

# **ROBIS: Uno strumento per valutare il rischio di bias nelle revisioni sistematiche**

## **Guida all'uso di ROBIS**

**PENNY WHITING** (KLEIJNEN SYSTEMATIC REVIEWS LTD, UNIVERSITÀ DI BRISTOL),

**JELENA SAVOVIĆ** (UNIVERSITÀ DI BRISTOL)

**JULIAN HIGGINS** (UNIVERSITÀ DI BRISTOL)

**DEBORAH CALDWELL** (UNIVERSITÀ DI BRISTOL)

**BARNABY REEVES** (UNIVERSITÀ DI BRISTOL)

**BEVERLEY SHEA** (UNIVERSITÀ DI OTTAWA)

**PHILIPPA DAVIES** (UNIVERSITÀ DI BRISTOL)

**JOS KLEIJNEN** (KLEIJNEN SYSTEMATIC REVIEWS LTD, UNIVERSITÀ DI MAASTRICHT)

**RACHEL CHURCHILL** (UNIVERSITÀ DI BRISTOL)

Traduzione in Italiano a cura di Laura Amato<sup>1</sup>, Federico Bertolini<sup>2</sup>, Silvia Minozzi<sup>1</sup>, Rosella Saulle<sup>1</sup>, Simona Vecchi<sup>1</sup>

<sup>1</sup>Dipartimento di Epidemiologia del SSR-ASL Roma1 Regione Lazio

<sup>2</sup>Dipartimento di Neuroscienze, Biomedicina e Scienze del Movimento, sezione di Psichiatria, Università di Verona

# Indice

## **Lo strumento ROBIS**

### **Definizione di rischio di bias**

### **Destinatari**

### **Fase 1: Valutare la rilevanza del quesito di ricerca (opzionale)**

- Esempio di valutazione della rilevanza

### **Fase 2: Identificare i punti critici del processo di revisione**

- Dominio 1: criteri di eleggibilità degli studi
  - Background
  - Esempi di valutazione
- Dominio 2: Identificazione e selezione degli studi
  - Background
  - Esempi di valutazione
- Dominio 3: Raccolta dati e valutazione degli studi
  - Background
  - Esempi di valutazione
- Dominio 4: Sintesi e risultati
  - Background
  - Esempi di valutazione

### **Fase 3: Valutare il rischio di bias della revisione**

- Esempi di valutazione

### **Problemi pratici nell'uso di ROBIS**

### **Bibliografia**

### **Appendice: Glossario**

## Lo strumento ROBIS

Lo strumento si completa in tre fasi: 1) valutare la rilevanza del quesito di ricerca (opzionale), 2) identificare i punti critici del processo di revisione e 3) valutare il rischio complessivo di bias (distorsioni sistematiche) della revisione. Sono state incluse domande chiave per aiutare a valutare specifici dubbi sui potenziali bias della revisione. Rispondere a queste domande chiave aiuta il valutatore a giudicare il rischio complessivo di bias. In questo documento consideriamo singolarmente queste tre fasi e forniamo indicazioni per la fase di valutazione complessiva e per ogni domanda chiave<sup>1-4</sup>.

## Definizione di rischio di bias

ROBIS valuta sia il rischio di bias (distorsione sistematica) di una revisione che (dove appropriato) la rilevanza del quesito della revisione per rispondere alla domanda in oggetto. Nello specifico, ROBIS affronta 1) il grado in cui i metodi utilizzati per lo sviluppo della revisione riescano a ridurre al minimo il rischio di distorsioni nella sintesi dei risultati e nelle conclusioni e 2) la misura in cui la domanda di ricerca affrontata dalla revisione corrisponde alla domanda di ricerca di chi deve utilizzarla (per es., un autore di un overview o un produttore di linee-guida).

I bias possono verificarsi se difetti sistematici o limitazioni nella progettazione, nella conduzione o nell'analisi di una revisione ne distorcono i risultati.

Le prove ottenute da una revisione possono essere di rilevanza limitata se il quesito della revisione non coincide con il quesito dell'overview/linea-guida.

## Destinatari

ROBIS è stato sviluppato avendo presenti tre specifici gruppi di possibili utilizzatori:

- autori di overview di revisioni sistematiche;
- produttori di linee-guida;
- autori di revisioni che vogliono valutare il rischio di bias delle loro revisioni dopo averle completate o minimizzare il rischio di bias nel progettare i metodi della revisione nella fase di stesura del protocollo.

Può essere utile per tutti coloro che vogliono valutare il rischio di bias di una revisione sistematica, non è stato progettato specificamente per altri destinatari. Altri potenziali utilizzatori di ROBIS includono le organizzazioni che supportano i processi decisionali (per es., NICE, IQWiG), i clinici interessati alla medicina basata sulle prove, gli editori di giornali, i revisori di articoli.

## Fase 1: Valutare la rilevanza del quesito di ricerca (opzionale)

I valutatori per prima cosa riportano il quesito cui cercano di dare risposta (per es., nella loro overview o linea-guida): abbiamo chiamato questo quesito "domanda target". Per revisioni sull'efficacia degli interventi, viene loro richiesto di definire tale domanda in termini di PICO (partecipanti, intervento, confronto ed esiti). Per revisioni che rispondono a differenti quesiti (per es., test diagnostici, fattori prognostici, eziologia, o modelli predittivi), sono previste altre categorie, se appropriato (vedi Box). I valutatori completano il PICO o l'equivalente per la revisione sistematica da valutare usando ROBIS e quindi devono chiedersi se le due domande (la domanda target e la domanda della revisione sistematica) coincidono. Se una o più delle categorie (PICO o equivalente) non coincidono, allora devono dare come valutazione "No". Se vi è una coincidenza parziale, allora possono giudicare come "Parziale". Per esempio se la domanda target si riferisce ad adulti e la revisione sistematica include solo partecipanti >60 anni. Se si sta solo valutando una revisione e non vi è una domanda target, allora questa fase di ROBIS può essere omessa.

Box. Esempi di quesiti e relativi PICO per diverse tipologie di revisione sistematica.		
Tipo di revisione	PICO relativo	Esempio
Intervento <sup>5</sup>	Pazienti/popolazione/i Intervento/i Confronto/i Esito/i	Adulti con epatite C cronica Terapia antivirale triplice con interferone pegilato Terapia antivirale duplice Risposta virologica sostenuta
Eziologia <sup>6</sup>	Pazienti/popolazione/i Esposizione/i e confronto/i Esito/i	Adulti Indice di massa corporea Cancro del colon-retto
Diagnosi <sup>7</sup>	Pazienti Test indice Standard di riferimento Condizione in studio	Adulti con sintomi che suggeriscono un cancro rettale Endoscopia ultrasonica Istologia chirurgica Stadio TO del cancro rettale
Prognosi <sup>8</sup>	Pazienti  Esito previsto  Uso previsto del modello  Momento temporale previsto	Donne in gravidanza, con o senza limitazione della crescita fetale, nessuna prova di rottura prematura delle membrane, nessuna evidenza di anomalie congenite o strutturali Esito avverso della gravidanza (peso alla nascita basso o alto, morte neonatale, mortalità perinatale) Per prevedere l'effetto delle indagini con ultrasuoni del liquido amniotico sull'esito della gravidanza Gravidanza avanzata (>37 <sup>a</sup> settimana di gestazione)

#### ESEMPIO DI VALUTAZIONE DELLA RILEVANZA

Per una revisione diagnostica:

Categoria	Condizione in studio	Revisione che si sta valutando <sup>7</sup>
Paziente/i:	Adulti con sintomi che suggeriscono un cancro rettale	Non chiaro
Test indice:	Endoscopia ultrasonica	Endoscopia ultrasonica
Standard di riferimento:	Istologia chirurgica	Chirurgia
Condizione in studio:	Cancro del retto in fase iniziale (T0)	Cancro del retto in fase iniziale (T0)

Valutazione della rilevanza		Spiegazione
La domanda di ricerca della revisione sistematica e la domanda cui si sta tentando di rispondere (per es., nella vostra overview o linea-guida) coincidono?	Non chiaro	La domanda coincide per quanto riguarda il test indice, lo standard di riferimento e la condizione in studio, ma nella revisione non viene definito il tipo di popolazione e quindi non è chiaro se il quesito della revisione coincida con il quesito target.

#### Fase 2: Identificare i punti critici del processo di revisione

La fase 2 ha l'obiettivo di identificare le aree dove possono essere stati introdotti bias nella revisione. Questa fase prevede la valutazione di quattro domini relativi ai processi chiave della revisione: eleggibilità degli studi; identificazione e selezione degli studi; raccolta dei dati e valutazione degli studi; sintesi e risultati. Questa fase di ROBIS identifica le aree potenzialmente critiche per il giudizio finale complessivo di rischio di bias. Ogni dominio comprende tre sezioni: informazioni utilizzate per giustificare il giudizio dato, domande chiave e giudizio relativo al rischio di bias. I domini dovrebbero essere considerati in modo consequenziale e non valutati come singole unità. Per esempio, questo significa che, quando si valuta il dominio 2 (identificazione e

selezione degli studi), il valutatore dovrebbe considerare le strategie di ricerca in relazione al quesito di ricerca specificato nel dominio 1.

Alle domande chiave si può rispondere “Sì”, “Probabilmente sì”, “No”, “Probabilmente no” e “Nessuna informazione”, dove il “Sì” indica che vi sono poche preoccupazioni. Il successivo livello di preoccupazione circa il rischio di bias associato a ciascun dominio è poi valutato come “Basso”, “Alto” o “Non chiaro”. Questo tipo di valutazione è stato preferito a una valutazione di “Sì”, “No” o “Non chiaro”, perché una tale valutazione avrebbe significato che “Sì” era una cosa negativa e questo sarebbe stato in contrasto con la terminologia utilizzata in tutte le domande chiave e avrebbe creato confusione. Inoltre, questa valutazione era coerente con la valutazione del rischio di bias nel dominio 3. Se le risposte a tutte le domande chiave relative a un certo dominio sono “Sì” o “Probabilmente sì” allora il livello di preoccupazione può essere giudicato come basso. Se a una domanda chiave si è risposto “No” o “Probabilmente no”, esiste una preoccupazione che vi sia un alto rischio di bias. La categoria “Nessuna informazione” dovrebbe essere utilizzata solo quando vi sono dati insufficienti nella revisione per permettere di formulare un giudizio. Registrando le informazioni utilizzate per raggiungere il giudizio (“motivazioni del giudizio”) aspiriamo a rendere trasparente la valutazione e, se necessario, a facilitare la discussione tra coloro che indipendentemente stanno completando la valutazione. È probabile che coloro che utilizzano ROBIS, per completare la valutazione, abbiano bisogno sia di conoscere la materia oggetto della revisione sia di una competenza metodologica.

## **FASE 2. DOMINIO 1: CRITERI DI ELEGGIBILITÀ DEGLI STUDI**

### *Background*

Il primo dominio ha lo scopo di valutare se vengono definiti a priori in modo chiaro e appropriato i criteri di eleggibilità degli studi utili per rispondere al quesito della revisione. Una revisione sistematica dovrebbe cominciare con una domanda o obiettivo chiaro e focalizzato<sup>9</sup>. Questo dovrebbe riflettersi nella definizione a priori dei criteri utilizzati per decidere la eleggibilità degli studi da includere nella revisione. La definizione a priori mira a garantire che le decisioni su quali studi includere siano prese in modo coerente piuttosto che sulla base di una conoscenza delle caratteristiche e dei risultati degli studi stessi. Normalmente è possibile valutare se i criteri di eleggibilità siano stati correttamente definiti a priori (e rispettati nella revisione) solo se è disponibile un protocollo o un documento registrativo precedente alla conduzione e al reporting della revisione. Quando un tale documento non è disponibile, il valutatore dovrà basare il suo giudizio relativamente a questo dominio dal modo in cui sono riportati i risultati, ma sarà difficile sapere se questi criteri erano stati stabiliti in precedenza e hanno guidato gli autori nel condurre la revisione o se invece sono stati decisi o modificati durante il processo di conduzione della revisione stessa.

I criteri di eleggibilità combinano gli aspetti relativi al quesito di ricerca della revisione con ulteriori dettagli relativi ai tipi di studi che la revisione prenderà in considerazione, incluso il disegno dello studio. Idealmente, le caratteristiche degli studi eleggibili saranno determinate in base alla popolazione/partecipanti coinvolti, all'intervento/esposizione valutati e ai confronti effettuati. Le informazioni relative alle caratteristiche degli studi dovrebbero essere abbastanza specifiche, evitando per quanto possibile qualsiasi ambiguità che avrebbe potuto consentire agli autori di imporre, durante il processo di revisione, giudizi post hoc sull'eleggibilità dello studio. Gli autori delle revisioni spesso definiscono dei limiti relativi ai criteri di eleggibilità, che possono o meno essere giudicati come appropriati. Per esempio, potrebbero limitare l'inclusione a studi con particolari caratteristiche, tipo a studi che riportano particolari esiti di interesse, a quelli che soddisfano specifici criteri di qualità o a quelli che sono stati condotti entro un certo intervallo temporale. Inoltre, possono introdurre limiti rispetto alle fonti di informazione che vogliono includere, come per esempio includere solo studi pubblicati come articoli completi o quelli disponibili in inglese. Nei casi in cui vengano applicati limiti all'inclusione degli studi, questi devono essere indicati in anticipo e deve essere fornita una giustificazione per tale limitazione. In alcune circostanze, i limiti possono introdurre bias nella selezione degli studi che possono essere inclusi.

La tabella 1 riassume le domande chiave incluse nel Dominio 1 insieme a una guida su come valutare ogni singola domanda. La tabella 2 fornisce una panoramica delle classificazioni per il dominio 1. Esempi di valutazioni delle revisioni in cui gli esaminatori avevano molte o poche preoccupazioni relative a questo dominio sono forniti nelle tabelle 3 e 4.

**Tabella 1.** Domande chiave per il Dominio 1 con guida per rispondere alle singole domande

Domanda chiave	Guida alla valutazione
1.1 La revisione aderisce agli obiettivi e ai criteri di eleggibilità predefiniti?	Una revisione sistematica dovrebbe cominciare con una domanda o un obiettivo chiaro e focalizzato, che dovrebbe riflettersi nei criteri utilizzati per decidere la eleggibilità degli studi da includere nella revisione. I dettagli dovrebbero essere specificati a priori nel protocollo della revisione a seconda del tipo di revisione che si intende condurre, ma generalmente dovrebbero includere il tipo di disegno di studio, il tipo di partecipanti e il tipo di interventi/esposizioni che si considerano per l'inclusione. Se il tipo di esiti considerati negli studi primari entreranno a far parte dei criteri utilizzati per l'eleggibilità, questo va chiaramente specificato. Qualsiasi tipo di criterio di esclusione va predefinito. Se esiste un protocollo della revisione in cui queste informazioni sono disponibili, la risposta a questa domanda sarà "Sì". Laddove non sia disponibile alcun protocollo ma vengano fornite informazioni su obiettivi predefiniti e criteri di eleggibilità dettagliati, e vi siano buone ragioni per ritenere che questi siano stati predefiniti in anticipo e rispettati durante tutta la revisione, i valutatori possono prendere in considerazione la risposta "Probabilmente sì". Qualsiasi modifica post hoc relativa ai criteri di eleggibilità o agli esiti deve tenere fede agli obiettivi della revisione ed essere debitamente giustificata e documentata. In assenza di un protocollo pre-pubblicato, nel caso in cui le informazioni sugli obiettivi predefiniti e i criteri di eleggibilità sono disponibili solo post hoc nella revisione pubblicata, a meno che non vi siano motivi per ritenere che tali dettagli siano stati predefiniti in anticipo e rispettati dall'inizio della revisione, a questa domanda si dovrebbe rispondere "Probabilmente no". Se mancano tutti o alcuni di questi dettagli, è necessario rispondere a questa domanda "No".
1.2 I criteri di eleggibilità erano appropriati rispetto al quesito della revisione?	I criteri di eleggibilità dovrebbero essere consequenziali rispetto al quesito della revisione e dovrebbero fornire sufficienti dettagli per permettere di valutare se gli studi inclusi sono appropriati rispetto al quesito. Le informazioni necessarie variano a seconda dell'argomento trattato nella revisione. Per esempio, per giudicare l'appropriatezza, il valutatore potrebbe aver bisogno di una descrizione chiara della popolazione in termini di fascia di età e diagnosi dei partecipanti allo studio, del contesto in cui è stato condotto lo studio, della dose di un farmaco o della frequenza di un'esposizione. Per rispondere a questa domanda è probabile che il valutatore debba avere una qualche conoscenza del tema trattato.
1.3 I criteri di eleggibilità erano privi di ambiguità?	Devono essere fornite informazioni specifiche sulle caratteristiche degli studi eleggibili, evitando per quanto possibile ogni ambiguità sul tipo di studio, popolazione, interventi, confronti ed esiti. I criteri devono essere sufficientemente dettagliati in modo da essere replicabili. Spesso nelle revisioni sistematiche mancano dettagli importanti sui criteri di eleggibilità degli studi. Per esempio, possono non esserci dettagli sulle modalità relative su come è stata fatta la diagnosi dei partecipanti allo studio. La diagnosi può essere effettuata utilizzando metodi diversi, alcuni dei quali potrebbero essere più validi o precisi di altri. Al fine di evitare l'introduzione di potenziali bias, gli autori delle revisioni dovrebbero aver deciso in anticipo quali metodi diagnostici sono appropriati per rispondere al quesito della loro revisione. Allo stesso modo, devono essere forniti dettagli specifici su interventi/esposizioni e confronti, includendo informazioni sui dosaggi dei farmaci, sulla frequenza di somministrazione, sui trattamenti concomitanti e così via. Per rispondere a questa domanda è probabile che il valutatore debba avere una qualche conoscenza del tema trattato, ma laddove permangano dubbi specifici riguardo ai criteri di eleggibilità dichiarati, solitamente è possibile formulare giudizi "No" o "Probabilmente no".
1.4 I limiti relativi ai criteri di eleggibilità basati sulle caratteristiche degli studi erano appropriati?	Qualsiasi limite applicato sulla base delle caratteristiche dello studio deve essere chiaramente descritto e deve esserne fornita una spiegazione razionale. Questi dettagli consentiranno ai valutatori di giudicare se tali limiti siano appropriati. Esempi possono riguardare il disegno di studio, la data di pubblicazione di uno studio, il luogo dove lo studio è stato condotto, alcune caratteristiche della qualità dello studio e il tipo di misure di esito disponibili. Questa domanda differisce dalla precedente che riguardava l'appropriatezza dei criteri di eleggibilità, rispetto al quesito della revisione. Quando sono disponibili sufficienti informazioni e il valutatore considera i limiti appropriati, a questa domanda si può rispondere "Sì" o "Probabilmente sì". Se i limiti relativi alle caratteristiche di uno studio non sono spiegati e non vi sono informazioni sufficienti per valutare la loro appropriatezza, a questa domanda si può rispondere "Probabilmente no" o "No". Quando i criteri di inclusione sono sufficientemente dettagliati e non è riportato alcun limite relativo alle caratteristiche degli studi, si può presumere che non ve ne siano e quindi rispondere "Sì" a questa domanda.
1.5 Eventuali limiti nei criteri di eleggibilità basati sulle fonti di informazione erano appropriati?	Qualsiasi limite relativo alle fonti di informazione deve essere chiaramente descritto e deve esserne fornita una spiegazione razionale. Questi dettagli consentiranno ai valutatori di giudicare l'appropriatezza di tali limiti. Esempi possono essere lo stato della pubblicazione o il suo formato, la lingua di pubblicazione e la disponibilità di dati. Questa domanda differisce da quella del Dominio 2 che è relativa alle strategie di ricerca. Quando i criteri di eleggibilità sono sufficientemente dettagliati e non è riportato alcun limite relativo alle fonti, si può presumere che non ve ne siano e quindi rispondere "Sì" a questa domanda.

Tabella 2. Preoccupazioni relative ai criteri di eleggibilità degli studi.	
Poche preoccupazioni	Sono stati fatti sforzi considerevoli per specificare chiaramente il quesito e gli obiettivi di revisione e per specificare e giustificare in modo appropriato e dettagliato i criteri di inclusione che sono stati rispettati durante la conduzione della revisione.
Molte preoccupazioni	È probabile che, a causa della mancanza di obiettivi e criteri di eleggibilità prestabiliti o a causa di limiti inappropriati, siano stati esclusi dalla revisione studi che sarebbero stati importanti e pertinenti per rispondere al quesito della revisione oppure che siano stati inclusi studi non appropriati.
Non chiaro	Non vi sono sufficienti informazioni per emettere un giudizio relativo a questo rischio di bias.

### Esempi di valutazione

#### a. Revisione giudicata ad alto rischio per quanto riguarda i criteri di eleggibilità degli studi<sup>7</sup>

Descrizione dal testo:

##### Abstract – Scopo

*“Valutare l’accuratezza di EUS nella stadiazione T del cancro rettale precoce”*

##### Criteri di Selezione degli studi

*“Sono stati selezionati solo gli studi in cui EUS era confermata dalla istologia chirurgica. Il criterio EUS utilizzato per T0 era tumore confinato alla mucosa. Da questo gruppo di studi sono stati inclusi solo gli studi per i quali è stato possibile costruire una tabella 2x2 per valori veri-positivi, falsi-positivi, falsi negativi e veri-negativi”*

Tabella 3. Esempio di valutazione per i criteri di eleggibilità degli studi giudicati ad alto rischio		
Domanda chiave	Valutazione	Motivazione
1.1 La revisione ha aderito a obiettivi e criteri di eleggibilità predefiniti?	Nessuna informazione	Non vi erano prove che gli obiettivi e i criteri di eleggibilità fossero stati predefiniti.
1.2 I criteri di eleggibilità erano appropriati rispetto al quesito della revisione?	Probabilmente no	I criteri di eleggibilità riportati erano appropriati ma mancavano alcuni dettagli, in particolare quelli relativi al tipo di popolazione. Quindi è stato impossibile giudicare l’appropriatezza dei criteri di eleggibilità.
1.3 I criteri di eleggibilità erano privi di ambiguità?	Probabilmente no	Non vi erano sufficienti dettagli sui criteri di eleggibilità. In particolare non veniva specificato quale tipo di popolazione o quale disegno di studio era considerato eleggibile.
1.4 I limiti relativi ai criteri di eleggibilità basati sulle caratteristiche degli studi erano appropriati?	Nessuna informazione	Non venivano forniti dettagli su limiti legati alle caratteristiche degli studi. In generale vi erano insufficienti dettagli sulla eleggibilità.
1.5 I limiti relativi all’eleggibilità basati sulle fonti di informazione erano appropriati?	Probabilmente sì	La revisione includeva solo studi per cui poteva essere costruita una tabella 2x2. Sebbene per questo non venisse fornita alcuna giustificazione, si tratta di una procedura solitamente utilizzata nelle revisioni di accuratezza diagnostica (DTA) e quindi questa decisione sembra essere ragionevole.
Preoccupazioni relative ai criteri di eleggibilità degli studi	Molte preoccupazioni (Alto livello di rischio di bias)	
Motivazione delle preoccupazioni	I criteri di eleggibilità erano insufficientemente dettagliati per permettere di giudicare se i criteri di inclusione degli studi erano o meno appropriati; in particolare non vi erano informazioni relative ai partecipanti eleggibili. Infine non era chiaro se i criteri erano stati prespecificati o adattati post hoc.	

**b. Revisione giudicata a basso rischio per quanto riguarda I criteri di eleggibilità degli studi<sup>7</sup>**

Descrizione dal testo:

**Obiettivo**

*“Valutare l’associazione e il valore predittivo delle indagini con ultrasuoni del volume del liquido amniotico rispetto ad esiti avversi della gravidanza.”*

**Metodi – Selezione degli studi**

*“1. Popolazione Donne in gravidanza, con o senza limitazione della crescita fetale, nessuna evidenza di rottura prematura delle membrane, nessuna evidenza di anomalie congenite o strutturali.*

*2. Test indice. Qualsiasi misura del liquido amniotico riportata dagli autori, compresi AFI, volume del liquido amniotico e tasca massima più profonda. È stata accettata qualsiasi soglia utilizzata e riportata dagli autori per definire basso o alto il liquido amniotico.*

*3. Esito. Qualunque standard di riferimento che consideri la compromissione del benessere fetale o neonatale; tra cui: pH anormale del cordone ombelicale alla nascita, punteggi Apgar, morte perinatale e esiti compositi come esito avverso perinatale. Qualsiasi standard di riferimento che consideri la bassa crescita fetale o feto piccolo per età gestazionale: peso alla nascita <10 °, <5 °, <3 ° percentile, soglie di peso assoluto alla nascita, indice ponderale.*

*4. Disegno di studio. Studi osservazionali in cui i risultati del test in studio sono confrontati con i risultati ottenuti da un test standard di riferimento, consentendo la produzione di una tabella 2x2 per calcolare indici di associazione e accuratezza del test per ciascuna soglia disponibile. Sono state escluse le serie di casi con dieci o meno studi e gli studi caso-controllo.”*

**Estrazione Dati**

*“Tutti gli studi dovevano dichiarare che hanno escluso la rottura delle membrane e anomalie congenite/strutturali dovute all’associazione di anomalie del tratto renale/urinario e anomalie cariotipiche con anomalie del volume del liquor.”*

**Tabella 4.** Esempio di valutazione per i criteri di eleggibilità degli studi giudicati a basso rischio.

Domanda chiave	Valutazione	Motivazione
1.1 La revisione ha aderito a obiettivi e criteri di eleggibilità predefiniti?	Sì	Gli autori hanno specificato chiaramente nell' <i>Abstract</i> e nel <i>Background</i> dell'articolo che gli obiettivi erano valutare l'associazione e il valore predittivo delle misurazioni ecografiche del volume del liquido amniotico rispetto a un esito negativo della gravidanza. Un protocollo separato forniva informazioni sul quesito specifico della revisione "Qual è l'accuratezza del volume del liquido amniotico per prevedere danni fetali/neonatali e una bassa crescita fetale?". Gli autori hanno fornito dettagli sui criteri di eleggibilità nella sezione <i>Selezione degli studi</i> organizzata per tipo di popolazione, test indice, esiti e disegno di studio.
1.2 I criteri di eleggibilità erano appropriati rispetto al quesito della revisione?	Sì	Il quesito della revisione indicava la necessità di individuare studi che valutavano l'accuratezza della misura del volume del fluido amniotico nel predire danni fetali/neonatali o bassa crescita fetale. I dettagli forniti nell'articolo sugli studi da considerare eleggibili per l'inclusione sembrano appropriati rispetto al quesito della revisione.
1.3 I criteri di eleggibilità erano privi di ambiguità?	Probabilmente sì	Il tipo di disegno di studio era chiaramente indicato. Sono stati inclusi studi osservazionali. Gli autori hanno escluso le serie di casi (con 10 o meno studi) e gli studi caso-controllo. Erano fornite sufficienti informazioni circa la popolazione di interesse (donne in gravidanza, con o senza limitazione della crescita fetale, nessuna evidenza di rottura prematura delle membrane, nessuna evidenza di anomalie congenite o strutturali), il test indice (qualsiasi misura del liquido amniotico riportata dagli autori, compresi AFI, volume del liquido amniotico e tasca massima più profonda), esiti (qualunque standard di riferimento che consideri la compromissione del benessere fetale o neonatale; venivano riportati vari esempi; qualsiasi standard di riferimento che consideri la bassa crescita fetale o feto piccolo per età gestazionale; di nuovo venivano riportati vari esempi).



segue **Tabella 4.** Esempio di valutazione per i criteri di eleggibilità degli studi giudicati a basso rischio.

Domanda chiave	Valutazione	Motivazione
1.4 I limiti relativi ai criteri di eleggibilità basati sulle caratteristiche degli studi erano appropriati?	Probabilmente sì	Erano chiaramente descritti i limiti relativi al disegno e alle dimensioni degli studi. Ulteriori limiti erano descritti nella sezione <i>Estrazione dati</i> . I limiti sembrano essere appropriati sebbene di essi non venga fornita alcuna giustificazione.
1.5 I limiti relativi all'eleggibilità basati sulle fonti di informazione erano appropriati?	Probabilmente sì	Non vi era alcun limite rispetto al linguaggio. Alcuni limiti potrebbero basarsi sulla disponibilità dei dati, sebbene non sia fornita alcuna giustificazione. Non sono stati descritti altri limiti relativi alle fonti di informazione. Quindi a questa domanda si può rispondere "Sì" o "Probabilmente sì".
Preoccupazioni relative ai criteri di eleggibilità degli studi	Poche preoccupazioni (Basso livello di rischio di bias)	
Motivazione delle preoccupazioni	A tutte le domande chiave si è risposto "Sì" o "Probabilmente sì", per cui non sono sorti dubbi sui criteri di eleggibilità. È stato fatto un considerevole sforzo per specificare chiaramente il quesito della revisione e gli obiettivi e per specificare in anticipo e giustificare definiti e appropriati criteri di eleggibilità.	

## FASE 2. DOMINIO 2: IDENTIFICAZIONE E SELEZIONE DEGLI STUDI

### Background

Questo dominio ha lo scopo di valutare se nella revisione siano stati esclusi studi che avrebbero soddisfatto i criteri di inclusione. Una ricerca esaustiva della letteratura mirata a individuare quanti più studi possibile è una componente chiave di ogni revisione sistematica. Idealmente, tale ricerca dovrebbe essere condotta o guidata da un esperto specialista dell'informazione (quello che in inglese viene definito come *information specialist*). Una selezione di studi basata sui risultati di una ricerca priva di bias assicura che le ricerche bibliografiche abbiano individuato tutti gli studi rilevanti da includere nella revisione. Le ricerche dovrebbero includere anche dati e altre fonti elettroniche (giornali indicizzati, atti di convegni, registri di studi in corso) per poter identificare studi pubblicati e non pubblicati, e prevedere altri metodi (per es., controllo delle referenze di revisioni già pubblicate, ricerca di citazioni, ricerca manuale) e l'utilizzo di una strategia di ricerca sensibile e appropriata. Le strategie di ricerca dovrebbero includere termini liberi (per es., nel titolo e nell'abstract) e termini controllati (per es., termini MESH per PubMed o Emtree per Embase) utili per identificare gli studi. Potrebbe essere utile cercare anche altri campi, come CAS Registry Number, e nome commerciale del farmaco o del dispositivo, a seconda dell'argomento e della risorsa da cercare. Le strategie sono in genere una combinazione di termini utili per catturare uno o più dei seguenti concetti: popolazione, intervento/test indice/esposizione. Per alcuni tipi di revisione, per esempio revisioni che includono RCT, può essere utile includere un filtro metodologico o per disegno di studio per alcune o per tutte le banche-dati ricercate. Tuttavia, per altri tipi di revisione, come quelle relative a studi di accuratezza diagnostica, è stato dimostrato che l'uso dei filtri può far perdere studi pertinenti e quindi dovrebbe essere evitato<sup>10,11</sup>. Può essere difficile valutare la sensibilità di una strategia di ricerca se si è privi di conoscenze metodologiche specifiche e se non si conosce l'argomento oggetto della revisione. In generale, i valutatori dovrebbero considerare se è stata inclusa una serie di termini appropriata per coprire tutti i possibili modi in cui i concetti utilizzati per catturare la domanda di ricerca potrebbero essere descritti. Per esempio, per una revisione sul cancro della mammella, non è sufficiente utilizzare come parola chiave solo il termine "tumore della mammella". Un esempio di possibile strategia per questo tipo di popolazione utilizzata per MEDLINE via OvidSP, è descritta di seguito:

1. *exp breast neoplasms/ or Carcinoma, Ductal, Breast/ (214761)*
2. *exp breast/ (31443)*
3. *exp neoplasms/ (2533252)*
4. *2 and 3 (17379)*
5. *(breast\$ adj5 (neoplasm\$ or cancer\$ or tumo?r\$ or metasta\$ or carcinoma\$ or adenocarcinoma\$ or sarcoma\$ or dcis or ductal or infiltrat\$ or intraductal\$ or lobular or medullary)).ti,ab,ot,hw. (256928)*
6. *(mammar:\$ adj5 (neoplas\$ or cancer\$ or tumo?r\$ or metasta\$ or carcinoma\$ or adenocarcinoma\$ or sarcoma\$ or dcis or ductal or infiltrat\$ or intra-ductal\$ or lobular or medullary)).ti,ab,ot,hw. (34861)*
7. *1 or 4 or 5 or 6 (277790)*

La tabella 5 riassume le domande chiave incluse nel Dominio 2 insieme a una guida su come valutare ogni singola domanda. La Tabella 6 fornisce una panoramica delle classificazioni per il Dominio 2. Esempi di valutazioni delle revisioni in cui gli esaminatori avevano molte o poche preoccupazioni relative a questo dominio sono forniti nelle tabelle 7 e 8.

Tabella 5. Domande chiave per il Dominio 2 con guida per rispondere alle singole domande.	
Domanda chiave	Guida alla valutazione
2.1 La ricerca bibliografica include un appropriato numero di banche dati/fonti elettroniche per individuare studi pubblicati e non pubblicati?	Il valutatore deve giudicare se le banche dati ricercate sono appropriate. Questo giudizio varierà in base all'argomento della revisione. Si prevede che la ricerca venga condotta almeno su MEDLINE e EMBASE. Vanno inoltre ricercati gli atti dei congressi e i registri di studi. Indicazioni sull'appropriata gamma di banche dati da ricercare sono reperibili nel <i>Cochrane Handbook</i> <sup>9</sup> o sul sito web del Centre for Reviews and Dissemination (CRD) ( <a href="http://www.york.ac.uk/inst/crd/finding_studies_systematic_reviews.htm">http://www.york.ac.uk/inst/crd/finding_studies_systematic_reviews.htm</a> ).
2.2 Oltre che sulle banche dati, sono state condotte altre ricerche per identificare articoli rilevanti?	Le ricerche dovrebbero includere anche la ricerca delle citazioni, il contatto con esperti del settore, il controllo delle referenze bibliografiche, la ricerca manuale degli studi.
2.3 La strategia di ricerca include struttura ed i termini delle strategie di ricerca erano tali da poter individuare quanti più studi possibile?	Per rispondere a questa domanda è necessario che sia disponibile una strategia di ricerca con sufficienti dettagli che permettano di replicarla. Se è disponibile solo una lista dei termini utilizzati senza specificare come essi sono stati combinati tra loro, il valutatore potrà rispondere a questa domanda "Probabilmente sì" o "Probabilmente no". Il valutatore dovrebbe considerare se la strategia di ricerca include una gamma appropriata di termini relativi all'argomento trattato, se è stata fatta una combinazione di termini controllati (come i Medical Subject Headings [MeSH] per MEDLINE) e ricerca con parole libere nel titolo e nell'abstract e se i filtri applicati erano appropriati. Per esempio, nelle revisioni DTA, è stato dimostrato che l'utilizzo dei filtri può far perdere studi rilevanti e quindi a questa domanda si dovrebbe rispondere "No" se la strategia include questi filtri <sup>10</sup> . Istruzioni per una valutazione critica delle strategie di ricerca è disponibile su PRESS Evidence-Based Checklist ( <a href="http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402">http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402</a> ).
2.4 I limiti basati sulla data di pubblicazione o sulla lingua di pubblicazione erano appropriati?	Se nessun limite è stato posto nelle strategie di ricerca, allora a questa domanda si dovrebbe rispondere "Sì". Questa domanda differisce da quella del dominio 1 (1.5) che riguarda i limiti dei criteri di selezione. Per rispondere a questa domanda sono necessarie informazioni relative a tutti e tre i componenti (data di pubblicazione, tipo di pubblicazione, lingua). Limiti legati alla lingua (per es., solo studi scritti in inglese) o al tipo di pubblicazione (per es., solo articoli disponibili in full text) sono raramente (se non mai) appropriati, e quindi se è presente uno di questi limiti, a questa domanda si dovrebbe di solito rispondere "No". I limiti relativi alla data di pubblicazione potrebbero essere appropriati se viene chiaramente motivata la ragione del limite stesso, in questo caso si può rispondere "Sì". Per esempio, se un farmaco o un test non era disponibile prima di una certa data, è ragionevole far partire le strategie di ricerca dalla data in cui il farmaco o il test è diventato disponibile.
2.5 Sono stati fatti sforzi per minimizzare gli errori nella selezione degli studi?	Questa domanda considera sia il processo di screening dei titoli e degli abstract sia la valutazione per l'inclusione dei full text degli articoli. Per poter rispondere a questa domanda sono necessarie informazioni su entrambe queste attività. Per poter rispondere "Sì" i titoli e gli abstract dovrebbero essere stati selezionati indipendentemente da almeno due autori e la valutazione dei full text degli articoli dovrebbe aver coinvolto almeno due autori (sia in doppio sia con uno che valuta e il secondo che controlla la decisione presa).

Tabella 6. Preoccupazioni relative ai metodi utilizzati per identificare e/o selezionare gli studi.	
Poche preoccupazioni	Considerando il quesito della revisione e criteri di eleggibilità valutati nel Dominio 1, un grande sforzo è stato fatto per identificare quanti più studi rilevanti è possibile attraverso varie modalità di ricerca utilizzando strategie di ricerca sensibili e appropriate e sono stati fatti vari passi per minimizzare le distorsioni e gli errori nella selezione degli studi.
Molte preoccupazioni	C'è la possibilità che nella revisione manchino alcuni studi da includere.
Non chiaro	Non vi sono sufficienti informazioni per emettere un giudizio relativo a questo rischio di bias.

## Esempi di valutazione

### a. Revisione giudicata ad alto rischio per quanto riguarda i metodi utilizzati per identificare e/o selezionare gli studi<sup>5</sup>

#### Descrizione dal testo:

*“Un documentalista ha ricercato Ovid MEDLINE dal 1947 fino ad agosto 2012, il database della Cochrane Library (primo trimestre del 2012), Embase (dal 1976 fino ad agosto 2012), Scopus (dal 1960 fino ad agosto 2012), PsycINFO (dal 1806 fino ad agosto 2012), registri di trial clinici e banche dati di studi finanziati. Almeno due autori hanno valutato indipendentemente gli studi per l’inclusione. Gli articoli scritti in altra lingua, oltre che in inglese, sono stati esclusi. Gli studi pubblicati come abstract di conferenze sono stati utilizzati solo nelle analisi di sensitività.”*

**Tabella 7.** Esempi di valutazione per una strategia di ricerca giudicata ad alto rischio.

Domanda chiave	Valutazione	Motivazione
3.1 La ricerca bibliografica include un appropriato numero di banche dati/fonti elettroniche per individuare studi pubblicati e non pubblicati?	Sì	MEDLINE, EMBASE, Scopus, PsycINFO e la Cochrane Library sono state ricercate (sebbene non sia chiaro quali banche dati incluse nella Cochrane Library siano state ricercate). Questa gamma di banche dati è stata giudicata appropriata. Sono stati inoltre ricercati registri di trial e banche dati di studi finanziati.
3.2 Oltre che sulle banche dati, sono state condotte altre ricerche per identificare articoli rilevanti?	No	No
3.3 I termini e la struttura delle strategie di ricerca erano tali da poter individuare quanti più studi possibile?	Nessuna informazione	Non viene riportata in modo completo la strategia di ricerca e non vi sono dettagli sui termini utilizzati; per cui non vi sono informazioni su cui basarsi per la valutazione.
3.4 I limiti basati sulla data di pubblicazione o sulla lingua di pubblicazione erano appropriate?	No	La revisione includeva solo studi in Inglese; questo rappresenta un potenziale bias di pubblicazione.
3.5 Sono stati fatti sforzi per minimizzare gli errori nella selezione degli studi?	Probabilmente sì	È riportato che la valutazione per l’inclusione è stata fatta indipendentemente da almeno due autori. Non è stato esplicitato se questa procedura è stata seguita sia per selezionare i risultati individuati della ricerca sia per la valutazione del full text degli articoli. Questo item è stato valutato come “Probabilmente sì”.
Preoccupazioni relative ai metodi utilizzati per identificare e/o selezionare gli studi	Molte preoccupazioni (Alto livello di rischio di bias).	
Motivazione delle preoccupazioni	Limitare l’inclusione ai soli articoli in lingua Inglese ci porta a pensare che vi sia un alto rischio di bias che studi rilevanti non siano stati inclusi nella revisione. Non possiamo giudicare l’appropriatezza della strategia di ricerca.	

### b. Revisione giudicata a basso rischio per quanto riguarda i metodi utilizzati per identificare e/o selezionare gli studi<sup>8</sup>

#### Descrizione dal testo:

*“Sono state ricercate le seguenti fonti dall’inizio fino ad ottobre 2011: MEDLINE; EMBASE; Cumulative Index To Nursing And Allied Health Literature (CINAHL); The Cochrane Central Register of Systematic Reviews; The Cochrane Central Register of Controlled Trials; DARE; MEDION; SIGLE; Index of Scientific and Technical Proceedings, Web of Science and ClinicalTrials.gov database. Le ricerche consistevano di parole chiave e di termini MeSH relativi ai test in esame combinati con termini MeSH di “Diagnosi prenatale”, “Ultrasonografia”, “Liquido amniotico” ed “Esito della gravidanza”. La strategia di ricerca è consultabile in Appendice S2. Le referenze bibliografiche di tutti gli studi primari inclusi e delle revisioni sono state esaminate per identificare articoli non catturati attraverso le ricerche elettroniche. Non vi sono limiti rispetto alla lingua. Il database è stato esaminato da due autori (RKM, CHM) e sono stati ottenuti tutti gli articoli che soddisfacevano i criteri di inclusione. Le traduzioni sono state ottenute per gli articoli non in inglese. Tre autori hanno preso le decisioni finali rispetto all’inclusione/esclusione in base all’aderenza ai seguenti criteri. “*

**Tabella 8.** Esempio di valutazione per una strategia di ricerca giudicata a basso rischio.

Domanda chiave	Valutazione	Motivazione
2.1 La ricerca bibliografica include un appropriato numero di banche dati/fonti elettroniche per individuare studi pubblicati e non pubblicati?	Sì	Sono state ricercate MEDLINE, EMBASE, CINAHL, SIGLE, Index of Scientific and Technical Proceedings, ClinicalTrials.gov database, la Cochrane Library, DARE e MEDION. È stata ricercata anche Web of Science sebbene non sia chiaro quali banche dati incluse in questa piattaforma siano state ricercate. Questa gamma di banche dati e di risorse è stata giudicata appropriata per individuare studi pubblicati e non pubblicati. Tuttavia il Cochrane Central Register of Systematic Reviews è probabilmente un errore di scrittura e gli autori hanno cercato CDSR.
2.2 Oltre che sulle banche dati, sono state condotte altre ricerche per identificare articoli rilevanti?	Sì	Le referenze bibliografiche di tutti gli studi primari inclusi e delle revisioni sono state esaminate per identificare articoli non catturati attraverso le ricerche elettroniche.
2.3 La struttura e i termini della strategia di ricerca erano tali da poter individuare quanti più studi possibile?	Sì	In un'appendice sul web sono disponibili dettagli sulle strategie di ricerca. Queste combinavano insieme termini riferiti alla popolazione (donne incinte) con i termini riferiti all'intervento (ultrasonografia) e sembravano essere sensibili senza avere limiti inappropriati (per es., il filtro per disegno di studio).
2.4 I limiti basati sulla data di pubblicazione o sulla lingua di pubblicazione erano appropriati?	Sì	Le banche dati sono state ricercate dall'inizio, non sono stati posti limiti di lingua e le ricerche includevano strategie per identificare la letteratura grigia. Pertanto è stato valutato che le strategie di ricerca non avevano limiti e a questa domanda si è risposto "Sì".
2.5 Sono stati fatti sforzi per minimizzare gli errori nella selezione degli studi?	Probabilmente sì	Sono riportate le procedure seguite per lo screening dei titoli e degli abstract e per la valutazione dei full text e queste procedure hanno visto coinvolti più autori. Tuttavia, non era esplicito se gli autori lavoravano in modo indipendente tra loro e quindi a questa domanda è stato risposto "Probabilmente sì".
Preoccupazioni relative ai metodi utilizzati per identificare e/o selezionare gli studi	Poche preoccupazioni (Basso livello di rischio di bias).	
Motivazione delle preoccupazioni	A tutte le domande chiave è stato risposto "Sì" o "Probabilmente sì" e quindi non sono state identificate potenziali aree di dubbio. È probabile quindi che la revisione abbia incluso un'alta percentuale di studi rilevanti.	

## FASE 2. DOMINIO 3: RACCOLTA DATI E VALUTAZIONE DEGLI STUDI

### Background

*Il terzo dominio ha lo scopo di valutare se sono stati introdotti bias durante la raccolta dati o durante il processo di valutazione. Una raccolta rigorosa dei dati dovrebbe prevedere una pianificazione già nella fase del protocollo e l'utilizzo di una scheda di raccolta dati strutturata che sia già stata sperimentata in precedenza. Dovrebbero essere raccolti tutti i dati che possono contribuire alla sintesi e all'interpretazione dei risultati. Questi dati dovrebbero includere sia i dati numerici e statistici che le caratteristiche più generali degli studi primari inclusi come tipo di disegno di studio, finanziamenti, setting, caratteristiche dei partecipanti, criteri di selezione, dettagli su intervento/esposizione/test indice e numero di partecipanti che hanno abbandonato precocemente lo studio. Il tipo di dati numerici da considerare dipende dal tipo di studi inclusi nella revisione. Per esempio, per una revisione DTA, di norma vengono estratti dati 2x2 relativi alle prestazioni del test. Per una revisione di studi osservazionali, possono essere estratti una stima dell'effetto aggiustata con un intervallo di confidenza del 95%, i dettagli delle variabili aggiustate e il metodo di analisi. Se i dati disponibili non erano presentati in forma utile per contribuire alla sintesi, gli autori della revisione dovrebbero riportare come eventualmente hanno ottenuto i dati utili per la sintesi. Per esempio possono essere stati contattati gli autori degli studi primari allo scopo di ottenere dati aggiuntivi. Per ricavare i dati utili alla sintesi, possono anche essere utilizzate appropriate trasformazioni statistiche. L'estrazione dei dati può potenzialmente essere suscettibile di errori. Gli errori potrebbero nascere quando si trascrivono i dati o se si sbaglia nel raccogliere le informazioni rilevanti presenti in uno studio. Possono inoltre verificarsi bias durante il processo di estrazione dei dati che è, per sua natura, soggettivo e aperto a interpretazioni. L'estrazione in doppio dei dati è quindi essenziale per salvaguardarsi contro errori e potenziali bias<sup>2</sup>.*

*La valutazione degli studi inclusi dovrebbe essere fatta utilizzando criteri appropriati al disegno dello studio primario incluso nella revisione<sup>2,9</sup>. Questa valutazione potrebbe essere fatta utilizzando strumenti validati sviluppati per valutare quel tipo di disegno di studio o potrebbe essere fatta utilizzando semplicemente una lista dei criteri rilevanti per individuare potenziali bias. Sia che si utilizzi uno strumento pubblicato sia che si utilizzi una lista di criteri ad hoc, il valutatore dovrebbe considerare se i criteri utilizzati sono sufficienti per identificare i potenziali bias negli studi inclusi. La decisione relativa alla sufficienza o meno dei criteri utilizzati dovrebbe basarsi su prove empiriche relative al rischio di bias dello specifico disegno di studio che si sta valutando<sup>12-17</sup>. Così come avviene per l'estrazione dei dati, bias o errori possono verificarsi durante questo processo di valutazione. Per cui la valutazione del rischio di bias dovrebbe coinvolgere idealmente due autori che lavorano in modo indipendente, o almeno che il secondo autore controlli le decisioni prese dal primo. È necessario riportare il processo svolto e le modalità di risoluzione di eventuali disaccordi sia per la valutazione del rischio di bias sia per l'estrazione dei dati.*

*La tabella 9 riassume le domande chiave incluse nel Dominio 3 insieme a una guida su come valutare ogni singola domanda. La tabella 10 fornisce una panoramica delle classificazioni per il Dominio 3. Esempi di valutazioni delle revisioni in cui gli esaminatori avevano molte o poche preoccupazioni relative a questo dominio sono forniti nelle tabelle 11 e 12.*

**Tabella 9.** Domande chiave per il Dominio 3 con guida per rispondere alle singole domande.

Domanda chiave	Guida alla valutazione
3.1 Sono stati compiuti sforzi per minimizzare gli errori nella raccolta dei dati?	Al fine di ridurre al minimo bias ed errori nel processo di raccolta dei dati, questo processo dovrebbe coinvolgere almeno due autori e bisognerebbe utilizzare schede di estrazione dei dati strutturate e validate. Idealmente, la raccolta dati dovrebbe essere fatta in modo indipendente, ma è accettabile anche l'estrazione da parte di un autore e il successivo controllo dettagliato da parte di un secondo autore. Il secondo autore dovrebbe leggere l'articolo in modo dettagliato per controllare non solo la precisione dei dati estratti, ma anche per garantire che nessuna informazione rilevante sia stata persa.
3.2 Le caratteristiche degli studi sono riportate in modo sufficientemente completo affinché sia gli autori della revisione sia i lettori possano interpretare i risultati?	Le informazioni sulle caratteristiche degli studi sono essenziali per interpretare i risultati della revisione in quanto consentono di valutare appropriatamente l'eterogeneità e l'applicabilità dei risultati. Queste informazioni possono essere riportate nella sezione dei metodi, nella parte relativa alle informazioni sulla raccolta dei dati, possono essere disponibili nelle tabelle relative alle caratteristiche degli studi inclusi o possono essere riassunte nel testo nella sezione risultati. Può essere difficile rispondere a questa domanda poiché spesso, per motivi di spazio, tutte le informazioni raccolte non sono riportate nel testo della pubblicazione. I valutatori potrebbero quindi aver bisogno di accedere a risorse aggiuntive come le appendici web.
3.3 Sono stati considerati tutti i dati rilevanti degli studi inclusi per utilizzarli nella sintesi?	Si dovrebbero estrarre tutti i risultati utili per essere inseriti in modo appropriato in una sintesi cumulativa. Idealmente, gli autori delle revisioni dovrebbero segnalare di quali dati necessitavano per la sintesi e in quale formato. Per esempio, per una revisione DTA vengono generalmente estratti i dati 2x2 delle prestazioni del test. Per una revisione di studi osservazionali, è possibile estrarre una stima dell'effetto aggiustata con un intervallo di confidenza del 95%, i dettagli delle variabili aggiustate e il metodo di analisi. Se questi dati non sono riportati esplicitamente nella sezione dei metodi, potrebbe essere possibile capire quali dati sono stati estratti dalle informazioni presenti nelle tabelle dei risultati, nei riepiloghi grafici (per es., grafici di forest plot) o dai dati riportati nel testo. Potrebbe anche essere necessario accedere a risorse aggiuntive come appendici web o al protocollo della revisione, se disponibile.  È molto raro che tutti gli studi primari inclusi in una revisione riportino i dati nel formato utile per contribuire alla sintesi cumulativa. Per esempio, in una revisione possono essere necessari dati continui che riguardano la differenza media rispetto al baseline con un intervallo di confidenza del 95%. Ci sono anche varie altre misure oltre a questa che possono essere utili, e spesso gli studi non riportano i dati relativi alla variazione prima e dopo il trattamento. Per rispondere "Sì" a questa domanda, nella sezione dei metodi, dovrebbero essere descritte le modalità seguite per ottenere i dati che negli studi non erano presentati in formato utilizzabile per essere inseriti in una sintesi statistica, per es. stimando/trasformando i dati o contattando gli autori per ottenere ulteriori informazioni.
3.4 I rischi di bias (o la qualità metodologica) sono stati formalmente valutati utilizzando criteri appropriati?	Una valutazione formale del rischio di bias è una componente essenziale di qualsiasi revisione. Se il rischio di bias non è stato valutato formalmente, si dovrebbe rispondere "No" a questa domanda. Se è stata effettuata una valutazione formale, i valutatori dovranno giudicare l'appropriatezza dei criteri utilizzati. Se è stato utilizzato uno strumento pubblicato e riconosciuto come valido, come il <i>Cochrane Risk of Bias</i> <sup>18</sup> per gli RCT o QUADAS-2 per gli studi DTA <sup>19</sup> , la valutazione sarà abbastanza semplice e a questa domanda si può rispondere "Sì". Se la revisione elenca semplicemente le domande valutate, utilizza uno strumento non pubblicato o uno strumento che non è più raccomandato, allora diventa più complicato rispondere a questa domanda. Il valutatore dovrà giudicare se i criteri valutati sono utili per identificare, considerando lo specifico disegno di studio, i potenziali bias degli studi primari inclusi. Per esempio, il punteggio Jadad è stato ampiamente utilizzato, ma non include il giudizio sulla mancanza o presenza di problemi relativi al mascheramento nell'assegnazione al braccio di trattamento o di controllo che è stato dimostrato essere un potenziale rischio di bias negli RCT <sup>20</sup> . Per rispondere "Sì" a questa domanda, le revisioni che hanno utilizzato la scala Jadad <sup>20</sup> dovrebbero pertanto aver valutato anche il mascheramento nell'assegnazione ai bracci di trattamento.
3.5 Sono stati fatti sforzi per minimizzare gli errori nella valutazione del rischio di distorsione sistematica?	Come per la raccolta dati, anche la valutazione del rischio di bias dovrebbe coinvolgere due autori sia in doppio sia con uno che valuta e il secondo che controlla la decisione presa.

**Tabella 10.** Preoccupazioni relative ai metodi utilizzati per la raccolta dati e la valutazione degli studi.

Poche preoccupazioni	Considerati gli studi inclusi nella revisione valutati nel Dominio 2, il rischio di bias è stato valutato utilizzando criteri appropriati, l'estrazione dei dati e la valutazione del rischio di bias hanno coinvolto due autori e sono state estratte le caratteristiche e i risultati degli studi inclusi.
Molte preoccupazioni	Durante la raccolta dati o la valutazione del rischio di bias potrebbero essere stati introdotti dei bias.
Non chiaro	Non vi sono sufficienti informazioni per emettere un giudizio relativo a questo rischio di bias.

*Esempi di valutazione***a. Revisione giudicata ad alto rischio per quanto riguarda i metodi utilizzati per la raccolta dati e la valutazione degli studi<sup>7</sup>**

## Descrizione dal testo:

*“Le tabelle 2x2 sono state costruite con i dati estratti da ogni studio. Due autori indipendentemente hanno raccolto ed estratto i dati utilizzando una scheda di estrazione. Qualsiasi differenza è stata risolta attraverso l'accordo... Non si è trovato un consenso su come valutare gli studi che non avevano un braccio di controllo. Quindi, questi criteri non si applicano agli studi senza un braccio di controllo. Pertanto, per questa meta-analisi e revisione sistematica, gli studi sono stati selezionati sulla base della completezza dei dati e dei criteri di inclusione.”*

**Commento: La tabella “Caratteristiche degli studi inclusi” contiene informazioni solo sul disegno di studio (retrospettivo o prospettico), tipo di cancro (tutti tumori del retto) e procedura di confronto (chirurgia per tutti).**

**Tabella 11.** Esempio di valutazione per la raccolta dati e la valutazione degli studi giudicata ad alto rischio.

Domanda chiave	Valutazione	Motivazione
3.1 Sono stati compiuti sforzi per minimizzare gli errori nella raccolta dei dati?	Sì	I dati degli studi sono stati estratti in modo indipendente da due autori che hanno utilizzato una scheda di estrazione strutturata. Le divergenze sono state risolte discutendone.
3.2 Le caratteristiche degli studi sono riportate in modo sufficientemente completo in modo che sia gli autori della revisione sia i lettori possano interpretare i risultati?	No	La tabella di sintesi include informazioni solo sul disegno dello studio (prospettivo o retrospettivo), tipo di tumore (rettale per tutti) e tipo di procedura di conferma (chirurgia per tutti). Nel testo non vengono riportati ulteriori dettagli sul disegno di studio o sul tipo di popolazione arruolata.
3.3 I risultati di tutti gli studi inclusi sono stati considerati per utilizzarli nella sintesi?	Sì	Dati 2x2 sono stati estratti da ogni studio incluso e questo è sufficiente per calcolare le misure di accuratezza diagnostica e i relativi intervalli di confidenza e per condurre la meta-analisi.
3.4 I rischi di bias (o la qualità metodologica) sono stati formalmente valutati utilizzando criteri appropriati?	No	La qualità degli studi non è stata formalmente valutata. Gli autori riportano che questo è dovuto al fatto che non sono riusciti ad accordarsi su come valutare gli studi che non avevano un braccio di controllo, tuttavia vi sono strumenti validati che avrebbero potuto utilizzare per valutare studi di DTA.
3.5 Sono stati fatti sforzi per minimizzare gli errori nella valutazione del rischio di distorsione sistematica?	No	La qualità degli studi non è stata formalmente valutata
Preoccupazioni relative ai metodi utilizzati per la raccolta dati e la valutazione degli studi	Molte preoccupazioni (Alto livello di rischio di bias).	
Motivazione delle preoccupazioni	La mancanza di una valutazione formale della qualità implica che il rischio di bias degli studi inclusi non è chiaro. Vi sono dettagli insufficienti per permettere al lettore di interpretare i risultati. Quindi in questa revisione vi è un alto rischio di bias sia per quanto riguarda la raccolta dei dati sia per quanto riguarda la valutazione della qualità degli studi inclusi.	

## b. Revisione giudicata a basso rischio per quanto riguarda i metodi utilizzati per la raccolta dati e la valutazione degli studi<sup>5</sup>

### Descrizione dal testo:

*“Un ricercatore ha estratto i dati relativi al disegno di studio, alla popolazione, al setting, agli interventi, alle analisi, ai follow-up e ai risultati. Un secondo ricercatore ha esaminato i dati per verificarne l'accuratezza. Due ricercatori hanno applicato in modo indipendente criteri predefiniti per valutare la qualità dello studio come buona, discreta o scarsa. Le discrepanze sono state risolte attraverso il consenso.”*

**Tabella 12.** Esempio di valutazione per la raccolta dati e la valutazione degli studi giudicata a basso rischio.

Domanda chiave	Valutazione	Motivazione
3.1 Vengono compiuti sforzi per minimizzare gli errori nella raccolta dei dati?	Probabilmente sì	Un autore ha estratto i dati dagli studi e un altro ha verificato l'accuratezza di tale estrazione. Non è chiaro se il secondo autore abbia controllato solo l'accuratezza dell'estrazione o abbia anche letto in dettaglio gli articoli per assicurarsi che non si siano perse informazioni rilevanti. Per cui a questa domanda si è risposto “Probabilmente sì” piuttosto che “Sì”.
3.2 Le caratteristiche degli studi sono riportate in modo sufficientemente completo in modo che sia gli autori della revisione sia i lettori possano interpretare i risultati?	Sì	In appendice sono disponibili tabelle con informazioni dettagliate sulle caratteristiche degli studi e sui risultati che consentono ai lettori di interpretare i risultati.
3.3 I risultati di tutti gli studi inclusi sono stati considerati per utilizzarli nella sintesi?	Sì	I dati dicotomici sono stati estratti come numero di eventi e totale dei partecipanti in entrambi i bracci di trattamento; con questi dati è stato calcolato il rischio relativo. Questa procedura è appropriata per la sintesi dei risultati.
3.4 I rischi di bias (o la qualità metodologica) sono stati formalmente valutati utilizzando criteri appropriati?	Probabilmente sì	Non sono disponibili dettagli sui criteri ROB ma gli autori citano i metodi della guida della <i>US Preventive Services Task Force</i> (2001) che include criteri appropriati.
3.5 Sono stati fatti sforzi per minimizzare gli errori nella valutazione del rischio di distorsione sistematica?	Sì	Due autori hanno valutato in modo indipendente la qualità degli studi.
Preoccupazioni relative ai metodi utilizzati per la raccolta dati e la valutazione degli studi	Poche preoccupazioni (Basso livello di rischio di bias).	
Motivazione delle preoccupazioni	A tutte le domande chiave si è risposto “Sì” o “Probabilmente sì” quindi non sono stati identificati potenziali bias. Per cui per quanto riguarda sia l'estrazione dei dati sia la valutazione della qualità metodologica degli studi, questa revisione non sembra avere bias.	

## FASE 2. DOMINIO 4: SINTESI E RISULTATI

### Background

Questo dominio ha lo scopo di valutare se gli autori della revisione hanno utilizzato metodi appropriati per combinare insieme i dati degli studi primari inclusi (in una sintesi quantitativa o non quantitativa). Gli approcci per sintetizzare i risultati dipendono dalla natura della domanda posta dalla revisione e dal tipo di studi primari che vengono sintetizzati. Per studi clinici randomizzati, un approccio comune è quello di utilizzare una media pesata delle stime dell'effetto del trattamento (per misure di rapporto dell'effetto del trattamento si usa una scala logaritmica), pesando per la precisione delle stime<sup>21</sup>. Per fare questo si possono utilizzare modelli a effetto fisso o random. In ogni caso, vi sono diversi modelli statistici per presentare i risultati relativi a un esito (per es., si può, per dati binari, utilizzare una regressione logistica<sup>22</sup>, oppure un modello che considera due o più esiti simultaneamente (meta-analisi bivariata o multivariata<sup>23</sup>), o un modello che valuti gli effetti di multipli trattamenti (network meta-analisi<sup>24</sup>), o infine un modello che consideri le variazioni degli effetti del trattamento (meta-regressione<sup>25</sup>); tutti questi diversi approcci possono anche essere combinati tra loro, rendendo la sintesi molto complessa. Opzioni simili sono disponibili anche per revisioni che hanno domande di ricerca diverse. Per revisioni sull'accuratezza diagnostica, un approccio bivariato è lo standard poiché con questo modello la sensibilità e la specificità sono

considerate simultaneamente tenendo conto della loro correlazione<sup>26</sup>. Per alcune revisioni, la sintesi statistica dei risultati può non essere appropriata e quindi i risultati vengono presentati in modo narrativo.

Alcuni dei più importanti aspetti da considerare in ogni sintesi (sia quantitativa sia non quantitativa) sono: (i) se l'approccio analitico utilizzato è appropriato rispetto al quesito di ricerca posto; (ii) se si è considerata la variabilità tra gli studi (eterogeneità); (iii) se si sono considerati i bias degli studi primari; (iv) se i dati provenienti dagli studi sintetizzati sono completi (particolarmente se vi è il rischio che i dati mancanti siano sistematicamente diversi dai dati disponibili, per es., a causa di pubblicazione selettiva degli esiti o di un bias di pubblicazione); e (v) se gli autori della revisione hanno introdotto dei bias nel riportare i loro risultati. Aspetti tecnici della meta-analisi, come la scelta del metodo da utilizzare per le stime dell'effetto, non sono considerati importanti. In ogni caso, possono esserci errori importanti, come interpretare l'errore standard come una deviazione standard, non riuscire ad aggiustare a seconda del tipo di disegno, per esempio in caso di dati abbinati o raggruppati per cluster, oppure applicando l'approccio standard di medie pesate a rapporti di rischio piuttosto che ai loro logaritmi.

La tabella 13 riassume le domande chiave incluse nel Dominio 4 insieme a una guida su come valutare ogni singola domanda. La tabella 14 fornisce una panoramica delle classificazioni per il Dominio 4. Esempi di valutazioni delle revisioni in cui gli esaminatori avevano molte o poche preoccupazioni relative a questo dominio sono forniti nelle tabelle 15 e 16.

**Tabella 13.** Domande chiave per il Dominio 4 con guida per rispondere alle singole domande.

Domanda chiave	Guida alla valutazione
<p>4.1 La sintesi include tutti gli studi che dovrebbe includere?</p>	<p>La sintesi dovrebbe cercare di includere tutti gli studi per cui sono stati raccolti dati utili per rispondere al quesito di ricerca. Nella sintesi possono mancare i risultati di singoli studi perché gli autori non sono a conoscenza dello studio (probabilmente a causa del bias di pubblicazione): la valutazione relativa alla mancanza di questi studi è stata fatta nel Dominio 2 e nella domanda chiave 4.5. La presente domanda si riferisce a tre ulteriori situazioni riguardanti gli studi identificati per l'inclusione nella revisione: (i) i risultati specifici di uno studio incluso non sono disponibili, (ii) gli autori non sono riusciti a raccogliere o elaborare i dati disponibili o (iii) gli autori hanno volutamente escluso i risultati di alcuni studi. Nel caso (i) deve essere chiaro che gli studi inclusi consideravano dati utili ma non riportavano i risultati richiesti. Per esempio, in un trial clinico possono essere considerati dati relativi a un esito di interesse, ma non essere riportati i risultati relativi. La distinzione importante che va fatta in questo caso è se la mancanza di risultati sia dovuta al tipo di risultati prodotti (per es., la significatività statistica) o se semplicemente essi non sono riportati per motivi pratici (per es., per ragioni legate alla lunghezza dell'articolo). Se i risultati di uno o più studi non sono disponibili poiché essi non erano statisticamente significativi, questo (non)reporting selettivo introdurrà bias nella sintesi degli studi che contribuiscono ai dati. Nel caso (ii) [vedi anche Dominio 3, domanda 3.3], gli studi potrebbero essere stati omessi per errore o perché i revisori non avevano familiarità con i calcoli statistici che avrebbero consentito la loro inclusione. Questo sarebbe un problema nel caso in cui gli studi omessi avessero risultati sistematicamente diversi da quelli inclusi. Per quanto riguarda il caso (iii), gli autori possono prendere decisioni inappropriate escludendo dalla sintesi alcuni studi. Un possibile esempio potrebbe essere l'esclusione basata solo su considerazioni statistiche di studi ritenuti responsabili di una grande eterogeneità (sebbene ciò possa essere ragionevole in un'analisi di sensibilità).</p> <p>Per affrontare questi temi, raccomandiamo che il valutatore esamini il numero di studi inclusi (per es., dalla flow chart) e il numero di studi inclusi nelle sintesi (per es., nei forest plot o nelle tabelle). Un disallineamento porterebbe a dover considerare il motivo per cui gli studi non sono inclusi nelle analisi. Nelle revisioni con un gran numero di studi (per es., più di 10 o 20 per una particolare sintesi), i funnel plot e le relative analisi statistiche possono fornire suggerimenti sulla possibilità di bias di selezione: vedere la domanda 4.5<sup>16</sup>.</p>
<p>4.2 Vengono riportate tutte le analisi definite a priori e, in caso negativo, vengono spiegati i motivi per cui non sono state eseguite?</p>	<p>Lo scopo di questa domanda è identificare i bias introdotti dagli autori nel processo di selezione degli studi da inserire nelle analisi e nella scelta dei metodi di analisi da utilizzare per assicurarsi che non vengano eliminati o sostituiti i risultati che non rispondono alle loro aspettative. Per rispondere "Sì" la revisione avrebbe dovuto seguire un protocollo pubblicato o comunque accessibile. Esempi includono un protocollo registrato su PROSPERO, un registro internazionale prospettico di revisioni sistematiche (<a href="http://www.crd.york.ac.uk/PROSPERO/">http://www.crd.york.ac.uk/PROSPERO/</a>) o un protocollo pubblicato di una revisione Cochrane (<a href="http://www.thecochranelibrary.com/">http://www.thecochranelibrary.com/</a>). Se ci sono indizi che sono state eseguite analisi predefinite, per esempio la sezione dei metodi appare rigorosa e tutte le analisi menzionate sono trattate nei risultati, il valutatore potrebbe rispondere "Probabilmente sì". In caso vi sia una esplicita assenza di un predefinito protocollo, il valutatore dovrebbe rispondere "No" a questa domanda. Tuttavia, se la revisione sistematica non fa riferimento all'esistenza o all'assenza di un protocollo, raccomandiamo che il valutatore risponda come "Nessuna informazione".</p>



segue **Tabella 13.** Domande chiave per il Dominio 4 con guida per rispondere alle singole domande.

Domanda chiave	Guida alla valutazione
<p>4.3 La sintesi è appropriata rispetto alla natura della domanda di ricerca, al disegno degli studi e agli esiti considerati negli studi inclusi?</p>	<p>Lo scopo principale di questa domanda è valutare la metodologia utilizzata per la sintesi dei risultati. La metodologia può essere scelta sulla base del disegno di studio e sulla base del quesito cui si intende rispondere, per cui è importante considerare questi aspetti nella valutazione.</p> <p>Se si è fatta una sintesi quantitativa (meta-analisi), questa domanda valuterà sia se è stato appropriato fare una sintesi statistica sia i metodi statistici utilizzati. Se invece è stata fatta una sintesi narrativa dei risultati, questa domanda valuterà sia se sarebbe stato più opportuno fare una sintesi quantitativa sia l'appropriatezza dell'approccio narrativo utilizzato.</p> <p>La decisione relativa all'appropriatezza della sintesi quantitativa dipende principalmente da quanto gli studi inclusi nella sintesi siano simili tra loro rispetto all'obiettivo di ricerca che si pongono, al disegno di studio e agli esiti (dove presenti). Ciò viene spesso definito "eterogeneità clinica" o "diversità clinica". È importante che il valutatore sia convinto che i risultati sintetizzati abbiano senso – per es., possono potenzialmente influenzare la pratica, la politica sanitaria o future ricerche. Il tema della eterogeneità statistica verrà considerato nella domanda successiva.</p> <p>Il valutatore dovrebbe rispondere "Sì" se è stato utilizzato un "accettabile" metodo statistico per combinare i risultati degli studi, compresa la valutazione appropriata del "peso" di ogni studio. Per esempio, in presenza di variabilità tra gli studi, ma dove si può presumere che gli effetti del trattamento provengano da una distribuzione comune, un modello di meta-analisi a effetto random potrebbe essere considerato appropriato mentre un modello a effetto fisso potrebbe non essere appropriato. Va notato tuttavia che una meta-analisi a effetto random non è appropriata in presenza di una forte correlazione tra dimensione dello studio ed effetto stimato ("effetti degli studi di piccole dimensioni" o "funnel plot asimmetria"). Si possono considerare approcci che utilizzano la regressione logistica o l'analisi per sottogruppi per esplorare la variabilità tra gli studi. In genere non raccomandiamo metodi per combinare i risultati degli studi che ignorano l'influenza della precisione sulla stima cumulativa – per es., quelli che danno lo stesso peso a tutti gli studi. Se vengono utilizzati questi metodi (per es., <i>vote counting</i>), suggeriamo di rispondere "No" a questa domanda.</p> <p>Per studi di accuratezza diagnostica, un metodo comunemente accettato per considerare la correlazione tra sensibilità e specificità è l'approccio bivariato. Potrebbe essere necessario ricorrere a una consulenza statistica per valutare le meta-analisi che utilizzano approcci multivariati, multi-level, bayesiani o network-meta-analisi.</p>
<p>4.4 La variabilità tra gli studi inclusi nelle sintesi era minima o, se presente, discussa?</p>	<p>Questa domanda si riferisce alla variabilità dei risultati degli studi inclusi piuttosto che alla variabilità delle loro caratteristiche. Tale variabilità può essere valutata visivamente (per es., con i forest plot); usando un test statistico (per es., il test <math>\chi^2</math> o Q della Cochrane); utilizzando una misura della varianza tra studi (spesso denominata <math>\tau^2</math>); o usando una misura della sovrapposizione degli intervalli di confidenza (per es., il test statistico <math>I^2</math> per l'inconsistenza dei risultati).</p> <p>Se una sostanziale eterogeneità viene ignorata in una meta-analisi, questo può portare a conclusioni fuorvianti e/o a imprecisioni. Se in presenza di eterogeneità viene utilizzato un modello a effetti fissi, è importante che l'autore riconosca che quell'analisi ignora l'eterogeneità. Il risultato dovrebbe essere integrato con i risultati relativi alla variabilità tra gli studi, e i risultati dell'analisi a effetti fissi dovrebbero essere interpretati in modo appropriato. Nella maggior parte dei casi, l'utilizzo di una meta-analisi a effetti fissi senza ulteriori analisi per esplorare la fonte della variabilità porta a rispondere "No" a questa domanda.</p> <p>Se è stato utilizzato in modo appropriato un modello a effetti random a causa dell'eterogeneità e se sono state fatte ulteriori analisi di sottogruppo/meta regressione per esplorare l'eterogeneità, il valutatore potrà rispondere "Sì" a questa domanda. Tuttavia, la meta-analisi a effetti random di studi estremamente diversi nelle caratteristiche o nei risultati può dare un risultato privo di significato, in particolare se i risultati degli studi indicano direzioni dell'effetto opposte.</p> <p>Se è stata fatta una sintesi narrativa poiché la sintesi statistica è stata giudicata inappropriata a causa della eterogeneità clinica, si può rispondere "Sì" a questa domanda, poiché l'eterogeneità è stata affrontata non combinando insieme i risultati.</p>

segue **Tabella 13.** Domande chiave per il Dominio 4 con guida per rispondere alle singole domande.

Domanda chiave	Guida alla valutazione
<p>4.5 I risultati erano robusti, per es., attraverso la dimostrazione con funnel plot o analisi di sensitività?</p>	<p>Questa domanda riguarda l'incertezza sui risultati delle sintesi statistiche e quanto questi risultati potrebbero modificarsi cambiando l'approccio utilizzato per la sintesi stessa. I funnel plot vengono usati per esaminare le relazioni tra le dimensioni dell'effetto e le dimensioni dello studio (spesso misurate statisticamente usando la precisione). L'asimmetria in un forest plot può essere dovuta a vari motivi che includono il bias di pubblicazione, la mancanza di dati riportati da uno studio, la differente importanza dei bias negli studi di grandi dimensioni rispetto a quelli di piccole dimensioni, oppure ad autentiche differenze negli effetti di studi di diverse dimensioni. Se il funnel plot è asimmetrico l'utilizzo di un metodo a effetto fisso o random produce risultati differenti. Pertanto può fornire informazioni interessanti utilizzare entrambi gli approcci come se fosse un'analisi di sensitività. Altrimenti possono essere utilizzati il metodo trim-and-fill, la regressione logistica o l'utilizzo di modelli selettivi per stimare il potenziale impatto sui risultati di differenti tipi di bias di pubblicazione<sup>27</sup>.</p> <p>Altri esempi di analisi di sensitività sono l'uso di metodi diversi per imputare i dati mancanti, o l'esclusione di studi singoli che si ritiene possano influenzare fortemente il risultato della sintesi. In alternativa, per verificare se i risultati sono robusti, può essere eseguita un'analisi di sensitività escludendo gli studi che non hanno utilizzato criteri diagnostici formali per identificare la malattia in studio o studi che sono stati valutati ad alto rischio di bias.</p> <p>L'uso di analisi di sensitività di per sé non è considerato sufficiente per ricevere una risposta "Sì" a questa domanda; i valutatori dovrebbero decidere se il loro uso ha dimostrato la solidità dei risultati. Se ci sono pochissimi studi o studi molto eterogenei, può essere chiaro che i risultati non sono robusti anche se gli autori della revisione non hanno fatto analisi di sensitività.</p> <p>Se è stata fatta una sintesi narrativa, il valutatore dovrebbe giudicare se diversi approcci per riassumere i risultati degli studi avrebbero potuto portare a conclusioni diverse e se singoli studi stanno influenzando le conclusioni.</p>
<p>4.6 I rischi di bias degli studi erano minimi o, se presenti, discussi nei risultati?</p>	<p>I valutatori sono incoraggiati a rispondere "No" a questa domanda se ritengono che negli studi inclusi vi siano bias importanti che non sono stati considerati dagli autori. Per esempio, se nella revisione non viene valutato il rischio di bias degli studi inclusi, il valutatore che utilizza ROBIS, dovrebbe rispondere "No". Allo stesso modo si dovrà rispondere "No" se il rischio di bias è stato valutato ma gli autori non lo considerano nelle sezioni risultati/conclusioni. Si può rispondere "Sì" a questa domanda se tutti gli studi inclusi sono stati giudicati a basso rischio di bias o se, quando gli studi erano ad alto rischio di bias, sono state fatte analisi di sensitività.</p> <p>Se i bias sono affrontati solo come parte della discussione dei risultati di una sintesi, l'esaminatore dovrebbe rispondere "No"; tali discussioni sono trattate più avanti nel giudizio complessivo sul rischio di bias della revisione (Fase 3).</p>

**Tabella 14.** Preoccupazioni relative ai metodi utilizzati per sintetizzare i risultati.

<p>Poche preoccupazioni</p>	<p>È improbabile che la sintesi produca risultati distorti, poiché qualsiasi limitazione dei dati è stata superata, oppure i risultati sono stati così convincenti che le limitazioni avrebbero avuto un impatto minimo.</p>
<p>Molte preoccupazioni</p>	<p>È probabile che la sintesi produca risultati distorti, in quanto (i) i potenziali bias sono stati ignorati (nei singoli studi e o tra gli studi); (ii) non è stata considerata una variabilità significativa tra gli studi; (iii) c'erano importanti inadeguatezze nella metodologia; o (iv) i risultati sono riportati in modo incompleto, tale da suscitare preoccupazioni.</p>
<p>Non chiaro</p>	<p>Non vi sono sufficienti informazioni per emettere un giudizio relativo a questo rischio di bias.</p>

*Esempi di valutazione***a. Revisione giudicata ad alto rischio per quanto riguarda la sintesi dei risultati<sup>28</sup>**

**Tabella 15.** Esempio di valutazione per la sintesi dei risultati giudicata ad alto rischio.

Domanda chiave	Valutazione	Motivazione
4.1 La sintesi include tutti gli studi che dovrebbe includere?	Nessuna informazione	L'autore dichiarava che 35 studi erano rilevanti per la revisione, solo i dati di 29 studi sono stati inseriti nella meta-analisi. Nessun'altra informazione era disponibile.
4.2 Vengono riportate tutte le analisi definite a priori e, in caso negativo, vengono spiegati i motivi per cui non sono state eseguite?	Nessuna informazione	Nessuna analisi era stata predefinita in un protocollo pubblicato. Nessun'altra informazione disponibile nel testo
4.3 La sintesi è appropriata rispetto alla natura della domanda di ricerca, al disegno degli studi e agli esiti considerati negli studi inclusi?	No	L'autore dichiarava che "è stata fatta una meta-analisi" (p.11). L'autore ha sommato gli eventi e le popolazioni e ha utilizzato i totali per generare l'odds ratio cumulativo, quindi non ha utilizzato i confronti all'interno degli studi. L'autore ha descritto gli studi da mettere in comune, ma sembrerebbe che vi sia una diversità clinica tra gli studi.
4.4 La variabilità tra gli studi inclusi nelle sintesi era minima o, se presente, discussa?	Probabilmente no	Non sono stati forniti dettagli sull'eterogeneità statistica. In ogni caso l'autore ha discusso le analisi di sottogruppo per gli studi ad alto rischio, gli studi partner e quelli con popolazione random. Inoltre, gli studi sono stati raggruppati in base alla prevalenza della circoncisione nella comunità. Non sono stati forniti ulteriori dettagli sui sottogruppi.
4.5 La robustezza dei dati era dimostrata, per es., attraverso funnel plot o analisi di sensibilità?	Probabilmente no	L'autore ha discusso alcuni studi che potevano essere problematici. "Sebbene possano essere assunzioni errate, l'inclusione o l'esclusione di questi studi non ha influito sul risultato complessivo" (P.11). Per queste analisi non sono riportati gli odds ratio ed i loro intervalli di confidenza.
4.6 I rischi di bias degli studi erano minimi o, se presenti, discussi nei risultati?	No	La qualità degli studi o il loro rischio di bias non è stato esplicitamente valutato. I bias non vengono discussi in modo esplicito nella sintesi.
Preoccupazioni relative ai metodi utilizzati per sintesi dei risultati	Molte preoccupazioni (Alto livello di rischio di bias).	
Motivazione delle preoccupazioni	Non è stato valutato il rischio di bias dei singoli studi, né il loro potenziale bias è stato considerato nella sintesi. L'autore ha sommato i gruppi di intervento attraverso gli studi piuttosto che analizzare i confronti all'interno degli studi, questo è un approccio pericoloso nella meta-analisi. Non è stata discussa o valutata l'eterogeneità nell'analisi.	

**b. Revisione giudicata a basso rischio per quanto riguarda la sintesi dei risultati<sup>8</sup>****Tabella 16.** Esempio di valutazione per la sintesi dei risultati giudicata a basso rischio.

Domanda chiave	Valutazione	Motivazione
4.1 La sintesi include tutti gli studi che dovrebbe includere?	Nessuna informazione	La flow chart riporta 43 studi inclusi nella revisione. Anche un'altra immagine che riporta le valutazioni QUADAS-2 mostra 43 studi. Anche un'appendice con dettagli sulle caratteristiche degli studi inclusi riporta dati su 43 studi. Tuttavia, i forest plot non presentano risultati per singolo studio, poiché gli studi potrebbero riportare più di un risultato. Non è possibile verificare se tutti gli studi siano stati inclusi in queste sintesi.
4.2 Vengono riportate tutte le analisi definite a priori e, in caso negativo, vengono spiegati i motivi per cui non sono state eseguite?	Probabilmente sì	Era disponibile un protocollo come informazione supplementare. Nessuna deviazione da esso è stata dichiarata o osservata. Sebbene il protocollo non specificasse i dettagli sulla valutazione del rischio di bias o sulla strategia da adottare per le analisi, la sezione dei metodi dell'articolo affronta entrambi questi temi in modo rigoroso. Tutte le analisi citate nella sezione metodi sono presentate nei risultati.
4.3 La sintesi è appropriata rispetto alla natura della domanda di ricerca, al disegno degli studi e agli esiti considerati negli studi inclusi?	Sì	Gli autori affermano che "i risultati sono stati raggruppati usando un modello di meta-analisi a effetti random in cui la definizione della misura del volume del liquido amniotico, la soglia utilizzata e la misura del risultato erano le stesse" (pag. 688). Per valutare la capacità predittiva dei test, gli autori hanno calcolato la sensibilità, la specificità e i rapporti di verosimiglianza, utilizzando i dati delle tabelle 2x2 sintetizzati utilizzando un modello di meta-analisi effetti random bivariati.
4.4 La variabilità tra gli studi inclusi nelle sintesi era minima o, se presente, discussa?	Sì	In questa revisione l'eterogeneità statistica è stata prevista e affrontata utilizzando un modello a effetti random. Sono state eseguite analisi di meta-regressione o di sottogruppo per esaminare potenziali modificatori di effetto. Sono stati calcolati, inoltre, gli intervalli predittivi quando $I^2 > 0\%$ .
4.5 La robustezza dei dati era dimostrata, per es., attraverso funnel plot o analisi di sensibilità?	Probabilmente no	Gli autori non hanno dichiarato se hanno eseguito delle analisi di sensibilità per valutare la robustezza dei loro risultati. Tuttavia, c'è una discussione sull'uso di un esito composito.
4.6 I rischi di bias degli studi erano minimi o, se presenti, discussi nei risultati?	Probabilmente sì	I bias sono stati valutati usando lo strumento QUADAS-2. Questa valutazione indicava che gli studi erano ad alto rischio di bias per la descrizione dello standard di riferimento e per la cecità. Altri aspetti di QUADAS-2 non suscitavano grande preoccupazione. La "qualità" è stata specificata in analisi per sottogruppi, ma le analisi per sottogruppi sono state riportate solo per uno dei due esiti principali. Per l'oligoidramnio e il risultato del benessere fetale gli autori hanno affermato che "Non c'era alcuna differenza in nessuna delle analisi di sottogruppo" (pag. 691).
Preoccupazioni relative ai metodi utilizzati per sintesi dei risultati	Poche preoccupazioni (Basso livello di rischio di bias).	
Motivazione delle preoccupazioni	Gli autori, nelle loro analisi, hanno valutato in modo soddisfacente l'eterogeneità e l'hanno esplorata attraverso analisi di sottogruppo. Il rischio di bias dei singoli studi è stato valutato e gli autori dichiaravano di averlo considerato in analisi di sottogruppo.	

### Fase 3: Valutare il rischio di bias della revisione

La fase finale considera se la revisione nel suo insieme è a rischio di bias. Questa valutazione utilizza la stessa struttura utilizzata nei singoli domini di fase 2, incluse le domande chiave e le informazioni relative a come si è formulato il giudizio, ma il giudizio relativo alle preoccupazioni relative alla presenza di bias è sostituito da un giudizio complessivo sul rischio di bias nella revisione. La prima domanda chiave per questa fase chiede se nell'interpretazione dei risultati sono stati considerati tutti i dubbi identificati nei domini da 1 a 4. Se non sono state identificate preoccupazioni, a questa domanda si risponde "Sì". Se uno o più dubbi sono stati identificati per uno qualsiasi dei precedenti domini, ma questi sono stati appropriatamente considerati nell'interpretare i risultati e nelle conclusioni, allora di nuovo si può rispondere "Sì" e, in base alla valutazione data alle altre domande chiave, la revisione può essere giudicata a "basso rischio di bias".

Questa fase include inoltre altre tre domande chiave correlate all'interpretazione dei risultati della revisione. Queste valutano se le conclusioni sono supportate dalle prove presentate, se è stata considerata la rilevanza degli studi inclusi e se gli autori hanno evitato di enfatizzare i risultati sulla base della significatività statistica. Questi sono tutti aspetti importanti dei risultati di una revisione attraverso cui possono essere introdotti rischi di bias.

La tabella 17 riassume le domande chiave incluse nel Dominio 4 insieme a una guida su come valutare ogni singola domanda. La tabella 18 fornisce una panoramica delle classificazioni per il Dominio 4.

Tabella 17. Domande chiave per la Fase 3 con guida per rispondere alle singole domande.	
Domanda chiave	Guida alla valutazione
A. L'interpretazione dei risultati considera tutte le preoccupazioni identificate nella valutazione di Fase 2?	Questa domanda fa riferimento alla valutazione delle preoccupazioni relative al processo di revisione eseguito nella Fase 2 e riassunte all'inizio della Fase 3. Se tutti i domini valutati durante la Fase 2 sono stati classificati come "Poche preoccupazioni" (basso livello), allora a questa domanda si può rispondere "Sì". Se a uno o più domini si è risposto "Molte preoccupazioni" (alto livello) o "Preoccupazioni non chiare" (livello non chiaro), i revisori devono considerare se gli autori della revisione hanno affrontato in modo appropriato, nell'interpretazione dei risultati, le preoccupazioni identificate durante la Fase 2. Per esempio, una revisione può essere stata giudicata ad alto rischio di bias perché ha applicato un filtro alle ricerche relativo al disegno di studio, che potrebbe aver portato alla perdita di studi rilevanti. Se questa revisione ha riconosciuto questo come potenziale limite e ha discusso di come ciò avrebbe potuto influire sui risultati nella discussione e nelle conclusioni della revisione, a questa domanda chiave si può rispondere "Sì".
B. È stata considerata in modo appropriato la rilevanza degli studi individuati rispetto al quesito di ricerca?	Un aspetto importante da considerare quando si interpretano i risultati di una revisione riguarda la rilevanza (applicabilità/validità esterna) degli studi identificati/inclusi rispetto al quesito della revisione. Se gli studi inclusi non sono direttamente applicabili al quesito cui si intende rispondere e se questo limite non viene discusso dagli autori nell'interpretazione dei risultati della revisione, allora vi può essere un rischio di bias. Alcuni autori di revisioni possono considerare la rilevanza degli studi inclusi come parte della valutazione formale della qualità. Questo è vero soprattutto per le revisioni di accuratezza diagnostica che per valutare la qualità hanno utilizzato lo strumento QUADAS-2 che include una valutazione formale dell'applicabilità degli studi primari al quesito della revisione, nonché la valutazione del rischio di bias degli stessi studi. Altri autori di revisioni possono utilizzare un approccio meno formale, discutendo la rilevanza degli studi nella discussione. Entrambi gli approcci possono essere appropriati a condizione che nelle conclusioni della revisione si rifletta adeguatamente sulle prove disponibili, compresa la rilevanza degli studi inclusi.
C. Gli autori della revisione hanno evitato di enfatizzare i risultati sulla base della significatività statistica?	Se in una revisione sono state eseguite più analisi, è importante che gli autori presentino tutte le analisi in modo equilibrato. A volte gli autori delle revisioni possono scegliere di evidenziare i risultati sulla base della loro significatività statistica. Per esempio, se una revisione valuta 10 esiti e non trova effetti per 9, e solo per uno i risultati sono a favore dell'intervento in modo statisticamente significativo, essi possono essere tentati di enfatizzare i risultati di quest'unico esito. Tuttavia, ciò fornirebbe un'immagine fuorviante dei veri risultati della revisione ed è quindi importante che ciò non avvenga.

**Tabella 18.** Rischi di bias introdotti dai metodi utilizzati per identificare e/o selezionare gli studi.

Basso rischio di bias	È probabile che i risultati della revisione siano affidabili. La Fase 2 non ha evidenziato alcun problema per quanto attiene al processo di conduzione della revisione o le preoccupazioni sono state adeguatamente considerate nelle conclusioni della revisione stessa. Le conclusioni sono state supportate dalle prove e hanno compreso l'esame della rilevanza degli studi inclusi.
Alto rischio di bias	Una o più delle preoccupazioni sollevate durante la valutazione di Fase 2 non è stata affrontata nelle conclusioni della revisione, le conclusioni della revisione non sono state supportate dalle prove, o le conclusioni non hanno considerato la rilevanza degli studi inclusi rispetto al quesito della revisione.

### Esempi di valutazione

#### a. Revisione giudicata ad alto rischio di bias<sup>7</sup>

**Tabella 19.** Riassunto delle preoccupazioni identificate durante la valutazione di fase 2.

Domanda chiave	Preoccupazione	Motivazione
Preoccupazioni relative ai criteri di eleggibilità degli studi	Molte preoccupazioni	I criteri di eleggibilità erano poco chiari e ambigui; non erano riportati dettagli sul tipo di popolazione eleggibile.
Preoccupazioni relative ai metodi utilizzati per identificare e/o selezionare gli studi	Molte preoccupazioni	È probabile che studi rilevanti siano stati persi dalle ricerche. Sebbene pochi dettagli vengano riportati, non sembra che sia stata cercata una gamma appropriata di banche dati, che la strategia di ricerca comprendesse termini per individuare studi diagnostici; non sono riportati metodi aggiuntivi oltre la ricerca delle banche dati, e la restrizione delle ricerche fino al 1980 non era giustificata.
Preoccupazioni relative ai metodi utilizzati per la raccolta dati e la valutazione degli studi	Molte preoccupazioni	La mancanza di una valutazione formale della qualità implica che non è chiaro il rischio di bias degli studi inclusi. Non erano disponibili dettagli sugli studi sufficienti per consentire al lettore di interpretare i risultati. Vi è quindi un alto rischio di bias sia nella raccolta dei dati sia nel processo di valutazione della qualità degli studi.
Preoccupazioni relative ai metodi utilizzati per sintesi dei risultati	Molte preoccupazioni	Sebbene i metodi di analisi utilizzati in questa revisione non fossero i più robusti dal punto di vista statistico, sembra probabile che siano affidabili. La preoccupazione principale di questa revisione è che non è stata investigata l'eterogeneità, e in particolare il risultato ad essa collegato. Poiché i dettagli disponibili dello studio che causava eterogeneità erano molto limitati e non è stata condotta una valutazione formale del rischio di bias, il lettore non può valutare se questo studio sia diverso dagli altri studi. Se i risultati di questo studio fossero più attendibili rispetto a quelli degli altri studi inclusi nella revisione, i risultati generali della revisione sarebbero a rischio di distorsione.

#### Descrizione dal testo:

*“Questa meta-analisi e revisione sistematica mostra che la sensibilità e la specificità combinate di EUS per l'invazione tumorale T0 è molto alta (circa il 97%) ... L'eterogeneità tra diversi studi è stata valutata non solo con test di eterogeneità, ma anche disegnando curve SROC e trovando la AUC, dal momento che diversi studi potrebbero utilizzare criteri leggermente diversi per la stadiazione. Un AUC di 1 per ogni test diagnostico indica che il test è eccellente. Le curve SROC per EUS hanno mostrato che il valore di AUC era molto vicino a 1, indicando che EUS è un eccellente test diagnostico per lo stadio T0 dei tumori rettali ... EUS ha un'eccellente sensibilità e specificità, che aiuta a diagnosticare accuratamente lo stadio T0 dei tumori rettali.”*

**Tabella 20.** Esempio di valutazione per la sintesi giudicata ad alto rischio di bias.

Domanda chiave	Valutazione	Motivazione
A. L'interpretazione dei risultati considera tutte le preoccupazioni identificate nella valutazione di Fase 2?	No	Gli autori della revisione non hanno valutato nessuno dei limiti considerati nella valutazione di Fase 2 e quindi essi non sono stati considerati nell'interpretazione dei risultati.
B. È stata considerata in modo appropriato la rilevanza degli studi individuati rispetto al quesito di ricerca?	No	La revisione non ha considerato la rilevanza degli studi inclusi rispetto al quesito di ricerca della revisione e non c'erano dettagli sufficienti, soprattutto in relazione alla popolazione, per consentire al lettore di fare questa valutazione.
C. Gli autori della revisione hanno evitato di enfatizzare i risultati sulla base della significatività statistica?	Sì	I dati enfatizzati nella discussione e su cui si basano le conclusioni sono tutti basati su stime riassuntive dell'analisi principale.
Rischio di bias	Alto	
Motivazione delle preoccupazioni	La valutazione di Fase 2 ha identificato una serie di preoccupazioni relative al processo di conduzione della revisione che non sono state affrontate dagli autori. Tra questi vi sono la mancanza di chiarezza nei criteri di inclusione, la possibilità di non individuare attraverso le strategie di ricerca studi potenzialmente rilevanti, la mancanza di una valutazione formale della qualità degli studi inclusi, dettagli insufficienti sugli studi inclusi e l'incapacità di considerare in modo appropriato le differenze tra gli studi nella sintesi. La revisione inoltre non considera la rilevanza degli studi identificati rispetto alla domanda di ricerca cui intendeva rispondere. Quindi questa revisione è ad alto rischio di bias.	

## b. Revisione giudicata a basso rischio di bias

**Tabella 21.** Riassunto delle preoccupazioni identificate durante la valutazione di fase 2.

Domanda chiave	Preoccupazione	Motivazione
1. Preoccupazioni relative ai criteri di eleggibilità degli studi	Poche preoccupazioni	A tutte le domande chiave si è risposto "Sì" o "Probabilmente sì" per cui non vi sono potenziali preoccupazioni relative ai criteri di eleggibilità degli studi. È stato fatto un considerevole sforzo per specificare con chiarezza quale fosse il quesito di ricerca della revisione e quali fossero gli obiettivi e per predefinire e giustificare appropriati e dettagliati criteri di eleggibilità
2. Preoccupazioni relative ai metodi utilizzati per identificare e / o selezionare gli studi	Poche preoccupazioni	Viene riportata la procedura seguita per lo screening dei titoli e degli abstract e per la valutazione dei full text degli articoli, ed entrambe le procedure coinvolgevano più autori. In ogni caso non era chiaro se gli autori lavorassero in doppio ed indipendentemente per cui a questa domanda si è risposto "Probabilmente sì".
3. Preoccupazioni relative ai metodi utilizzati per la raccolta dati e la valutazione degli studi	Poche preoccupazioni	Tutti gli articoli sono stati valutati da un minimo di due autori e i dati necessari sono stati estratti in modo indipendente. La qualità degli studi è stata valutata utilizzando uno strumento appropriato.
4. Preoccupazioni relative ai metodi utilizzati per sintesi dei risultati	Poche preoccupazioni	Gli autori hanno affrontato in modo soddisfacente nelle loro analisi la eterogeneità e l'hanno esplorata con analisi di sottogruppo. Hanno considerato il rischio di bias dei singoli studi e hanno dichiarato di averlo considerato nelle analisi di sottogruppo.

**Descrizione dal testo:****Conclusioni**

*“I dati attuali suggeriscono che l'oligohidramnios (scarso liquido amniotico) è fortemente associato a feti piccoli per età gestazionale e a mortalità, e il polyhidramnios (molto liquido amniotico) a feti con peso alla nascita > 90 centile. Nonostante le forti associazioni con risultati sfavorevoli, le prove disponibili non sono sufficienti per prevedere con precisione il rischio per i singoli individui”.*

**Discussione**

*“La nostra valutazione della qualità ha rivelato preoccupazioni per quanto riguarda la possibilità di bias relativi alla selezione dei pazienti, alle prestazioni del test indice e dello standard di riferimento. Non siamo stati in grado di eseguire analisi di sottogruppi per le gravidanze pretermine rispetto a quelle a termine e alcuni studi hanno riportato dati insufficienti per determinare se le soglie per la misurazione del liquido amniotico sono state aggiustate per tempo di gestazione. Dove possibile abbiamo utilizzato i risultati ottenuti più vicini al parto e abbiamo eseguito analisi di sottogruppi per quelli in cui il test è stato eseguito entro 7 giorni dal parto. In particolare, vi erano molte poche informazioni sull'esatta metodologia degli standard di riferimento e se eventualmente è stato somministrato qualche trattamento tra la misura effettuata col test indice e quella fatta col test di riferimento. Una delle maggiori preoccupazioni riguarda quindi il numero di gravidanze indotte al parto a causa della scarsità di liquido amniotico, cosa che influenza i risultati relativi all'esito della gravidanza, il cosiddetto bias dovuto all'intervento. Questo bias può essere veramente rimosso solo eseguendo un RCT, ma questo sarebbe impossibile da eseguire in quanto le misurazioni del volume del liquido amniotico sono diventate lo standard nella sorveglianza fetale e nella gestione delle gravidanze ad alto rischio e quindi sarebbe molto difficile reclutare pazienti per questo tipo di studio. Infine, le misure di esito utilizzate in questa revisione sono state quelle riportate dagli autori degli studi inclusi, si riconosce che molte delle misure di esito sono soggettive (ad esempio l'ammissione all'unità di terapia intensiva neonatale, la necessità di rianimazione). Non veniva riportata negli studi l'unica misura oggettiva di esiti negativi per il feto che è rappresentata dalla correlazione tra campioni di pH del cordone ombelicale e risultati a lungo termine come la paralisi cerebrale”.*

**Tabella 22.** Esempio di valutazione per una sintesi giudicata a basso rischio di bias.

Domanda chiave	Valutazione	Motivazione
A. L'interpretazione dei risultati considera tutte le preoccupazioni identificate nella valutazione di Fase 2?	Sì	Non sono state identificate preoccupazioni durante la valutazione della Fase 2.
B. È stata considerata in modo appropriato la rilevanza degli studi individuati rispetto al quesito di ricerca?	Probabilmente sì	Sebbene non esplicitamente, le implicazioni dei risultati della revisione a livello dei singoli studi sono state discusse in dettaglio. Le potenziali fonti di bias relative al tipo di popolazione inclusa e agli esiti considerati negli studi sono state discusse in dettaglio anche rispetto alla loro rilevanza.
C. Gli autori della revisione hanno evitato di enfatizzare i risultati sulla base della significatività statistica?	Sì	Le conclusioni della revisione riflettono sia i risultati statisticamente significativi sia quelli non significativi.
Rischio di bias	Basso	
Motivazione delle preoccupazioni	La valutazione di Fase 2 non ha evidenziato preoccupazioni relative al processo di conduzione della revisione. I potenziali limiti degli studi inclusi in termini di rischio di bias sono stati affrontati in dettaglio nella discussione. Le conclusioni della revisione riflettono adeguatamente i risultati.	

**Problemi pratici nell'uso di ROBIS**

Non sempre è semplice valutare il rischio di bias delle revisioni sistematiche. ROBIS è stato disegnato per fornire un approccio strutturato per valutare il rischio di bias, con una valutazione facoltativa dell'applicabilità. Mentre ROBIS è stato progettato per essere utilizzabile da autori con background diversi, è probabile che sia richiesta una certa competenza metodologica e di contenuto. Si raccomanda che una valutazione ROBIS sia completata da due revisori, idealmente indipendentemente, oppure da un revisore con un secondo revisore che controlla la valutazione. Raccomandiamo di considerare, per la valutazione con ROBIS, tutte le domande chiave. ROBIS è stato disegnato per essere genericamente applicabile a revisioni che considerano una varietà di argomenti



e quindi tutte le domande chiave dovrebbero essere rilevanti per tutti i tipi di revisione. ROBIS implica che si formulino dei giudizi. Il nostro scopo era di rendere questo processo il più trasparente possibile chiedendo ai revisori di riportare le informazioni utilizzate per formulare i giudizi, rispondere alle domande chiave e di esplicitare il razionale per la formulazione del giudizio complessivo.

Una completa valutazione ROBIS richiede che vengano valutate tutte le domande chiave per tutti i domini. Tuttavia, in determinate situazioni potrebbe essere opportuno “fermarsi” una volta raggiunto un determinato punteggio. Questo dipende dallo scopo della valutazione. Sarà necessaria una valutazione completa se lo scopo della valutazione ROBIS è di fornire una valutazione del rischio di bias nella revisione e le ragioni di eventuali bias. Tuttavia, se lo scopo è semplicemente identificare se la revisione è in generale ad alto rischio di bias o se ci sono preoccupazioni rispetto a specifici domini, i valutatori possono scegliere di fermarsi una volta che è stato identificato un alto rischio di bias. Per esempio, se una revisione non ha ricercato un numero appropriato di banche dati (domanda chiave 2.1), la revisione può essere considerata a elevato rischio per questo dominio. Gli esaminatori interessati solo a verificare se la revisione è a rischio di bias possono quindi scegliere di non valutare il resto delle domande chiave relative a questo dominio o in uno qualsiasi degli altri domini, semplicemente assegnando una valutazione di “alto” rischio di bias.

## Bibliografia

- Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological Expectations of Cochrane Intervention Reviews (MECIR): methodological standard for the conduct of new Cochrane Intervention Reviews. 2012; Version 2.2, 17 December 2012: <http://www.editorial-unit.cochrane.org/mecir>.
- Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]. York: University of York; 2009 [accessed 23.3.11].
- Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD January 2014. AHRQ Publication No. 10(14)-EHC063-EF.
- Institute of Medicine. finding what works in health care: standards for systematic reviews. Washington, DC: The National Academies Press, 2011.
- Chou R, Hartung D, Rahman B, Wasson N, Cottrell EB, Fu R. Comparative effectiveness of antiviral treatment for hepatitis C virus infection in adults: a systematic review. *Ann Intern Med* 2013; 158: 114-23.
- 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.
- Puli SR, Bechtold ML, Reddy JB, Choudhary A, Antillon MR. Can endoscopic ultrasound predict early rectal cancers that can be resected endoscopically? A meta-analysis and systematic review. *Dig Dis Sci* 2010; 55: 1221-9.
- Morris RK, Meller CH, Tamblyn J, et al. Association and prediction of amniotic fluid measurements for adverse pregnancy outcome: systematic review and meta-analysis. *Bjog* 2014; 121: 686-99.
- Higgins JPT, Green S (eds). Cochrane handbook for systematic reviews of interventions [Internet]. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration; 2011 [accessed 23.3.11].
- Beynon R, Leeflang MM, McDonald S, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. The Cochrane database of systematic reviews. 2013; 9: MR000022.
- Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol* 2011; 64: 602-7.
- Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061-6.
- Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997; 315: 617-9.
- Rutjes AW, Reitsma JB, Di NM, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174: 469-76.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-12.
- Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002; 21: 1513-24.
- Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336: 601-5.
- Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. *BMJ* 2011; 343: d5928.
- 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.
- Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17: 1-12.
- Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to meta-analysis. Chichester, UK: Wiley, 2009.
- Simmonds MC, Higgins JP. A general framework for the use of logistic regression models in meta-analysis. *Stat Methods Med Res* 2016; 25: 2858-77.
- Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Stat Methods Med Res* 2013; 22: 133-58.
- Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23: 3105-24.
- Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21: 1559-73.
- Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working G. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; 149: 889-97.
- Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; 56: 455-63.
- Van Howe RS. Circumcision and HIV infection: review of the literature and meta-analysis. *Int J STD AIDS* 1999; 10: 8-16.
- Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529-36.

## Appendice: Glossario

Questa sezione fornisce una panoramica di alcuni dei termini utilizzati in ROBIS con le definizioni relative a come questi termini sono utilizzati all'interno di ROBIS.

Termine	Definizione
<b>Autori di una revisione</b>	Gli autori di una revisione sistematica; a volte semplicemente definiti come "autori".
<b>Autori di overview</b>	Gli autori di una overview.
<b>Banche dati</b>	Banche dati bibliografiche che indicizzano i rapporti degli studi.
<b>Bias</b>	Limiti o distorsioni sistematiche nella progettazione, nell'esecuzione o nell'analisi di una revisione (o di uno studio primario) che falsano i risultati.
<b>Criteri di eleggibilità</b>	I criteri utilizzati per determinare se uno studio dovrebbe essere incluso o escluso da una revisione sistematica. Sono noti anche come criteri di inclusione/esclusione.
<b>Domanda target</b>	La domanda cui i valutatori stanno provando a rispondere, per esempio, nella loro overview o nella loro linea-guida.
<b>Eterogeneità</b>	La variabilità tra gli studi. L'eterogeneità può essere utilizzata in senso generale per descrivere le differenze tra gli studi in termini di popolazione, intervento, confronto, qualità metodologica o disegno dello studio. L'eterogeneità statistica si verifica quando le stime degli effetti tra gli studi differiscono in misura maggiore di quanto ci si aspetterebbe a causa della variabilità del campione (caso).
<b>Fonti elettroniche</b>	Fonti elettroniche (diverse dalle banche dati bibliografiche) utilizzate per individuare report di studi da includere nelle revisioni sistematiche.
<b>Lettori</b>	Le persone che leggono la revisione sistematica.
<b>Linea-guida</b>	Raccomandazioni sviluppate sistematicamente per aiutare i medici e i pazienti a prendere decisioni in merito all'assistenza sanitaria appropriata per specifiche condizioni cliniche. Esse dovrebbero basarsi sulle migliori prove disponibili. Ciò significa che coloro che sviluppano linee-guida devono generalmente includere e valutare in esse le revisioni sistematiche esistenti.
<b>Meta-analisi</b>	Il metodo statistico utilizzato per combinare insieme i risultati di più studi. Viene eseguita calcolando una media ponderata delle stime degli effetti dei singoli studi. Una meta-analisi è spesso condotta come parte di una revisione sistematica, ma non sempre è possibile o appropriato condurla.
<b>Overview</b>	Una revisione di revisioni sistematiche.
<b>Raccolta dati</b>	Il processo di raccolta dei dati dagli studi primari e da altre fonti, talvolta denominato estrazione di dati. Può comportare la raccolta di informazioni da fonti aggiuntive (per esempio contattando gli autori) o la semplice estrazione di dati degli studi primari.
<b>Report</b>	Una sintesi dei metodi degli studi, dei risultati e delle conclusioni; possono riferirsi a studi sia pubblicati sia non pubblicati. Un singolo studio può essere riportato in più report diversi; alcuni singoli report possono riportare più studi.
<b>Revisione</b>	Abbreviazione di revisione sistematica. Può essere utilizzato come termine per indicare una revisione non sistematica ma questo non accade nel contesto di ROBIS.
<b>Revisione sistematica</b>	Un approccio sistematico per riassumere e sintetizzare le prove derivanti dagli studi. Seguono una struttura definita per identificare, valutare e sintetizzare tutte le prove disponibili relative a un particolare quesito di ricerca. Possono includere o meno una sintesi statistica (meta-analisi).
<b>Rilevanza</b>	La misura in cui il quesito della revisione coincide con il quesito cui siete interessati a rispondere.
<b>Significatività statistica</b>	È un valore, chiamato <i>valore p</i> , che indica la probabilità che si verifichi un dato risultato, posto che una determinata affermazione (detta <i>ipotesi nulla</i> ) sia vera. Se il valore <i>p</i> è sufficientemente piccolo, lo sperimentatore può tranquillamente affermare che l'ipotesi nulla è falsa. Spesso viene impostata una soglia per il rifiuto dell'ipotesi nulla del 5%, ma è possibile scegliere qualsiasi altro valore. Quindi, se si trova un valore di <i>p</i> inferiore a 0,05 (o al livello specificato), il risultato può essere considerato statisticamente significativo e l'ipotesi nulla viene rifiutata.
<b>Sintesi</b>	L'analisi, sia quantitativa sia non quantitativa, utilizzata per sintetizzare i risultati degli studi primari inclusi nella revisione sistematica.
<b>Stima cumulativa</b>	Una stima unica dell'effetto ottenuta con la meta-analisi, combinando i risultati di più studi.
<b>Strategia di ricerca</b>	I termini esatti e le loro combinazioni utilizzati per la ricerca nelle banche dati bibliografiche.
<b>Studi</b>	Gli studi primari inclusi nella revisione sistematica. Un singolo studio può essere riportato in più pubblicazioni.