup>.

AMSTAR 2 fornisce un'ampia valutazione della qualità delle RS, incluse le criticità che possono originare da una inadeguata conduzione, con un impatto incerto sui risultati. In tal senso differisce dal *Risk Of Bias In Systematic reviews* (ROBIS)<sup>62</sup>, un sofisticato strumento in tre step che si concentra specificamente sul rischio di presenza dei bias introdotti dalla conduzione della RS. ROBIS copre la maggior parte dei quesiti di ricerca (diagnosi, prognosi e eziologia), mentre AMSTAR 2 è finalizzato alla valutazione di RS sull'efficacia degli interventi sanitari. Inevitabilmente esiste una sovrapposizione di item tra ROBIS e AMSTAR 2: peraltro due ricercatori (BCR, BJS) sono stati coinvolti nello sviluppo di entrambi gli strumenti.

AMSTAR 2 mantiene l'approccio ben noto di checklist guidata, con l'aggiunta e la modifica di item, e sarà familiare agli utilizzatori dello strumento originale, anche se più impegnativo da utilizzare per le ragioni sopra esposte. Poiché AMSTAR 2 è strutturato secondo gli step principali della conduzione di una RS, può essere utilizzato come un supporto didattico o come checklist di verifica per gli autori di RS.

A differenza dello strumento originale, AMSTAR 2 identifica alcuni item critici (box 1) che potrebbero ridurre l'affidabilità di una RS e chiede agli utenti di spe-

cificare eventuali variazioni in relazione all'argomento della revisione. Verosimilmente ci sarà un dibattito sugli item di questo elenco e gli utilizzatori potrebbero voler pre-definire un set differente di item critici in relazione ai quesiti PICO o al setting.

Non è stata eseguita una validazione approfondita di AMSTAR 2. Rispetto allo strumento originale sono stati mantenuti 10 item, anche se con alcune modifiche nella formulazione in seguito ai feedback e all'esperienza d'uso. A 2 item è stata data una copertura più dettagliata: la selezione degli studi in doppio e l'estrazione dei dati, originariamente incluse nello stesso item, sono ora item differenti; sono state inoltre aggiunte considerazioni più dettagliate e distinte del RoB per RCT e NRSI. Gli item secondari sono derivati da strumenti Cochrane ampiamente utilizzati. Un item è stato rimosso; la considerazione della letteratura grigia, che precedentemente era un item separato, è ora inclusa nell'item relativo ricerca bibliografica. In totale, sono stati aggiunti 4 item. Due di questi provengono direttamente dallo strumento ROBINS-I, ovvero l'elaborazione del PICO nella revisione e le modalità con cui il RoB è stato gestito durante la sintesi delle evidenze<sup>43</sup>. Un altro dei nuovi item, la discussione delle possibili cause e l'importanza dell'eterogeneità, è un'elaborazione di quanto contenuto nello strumento originale<sup>22</sup>. L'item finale, ovvero la motivazione della scelta dei disegni di studio, è giustificato dalla necessità di adattare AMSTAR ai NRSI. Non riteniamo che questo necessiti di validazione in quanto gli autori di RS dovrebbero sempre giustificare il motivo per cui hanno incluso disegni di studio più esposti a bias.

I livelli di accordo raggiunti dalle tre coppie di valutatori variavano tra i vari item, ma erano da moderati a rilevanti per la maggior parte degli essi: in particolare, l'accordo tra due valutatori coinvolti nello sviluppo di AMSTAR 2 non era maggiore di quello raggiunto tra valutatori esperti non coinvolti nel suo sviluppo. Non ci si attendeva un accordo perfetto: le differenze tra valutatori riflettono infatti la complessità di giudizio sui singoli item e dovrebbero stimolare la discussione di gruppo su motivazioni e rilevanza, e, se necessario, il confronto con esperti della materia e di metodologia.

Nello sviluppo di AMSTAR 2 ci si è basati sul consenso del panel, ma anche sull'ampio riscontro da parte degli utilizzatori dello strumento originale sotto forma di comunicazioni dirette, commenti sul sito web e valutazioni effettuate in occasione dei workshop didattici ai Cochrane Colloquium. Nelle fasi successive di sviluppo di AMSTAR 2 sono state analizzate e discusse alcune critiche mosse ad AMSTAR recentemente pubblicate<sup>25-31</sup>.

La nostra esperienza di pubblicazione e utilizzo dello strumento AMSTAR originale è che tutti i giudizi devono essere espressi e i valutatori possono decidere se apportare eventuali modifiche allo strumento<sup>25,26,30</sup>.

Incoraggiamo gli utilizzatori a fornire feedback e, nel caso adattino lo strumento a particolari setting, a segnalare la loro esperienza su [www.amstar.ca](http://www.amstar.ca).

### Ringraziamenti

Ringraziamo per la loro assistenza nello sviluppo degli strumenti di valutazione critica AMSTAR: Lex Bouter, Maarten Boers, Alonso Carrasco-Labra, Jeremy Grimshaw, Ranjeeta Mallick, Jordi Pardo-Pardo e Larissa Shamseer. Per la conduzione degli studi di affidabilità riportati in questo manoscritto ringraziamo Brian Hutton, Pauline Barbeau, Fatemeh Yazdi, Vesa Basha e Roxanne Ward.

### Contributo degli autori

BJS, DAH, GW e PT hanno concepito il progetto. BJS e DAH hanno supervisionato il progetto. BJS e DAH hanno guidato il gruppo di lavoro. Tutti gli autori hanno contribuito allo sviluppo di AMSTAR 2 e alla stesura delle relative linee guida. BJS, DAH, BCR e PT hanno contribuito alla redazione e alla riformulazione del manoscritto. Tutti gli altri autori hanno esaminato e commentato le bozze del manoscritto. BJS e DAH sono i garanti.

### NOTE ALLA VERSIONE ITALIANA

La Fondazione GIMBE ha realizzato la traduzione italiana dell'articolo senza alcun finanziamento istituzionale o commerciale.

### TEAM CHE HA REALIZZATO LA VERSIONE ITALIANA

#### Responsabile scientifico

Antonino Cartabellotta, Fondazione GIMBE

#### Coordinamento editoriale

Marco Mosti, Fondazione GIMBE

#### Traduzione

Tiziano Innocenti, Fisioterapista

#### Revisione editoriale

Elena Cottafava, Fondazione GIMBE

### MATERIALE SUPPLEMENTARE

Checklist AMSTAR 2 in formato Word

### BIBLIOGRAFIA

1. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010;7:e1000326.
2. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;309:597-9.
3. Moher D, Alessandro Liberati, Tetzlaff J, Altman DG, and the PRISMA Group. Preferred Reporting Items for (SR) and Meta-Analyses: The PRISMA Statement. *Ann Intern Med* 2009;6:264-9.
4. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *Meta-analysis Of Observational Studies in Epidemiology*

(MOOSE) group. *JAMA* 2000;283:2008-12.

5. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook). Ultimo accesso: 18 febbraio 2020.
6. Dechartres A, Charles P, Hopewell S, Ravaut P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. *J Clin Epidemiol* 2011;64:136-44.
7. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
8. Wong WC, Cheung CS, Hart GJ. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. *Emerg Themes Epidemiol* 2008;5:23.
9. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998;51:1235-41.
10. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-84.
11. Murray J, Farrington DP, Eisner MP. Drawing conclusions about causes from systematic reviews of risk factors: The Cambridge Quality Checklists. *J Exp Criminol* 2009;5:1-23.
12. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651-7.
13. Bérard A, Andreu N, Tétrault J, Niyonsenga T, Myhal D. Reliability of Chalmers' scale to assess quality in meta-analyses on pharmacological treatments for osteoporosis. *Ann Epidemiol* 2000;10:498-503.
14. Thompson S, Ekelund U, Jebb S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. *Int J Epidemiol* 2011;40:765-77.
15. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. *Systematic reviews in health care: meta-analysis in context*. Wiley & Sons, 2008;285-312.
16. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987;316:450-5.
17. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44:1271-8.
18. Oxman AD, Cook DJ, Guyatt GH. Evidence-Based Medicine Working Group. Users' guides to the medical literature. VI. How to use an overview. *JAMA* 1994;272:1367-71.
19. Scottish Intercollegiate Guidelines Network. *Methodology Checklist 1: Systematic Reviews and Meta-analyses*. Disponibile a: <https://www.sign.ac.uk/media/1721/srchecklist.doc>. Ultimo accesso: 18 febbraio 2020.
20. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews:

methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc* 2015;13:132-40.

21. Whiting P, Savović J, Higgins JP, et al. ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.

22. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.

23. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62:1013-20.

24. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* 2007;2:e1350.

25. Kung J, Chiappelli F, Cajulis OO, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. *Open Dent J* 2010;4:84-91.

26. Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol* 2015;68:574-83.

27. Faggion CM Jr. Critical appraisal of AMSTAR: challenges, imitations, and potential solutions from the perspective of an assessor. *BMC Med Res Methodol* 2015;15:63.

28. Teich ST, Heima M, Lang L. Dental Students' Use of AMSTAR to Critically Appraise Systematic Reviews. *J Dent Educ* 2015;79:1031-9.

29. Burda BU, Holmer HK, Norris SL. Limitations of a measurement tool to assess systematic reviews (amstar) and suggestions for improvement. *Syst Rev* 2016;5:58.

30. Wegewitz U, Weikert B, Fishta A, Jacobs A, Pieper D. Resuming the discussion of AMSTAR: What can (should) be made better? *BMC Med Res Methodol* 2016;16:111.

31. Dahm P. Raising the bar for systematic reviews with Assessment of Multiple Systematic Reviews (AMSTAR). *BJU Int* 2017;119:193.

32. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998;316:140-4.

33. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371:569-78.

34. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13:e1002028.

35. Shapero S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;140:771-8.

36. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol* 1991;44:127-39.

37. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889-97.

38. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of

individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.

39. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313-24.

40. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review - a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10(Suppl 1):21-34.

41. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8:19-32.

42. Higgins JP, Altman DG, Gøtzsche PC, et al. Cochrane Bias Methods Group. Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. *BMJ* 2011;343:d5928.

43. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

44. National Collaborating Centre for Methods and Tools (2017). Appraising the risk of bias in randomized trials using the Cochrane Risk of Bias Tool. Hamilton, ON: McMaster University. (Updated 1 September, 2017). Disponibile a: [www.nccmt.ca/resources/search/280](http://www.nccmt.ca/resources/search/280). Ultimo accesso: 18 febbraio 2020.

45. DeAngelis CD, Fontanarosa PB. Impugning the integrity of medical science: the adverse effects of industry influence. *JAMA* 2008;299:1833-5.

46. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.

47. Yaphe J, Edman R, Knishkowsky B, Herman J. The association between funding by commercial interests and study outcome in randomized controlled drug trials. *Fam Pract* 2001;18:565-8.

48. Bilandzic A, Fitzpatrick T, Rosella L, Henry D. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13:e1001987.

49. Nüesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;341:c3515.

50. Bero L, El-Hachem P, Abou-Haidar H, Neumann I, Schünemann HJ, Guyatt GH. What is in a name? Nonfinancial influences on the outcomes of systematic reviews and guidelines. *J Clin Epidemiol* 2014;67:1239-41.

51. Committee on Publication Ethics. Code of conduct and best practice guidelines for journal editors. Disponibile a: [https://publicationethics.org/files/Code\\_of\\_conduct\\_for\\_journal\\_editors\\_Mar11.pdf](https://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf). Ultimo accesso: 18 febbraio 2020.

52. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol* 1999;149:981-3.

53. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158:915-20.

54. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008;167:492-9.

55. Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001;2:463-71.

56. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.

57. Shea, BJ Assessing the Methodological Quality of Systematic Reviews: The Development of AMSTAR, PhD, Vrije Universiteit Amsterdam 2008: page 70. Disponibile a: <https://research.vu.nl/en/publications/assessing-the-methodological-quality-of-systematic-reviews-the-de>. Ultimo accesso: 18 febbraio 2020.
58. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666-76.
59. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of observational studies in metaanalyses. Disponibile a: [www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). Ultimo accesso: 18 febbraio 2020.
60. Scottish Intercollegiate Guidelines Network SIGN 50: Methodology Checklist 3: Cohort Studies. Disponibile a: [https://www.sign.ac.uk/media/1712/checklist\\_for\\_cohort\\_studies.rtf](https://www.sign.ac.uk/media/1712/checklist_for_cohort_studies.rtf). Ultimo accesso: 18 febbraio 2020.
61. McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. *PLoS Med* 2011;8:e1001098.
62. Whiting P, Savović J, Higgins JP, et al. ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.

### APPENDICE 1. GUIDA ALLA COMPILAZIONE DI AMSTAR 2

Molti degli item di AMSTAR 2 sono formulati in maniera auto-esplicativa. Tuttavia, gli aspetti valutati sono spesso complessi e soggetti a interpretazioni diverse, in particolare quando i giudizi vengono formulati per un ampio spettro di interventi. Questa appendice fornisce ulteriori raccomandazioni per l'uso di AMSTAR 2. Le sovrapposizioni con il contenuto dell'articolo sono intenzionali, perché questa appendice può essere utilizzata in maniera indipendente.

Questa guida fornisce un'indicazione delle modalità con cui applicare i criteri quando le RS sono condotte su interventi sanitari (di solito clinici) ben definiti. Gli utenti individuali di AMSTAR 2, naturalmente, potrebbero ritenere necessario non seguire le indicazioni fornite, sia sui singoli item sia nella valutazione complessiva di una RS. È opportuno che queste variazioni vengano documentate, così da permettere ad altri utenti di beneficiare dalle loro esperienze.

AMSTAR 2 non è progettato per generare uno score complessivo. Infatti, dietro un punteggio elevato potrebbero nascondersi criticità su specifici item (es. ricerca bibliografica inadeguata o incapacità di valutare il ROB negli studi inclusi nella RS). Nell'eseguire una valutazione complessiva della RS, è importante tenere conto di eventuali problematiche negli item critici, che potrebbero ridurre notevolmente l'affidabilità dei risultati di una RS.

#### Item 1. I quesiti di ricerca e i criteri di inclusione della revisione comprendono gli elementi del PICO?

Il PICO (popolazione, intervento, controllo, outcome) viene frequentemente utilizzato come framework pratico e facilmente memorizzabile per i quesiti di ricerca. Talora è necessario prevedere anche un arco temporale per essere certi che lo studio rilevi outcome clinici rilevanti (es. quando l'effetto dell'intervento è prevedibile solo dopo diversi anni).

Il PICO identifica gli elementi che dovrebbero essere descritti in dettaglio nel report della RS, consentendo al valutatore di giudicare la selezione degli studi e la loro combinabilità, e al lettore di determinare l'applicabilità dei risultati. Gli autori di RS non sempre rendono espliciti gli elementi del PICO, che dovrebbero comunque essere rilevabili attraverso un'attenta lettura delle sezioni abstract, introduzione e metodi. Per assegnare "Sì", i 4 elementi del PICO dovrebbero essere descritti in una specifica sezione del report.

#### Item 2. La revisione sistematica dichiara esplicitamente che i metodi sono stati definiti prima della sua conduzione, motivando tutte le violazioni significative del protocollo?

Le RS sono una forma di ricerca osservazionale e i metodi per la loro conduzione dovrebbero essere sempre defini-

ti prima del loro avvio. Gli autori dovrebbero dimostrare di aver condotto la RS sulla base di un protocollo scritto e verificato da un ente indipendente, in quanto l'aderenza a un protocollo riduce il rischio di errori nella conduzione della RS. Il protocollo può essere pubblicato in un registro (es. PROSPERO - [www.crd.york.ac.uk/PROSPERO](http://www.crd.york.ac.uk/PROSPERO)), su una rivista open access (es. BMJ Open) o sottomesso, con evidenza di data, presso un centro di ricerca o un comitato etico. I quesiti di ricerca e i metodi dovrebbero essere pianificati prima di condurre la RS. Come minimo, questo dovrebbe essere indicato nel report finale (punteggio "Sì, in parte"). Per assegnare "Sì" gli autori dovrebbero dimostrare di aver lavorato sulla base di un protocollo scritto e verificato da un ente indipendente (un registro, un altro organismo indipendente quale un comitato etico o un centro di ricerca) prima di avviare la RS. I valutatori dovrebbero confrontare il report pubblicato della RS con il protocollo registrato, se disponibile. Quando vengono rilevate deviazioni rispetto al protocollo, i valutatori dovrebbero identificare se sono state segnalate e giustificate dagli autori della RS. Evidenti discrepanze non motivate comportano un declassamento del giudizio.

#### Item 3. Gli autori motivano la scelta del disegno degli studi inclusi nella revisione?

Considerato che la selezione dei tipi di studio da includere nelle RS non dovrebbe essere arbitraria: di conseguenza, gli autori dovrebbero indicare di avere seguito una strategia. La regola generale (che può essere dedotta da ciò che gli autori hanno effettivamente scritto) è chiedersi innanzitutto se limitare la revisione ai RCT può fornire una sintesi incompleta degli effetti di un trattamento per vari ragioni: indisponibilità di RCT pertinenti o risultati mancanti (di solito relativi agli eventi avversi) in quelli disponibili, insufficiente potenza statistica, popolazioni estremamente selezionate, trattamenti di controllo/intervento non adeguati. Se la risposta a questa domanda è positiva è giustificata l'inclusione di NSRI. Al contrario, per giustificare la restrizione della RS ai soli RCT, gli autori dovrebbero sostenere che essi da soli possono fornire un quadro completo degli effetti del trattamento. La restrizione di una revisione a soli NRSI è giustificata quando gli RCT non possono fornire gli outcome necessari, o quando siano già state completate nel caso RS di RCT e quindi una revisione di NRSI integra quanto già noto. L'inclusione di RCT e NRSI può essere giustificata per un quadro completo dell'efficacia e degli eventi avversi associati ad un intervento sanitario. In questa situazione si consiglia che questi due tipi di studi siano valutati e combinati in modo indipendente (se la meta-analisi è appropriata). In ogni caso, si tratta di un'area alquanto trascurata e anche con una guida di riferimento può essere difficile giudicare la misura in cui una RS soddisfa i criteri di valutazione.

La motivazione della selezione dei disegni di studio può essere dedotta solo dopo attenta lettura dell'intera RS.

#### **Item 4. Gli autori hanno effettuato una ricerca sistematica della letteratura?**

La ricerca bibliografica dovrebbe essere eseguita utilizzando almeno due database. Il report dovrebbe includere range temporale e i database utilizzati (es. Central, EMBASE e MEDLINE). Parole chiave e/o termini MeSH devono essere riportati e la strategia di ricerca completa deve essere disponibile su richiesta. Le ricerche bibliografiche sui database dovrebbero essere integrate consultando le revisioni già pubblicate, i registri di trial, gli esperti del settore e analizzando le voci bibliografiche degli studi reperiti. Talora è necessario rivolgersi agli autori degli studi originali per chiarimenti sui risultati o aggiornamenti/correzioni. Dovrebbero essere ricercate le pubblicazioni in tutte le lingue rilevanti ed eventuali restrizioni linguistiche devono essere motivate. Inoltre, in alcuni casi è necessario ricercare nella letteratura grigia, dal momento che alcune pubblicazioni potrebbero non essere state sottoposte a *peer review*. Questo è particolarmente rilevante per policy report e i programmi di valutazione, disponibili solo sul web (es. agenzie governative, non governative o di HTA). Se la letteratura grigia è considerata importante, gli autori dovrebbero cercare fonti appropriate, come registri di trial, abstract di conferenze, tesi e report non pubblicati disponibili online (es. siti web di Università, ResearchGate). Inoltre, i trial sugli interventi sanitari potrebbero non essere stati pubblicati su riviste *peer review*, ma potrebbero essere richiesti direttamente agli sponsor o ai ricercatori. Per attribuire "Sì", i valutatori dovrebbero verificare che gli autori della RS abbiano tenuto in considerazione tutti gli aspetti rilevanti della ricerca bibliografica.

#### **Item 5. La selezione degli studi è stata effettuata da almeno due autori in maniera indipendente?**

La best practice prevede che due revisori indipendenti valutino l'eleggibilità dei singoli studi nelle RS, verificandone le caratteristiche (da titolo, abstract e full text) rispetto alle componenti del PICO. Gli autori dovrebbero riportare sia l'accordo tra valutatori in un campione di studi considerati per l'inclusione nella RS, sia il metodo utilizzato per raggiungere il consenso in caso di disaccordo nella selezione degli studi. Se la selezione di tutti gli studi è stata effettuata da un solo autore, con un secondo che ha verificato l'accordo su un campione di studi, si raccomanda di raggiungere un accordo forte (k score  $\geq 0.80$ ). La selezione degli studi dovrebbe essere effettuata almeno da due valutatori indipendenti e, in caso di disaccordo dovrebbe essere utilizzato un processo di consenso. Quando la selezione degli studi è effettuata da un solo autore, un secondo dovrebbe verificare l'accordo su un campione di studi rappresentativi raggiungendo un accordo forte (k score  $\geq 0.80$ ).

#### **Item 6. L'estrazione dei dati è stata effettuata da almeno due autori in maniera indipendente?**

L'estrazione dei dati dovrebbe essere effettuata da almeno due autori in maniera indipendente e utilizzato un processo di consenso in caso di disaccordo. Nel caso in cui l'estrazione dei dati sia effettuata da un solo autore, un secondo dovrebbe verificare l'accordo su un campione di studi rappresentativi raggiungendo un accordo forte (k score  $\geq 0.80$ ).

#### **Item 7. Gli autori forniscono l'elenco degli studi esclusi giustificando le motivazioni?**

Gli autori devono fornire l'elenco completo degli studi esclusi, anche se potenzialmente rilevanti, riportando le relative motivazioni. Tale esclusione può essere resa necessaria per l'inappropriatezza o la non rilevanza di popolazioni, interventi e controlli. Il rischio di bias (RoB), trattato separatamente e successivamente nel processo di revisione, non rappresenta in questa fare un criterio di esclusione. Considerato che l'esclusione non giustificata può influenzare i risultati della RS, nelle prime fasi di una RS si raccomanda un approccio più inclusivo.

#### **Item 8. Gli autori descrivono con sufficiente livello di dettaglio gli studi inclusi?**

La RS dovrebbe riportare la descrizione di partecipanti, interventi, controlli, risultati, disegno, analisi e setting dei singoli studi. Il livello di dettaglio dovrebbe essere tale da permettere ad un valutatore/utilizzatore di stabilire se gli studi sono stati selezionati in maniera appropriata (in relazione alla struttura del PICO) e se le popolazioni e gli interventi sono rilevanti per la propria pratica. Tali descrizioni forniscono anche un framework per studiare l'eterogeneità negli effetti dell'intervento (es. dose, fascia d'età, setting assistenziale, etc.).

#### **Item 9: Gli autori hanno utilizzato un metodo adeguato per analizzare il rischio di bias dei singoli studi inclusi nella revisione?**

Questa è una sezione cruciale della valutazione di qualsiasi RS, in particolare di quelle che includono NRSI. La domanda chiave per la valutazione è se gli autori delle revisioni hanno preso in considerazione il RoB nella sintesi e nell'interpretazione dei risultati. Quando la revisione include solo RCT, si consiglia di consultare l'Handbook Cochrane per determinare se gli autori della RS hanno effettuato una valutazione adeguata del RoB nei singoli studi. Questa sezione rappresenta invece una sfida ardua per il RoB nei NRSI.

Gli autori delle RS dovrebbero utilizzare un approccio sistematico per la valutazione del RoB, preferibilmente utilizzando uno strumento standardizzato. In caso negativo, bisognerebbe essere certi che lo strumento usato sia stato in grado di rilevare gravi limiti metodologici. Esistono diversi strumenti per la valutazione del RoB (per singoli studi), tra cui la *Newcastle Ottawa Scale*,

il SIGN e il *Mixed Methods Appraisal Tool* (MMAT). Lo strumento di valutazione più completo è quello recentemente proposto dalla Cochrane: il ROBINS-I. Tuttavia, proprio per la sua recente introduzione, lo strumento potrebbe non essere stato disponibile al momento della conduzione della RS.

Nello sviluppo di AMSTAR 2 è stato seguito l'esempio degli strumenti di analisi del RoB della Cochrane, sia per gli RCT ([http://handbook.cochrane.org/chapter\\_8/8\\_5\\_the\\_cochrane\\_collaborations\\_tool\\_for\\_assessing\\_risk\\_of\\_bias.htm](http://handbook.cochrane.org/chapter_8/8_5_the_cochrane_collaborations_tool_for_assessing_risk_of_bias.htm)) che per i NRSI ([www.riskofbias.info](http://www.riskofbias.info)), da cui sono stati ripresi gli item di valutazione. Qualunque sia lo strumento utilizzato dagli autori della RS, i valutatori dovrebbero considerare se soddisfa gli elementi aspetti elencati nell'item 9 dell'AMSTAR 2.

Considerato che AMSTAR 2 affronta solo i domini di bias più comuni, le raccomandazioni qui formulate non sono esaustive e una valutazione più approfondita del RoB richiede un contributo specialistico. Per giudicare le modalità con cui il RoB è stato valutato dagli autori delle RS, è necessario rivolgersi ad esperti dell'argomento della RS, se non inclusi nel gruppo di lavoro. Gli esperti dovrebbero fornire il loro contributo specifico per identificare eventuali fattori confondenti significativi e i bias di selezione e di misurazione rilevanti per la RS in esame. Inoltre, sarebbe opportuno approfondire le tecniche appropriate per l'aggiustamento dei risultati rispetto alle variabili confondenti.

Lo strumento ROBINS-I/ACROBAT-NRSI indica i seguenti domini di bias come maggiormente rilevanti per le RS che includono NRSI:

- **Variabili confondenti.** Si verificano quando gli effetti di due interventi o due esposizioni associate (es. fumo e consumo di alcol) non sono stati distinti nell'analisi, con conseguente potenziale attribuzione di un effetto alla variabile errata. Negli studi sull'efficacia degli interventi sanitari, la variabile confondente può anche essere correlata all'indicazione per il trattamento, ad esempio quando un farmaco viene somministrato preferenzialmente a pazienti con tassi di comorbidità più elevati rispetto a quelli che assumono il farmaco di confronto e le comorbidità sono associate all'outcome di interesse. Questo tipo di valutazioni sono tipicamente riportate nei dati baseline dei singoli studi. Il potenziale fattore confondente di base può essere gestito con diverse modalità: nel disegno di studio (es. matching per il *propensity score*), mediante aggiustamento (es. regressione logistica) e con altre tecniche quale *l'instrumental variable analysis* e l'inclusione di esposizioni "traccianti". Considerato che anche le tecniche più sofisticate non siano in grado di aggiustare tutte le variabili confondenti, le associazioni deboli, anche se statisticamente significative, dovrebbero essere sempre interpretate con cautela.
- **Bias di selezione del campione.** Si verifica quando il

campione viene selezionato in modo da distorcere direttamente la relazione reale tra esposizione ed esito ed è indipendente dalle variabili confondenti. Ad esempio, non si dovrebbe studiare l'associazione tra fumo e malattie cardiache reclutando soggetti da una clinica per smettere di fumare. La selezione dei soggetti con l'esposizione di interesse non dovrebbe essere correlata al loro esito. Allo stesso modo, la selezione dei soggetti con l'outcome di interesse non dovrebbe essere correlata alla loro esposizione. Anche le tempistiche di selezione possono essere importanti: se i soggetti hanno utilizzato un farmaco prima di essere arruolati (*prevalent users*), avranno un minor rischio di effetti avversi. Per tale ragione, gli studi farmaco-epidemiologici tendono a reclutare solo "new users" di farmaci, analogamente a quanto accade in un RCT dopo la randomizzazione. Altri bias sul timing di campionamento (*immortal time bias* e *inception bias*) possono risultare rilevanti. Per ulteriori informazioni consultare il documento ROBINS-I.

- **Bias nella misurazione di esposizioni ed esiti.** La misurazione di un'esposizione o di un trattamento può essere classificata in modo errato se non accuratamente registrata in tempo reale. In genere, i moderni studi farmaco-epidemiologici utilizzano prescrizioni o documenti di somministrazione come surrogato dell'assunzione del farmaco. Tuttavia, considerato che l'aderenza al trattamento non è mai del 100%, con questo metodo il consumo effettivo sarà erroneamente classificato. Se questo errore non è trascurabile, rappresenta un *bias toward the null*, ovvero la riduzione dell'effetto apparente (o apprezzabile) sull'outcome di salute. In alcuni settori di ricerca i ricercatori si affidano alla memoria (es. esposizione ultravioletta e relazione con melanoma), che può determinare diversi errori di classificazione. Per esempio, i genitori di un bambino deceduto per morte improvvisa - *sudden infant death syndrome* (SIDS) - possono dare particolare rilevanza al ricordo di qualsiasi farmaco somministrato al bambino prima dell'evento. La misurazione degli outcome può anche essere influenzata da classificazioni erranee (*misclassification*) che, se di tipo non-differenziale, generano un *bias toward the null*. Ad esempio, se l'ecografia degli arti inferiori viene eseguita frequentemente in donne con gambe gonfie e dolorose che assumono un contraccettivo orale, la selezione di persone da una clinica ecografica può distorcere i risultati degli studi che valutano l'associazione tra trombosi venosa profonda e contraccettivi orali.
- **Reporting selettivo di outcome e analisi.** Studi osservazionali di grandi dimensioni spesso analizzano database che registrano molti outcome in una popolazione definita. Se gli outcome non sono predefiniti (preferibilmente in un protocollo registrato), i ricercatori possono essere tentati di analizzare altri outcome e segnalare in modo selettivo quelli che sembrano essere differenti tra individui esposti e non esposti. Inoltre, esistono in genere

diversi metodi per analizzare un set di dati non randomizzato (inclusi, ad esempio, diversi modi di categorizzare l'intervento o diversi modelli di aggiustamento multi-variabile). Se il protocollo delle analisi non viene specificato prima dello studio, potrebbero essere effettuate analisi che suggeriscono una differenza statisticamente significativa che potrebbe non risultare in altre analisi. Gli autori di RS dovrebbero determinare se risultati e metodi di analisi negli studi inclusi sono stati o meno predefiniti. Il raggiungimento di questo obiettivo è indubbiamente facilitato dalla registrazione di un numero sempre più elevato di studi prima della loro conduzione.

### **Item 10: Gli autori riportano le fonti di finanziamento degli studi inclusi nella revisione?**

Numerosi studi hanno dimostrato che gli studi sponsorizzati, rispetto a quelli indipendenti, hanno maggiori probabilità di ottenere risultati che favoriscono il prodotto di uno sponsor. Gli autori delle RS dovrebbero riportare le fonti di finanziamento per ogni studio incluso nella RS o che tali informazioni non sono incluse nei report degli studi. La disponibilità di queste informazioni permette di analizzare separatamente i risultati degli studi sponsorizzati e quelli indipendenti.

### **Item 11: Se è stata condotta una meta-analisi, gli autori hanno utilizzato metodi appropriati per la combinazione statistica dei risultati?**

Gli autori di RS dovrebbero dichiarare esplicitamente nel protocollo della RS i principi su cui hanno basato la decisione di eseguire la meta-analisi dei dati degli studi inclusi. Ad esempio, il desiderio di ottenere una stima cumulativa degli effetti (es. da un numero di studi comparabili ma sottodimensionati), oppure il fatto che gli studi possono essere combinati perché hanno caratteristiche comparabili (in termini di controlli e interventi).

Se la meta-analisi è stata considerata appropriata, gli autori dovrebbero spiegare la decisione di utilizzare modelli a effetto fisso o effetto random nel caso di RS di RCT e riportare i metodi per indagare l'eterogeneità.

Nei NRSI il campione di popolazione in studio può essere molto variabile da piccole coorti (di decine o centinaia di partecipanti) a studi di centinaia di migliaia di individui e migliaia di eventi. Se questi risultati vengono combinati con quelli di RCT più piccoli, le stime aggregate dell'effetto saranno dominate dai dati dei NRSI. Inoltre, i risultati di NRSI possono essere influenzati da un'ampia gamma di bias (vedi sopra), determinando una distorsione delle stime cumulative, seppure precise.

Gli autori delle RS dovrebbero riportare le stime cumulative degli effetti in modo separato per i diversi tipi di studio. Nel caso di NRSI, l'aggregazione può determinare una stima dell'effetto molto precisa e statisticamente significativa, ma parziale. Tuttavia, l'intervallo di confidenza è calcolato partendo dal presupposto che non vi sia

alcun bias (es. le stime sono accurate come se fossero ottenute da un RCT di elevata qualità con lo stesso numero di partecipanti). È raro che un NRSI abbia un RoB basso come un RCT di elevata qualità e gli intervalli di confidenza per i NRSI (e le stime aggregate basate su essi) devono essere considerati con cautela. Questo problema è importante quando si considera il diverso RoB e l'incertezza sul RoB nei NRSI.

L'eterogeneità è un aspetto rilevante in ogni meta-analisi, ma è particolarmente importante in quelle di NRSI, a causa dei differenti metodi utilizzati nei diversi studi. Oltre alle solite fonti di eterogeneità [confronti differenti, variazioni nel rischio basale degli outcome o altre caratteristiche della popolazione dello studio, interventi diversi (es. relazione dose-risposta, contesto/setting, competenza professionale) e diverse definizioni degli outcome], è importante considerare l'eterogeneità in termini di partecipanti, completezza dei dati e metodi di gestione e analisi dei dati. L'aggiustamento statistico degli effetti dell'intervento per i fattori confondenti può portare a stime molto diverse dalla stima non aggiustata derivata dai dati grezzi.

In generale, quando si combinano i risultati di NRSI, gli autori dovrebbero aggregare le stime di efficacia aggiustate, non i dati grezzi. Se lo fanno, dovrebbe esserci una chiara motivazione. Tuttavia, è molto probabile che studi diversi riportino gli effetti del trattamento aggiustati per diversi gruppi di covariate (o covariate misurate e adattate in modi diversi); questa diversità rappresenta un'altra fonte di potenziale eterogeneità.

### **Item 12: Se è stata condotta una meta-analisi, gli autori analizzano il potenziale impatto del rischio di bias dei singoli studi nei risultati della meta-analisi o nelle altre sintesi delle evidenze?**

Nei casi in cui gli autori di revisioni abbiano scelto di includere solo RCT di elevata qualità, non dovrebbero esserci dubbi rilevanti sul potenziale impatto dei bias sui risultati. Tuttavia, se sono stati inclusi RCT di variabile qualità, dovrebbero valutarne l'impatto mediante analisi di regressione o stimando le dimensioni aggregate dell'effetto solo negli studi a basso RoB. Nel caso dei NRSI, dovrebbero stimare le dimensioni aggregate dell'effetto includendo solo gli studi a RoB basso o moderato e/o solo quelli a basso RoB (se esistenti). Se non sono state eseguite meta-analisi (o altre tecniche statistiche per sintetizzare i dati come l'analisi di regressione), gli autori dovrebbero comunque commentare il probabile impatto del RoB sui risultati degli studi individuali.

### **Item 13: Gli autori tengono in considerazione il rischio di bias nei singoli studi quando interpretano/discutono i risultati della revisione?**

Anche se non è stata condotta una meta-analisi, gli autori della RS dovrebbero discutere l'impatto del RoB nell'in-

interpretazione dei risultati. Questo è particolarmente importante quando le revisioni includono RCT con RoB variabile e/o NRSI. La discussione non dovrebbe essere limitata all'impatto del RoB sulle stime aggregate, ma dovrebbe anche considerare se questo possa spiegare le differenze tra i risultati dei singoli studi. Gli autori dovrebbero prendere in considerazione esplicitamente il RoB se formulano raccomandazioni che potrebbero avere un impatto sulla pratica clinica o sulle politiche sanitarie.

**Item 14: Gli autori spiegano e discutono in maniera soddisfacente ogni eterogeneità osservata nei risultati della revisione?**

Ci sono molte potenziali cause di eterogeneità nei risultati di NRSI rispetto ai RCT, tra cui molti fattori considerati in questo strumento, come la diversità di disegni di studio, metodi di analisi, popolazioni e intensità dell'intervento (dosaggi nel caso di farmaci). Sia gli elementi del PICO che i domini di bias elencati nell'item 9 dovrebbero essere considerati come rilevanti fonti potenziali di eterogeneità nei risultati. Gli autori delle RS dovrebbero prendere in considerazione questo aspetto e discuterne l'impatto sui risultati e sulle eventuali raccomandazioni.

**Item 15: Se è stata effettuata una meta-analisi, gli autori hanno esplorato adeguatamente il bias di pubblicazione e discusso il potenziale impatto sui risultati della revisione?**

Questo aspetto è molto rilevante, ma può essere difficile gestirlo per gli autori e gli esperti di RS. In genere, vengono utilizzati test statistici o layout grafici la cui positività indica la presenza di bias di pubblicazione. Tuttavia, i test negativi non sono una garanzia dell'assenza di bias in ragione della loro scarsa sensibilità. L'importanza del bias di pubblicazione dipende dal contesto e dal setting: ad esempio, vari studi finanziati dall'industria (es. farmaci, dispositivi) potrebbero essere maggiormente influenzati dal bias di pubblicazione, rispetto a studi simili non sponsorizzati. Il punto fondamentale è verificare se gli autori della RS hanno fatto del loro meglio per identificare il bias di pubblicazione tramite ricerche bibliografiche più approfondite e analitiche (se opportuno e in relazione al setting), se hanno mostrato una consapevolezza del probabile impatto del bias di pubblicazione nell'interpretazione e discussione dei risultati e se hanno eseguito un'analisi di sensibilità per stimare quanti studi "nulli" mancanti sarebbero necessari per invalidare i risultati ottenuti.

**Item 16: Gli autori hanno riportato ogni fonte potenziale di conflitto di interessi, includendo anche eventuali finanziamenti ricevuti per condurre la revisione?**

Come sopra riportato, gli studi sponsorizzati, rispetto a quelli indipendenti, possono generare risultati più favorevoli all'intervento. Lo stesso concetto si applica alle RS e gli autori dovrebbero dichiarare le fonti di finanziamento

diretto, come generalmente richiesto dalle riviste. Tuttavia per la valutazione dei conflitti di interesse degli autori di una RS questo non è sufficiente e dovrebbero essere segnalati eventuali altre relazioni finanziarie e non. La RS potrebbe essere finanziata in modo indipendente, ma gli autori potrebbero avere legami con le aziende che producono prodotti inclusi nella RS. I conflitti di interesse professionali sono rilevanti, ma difficili da distinguere perché vengono segnalati raramente. Quando i ricercatori hanno investito la loro carriera in un determinato settore, una RS in conflitto con le loro radicate convinzioni può essere un problema. Potenziali conflitti di interesse di questo tipo, difficili da identificare, possono essere dedotti dal fatto che gli autori della RS potrebbero aver pubblicato estensivamente nel settore di ricerca in esame e i loro studi potrebbero essere inclusi nella RS. Anche se gli effetti dei conflitti di interesse potrebbero manifestarsi negli altri domini di bias, riteniamo che questo item debba sempre essere valutato separatamente.