Evidence-based Guideline:

Premature ovarian insufficiency

Integrity assessments using the Research Integrity in Guidelines and evIDence synthesis (RIGID) framework

September 2024

ESHRE, ASRM, CRE WHIRL and IMS Guideline Group on POI

eshre SCIENCE MOVING PEOPLE MOVING SCIENCE
asrm American Society for Reproductive Medicine
CREWHiRL Centre for Research Excellence in Women's Health in Reproductive Life
International IMS Menopause Society Promoting education and research on midlife women's health

Integrity assessments using the Research Integrity in Guidelines and evIDence synthesis (RIGID) framework

Background

Clinical guidelines rely on sound evidence to underpin recommendations for patient care. In recent years, the scientific community has seen an increase in the frequency and/or detection of "problematic studies", the most visible of which are retracted studies [1]. Problematic studies refer to studies with questionable data or findings, which may result from scientific misconduct, plagiarism, poor research practices, or naïve but honest error [1]. Regardless of the cause, erroneous conclusions arising from flawed data can have important and far-reaching consequences, jeopardising the validity of evidence synthesis. This is especially problematic when flawed evidence is used to inform clinical guidelines, which can directly impact patient care. Incorporating integrity assessments into guidelines and broader evidence synthesis processes is therefore critical to ensure the authenticity and accuracy of evidence used in these contexts and to safeguard patient and public trust in the scientific enterprise.

To identify and manage problematic studies encountered during the development of this guideline, we implemented the *Research Integrity in Guidelines and evIDence synthesis* (RIGID) framework [2] - a transparent, unbiased, and rigorous method for incorporating integrity assessments into evidence synthesis processes. This framework has been successfully used to inform prior international guidelines including the International Evidence-based Guidelines in Polycystic Ovary Syndrome [3,4] and the Australian adaptation of the ESHRE Guideline for Unexplained Infertility [5].

Methods

Detailed methodology for the RIGID framework are published elsewhere [2]. Briefly, an integrity committee is formed with a minimum of five members comprising guideline/ project leads and integrity or methodology experts. These members include two independent reviewers who will carry out the integrity assessments. The following 'READER' steps are then applied:

1) **Review:** The framework starts with the search and screening of studies for inclusion as per standard evidence synthesis processes. Once full text screening is complete and a list of eligible studies is determined, integrity assessments commence.

2) **Exclude:** Independent reviewers identify and exclude studies listed on the Retraction Watch database (tabulated with reasons);

3) **Assess:** Independent reviewers assess the remaining eligible studies using a research integrity tool/checklist (e.g., RIA[6], TRACT[7], etc.), and assign an initial risk rating of low, moderate or high. Areas of concern are documented and the scoresheet and studies are circulated to the remaining committee members.

4) **Discuss:** Integrity committee members review the studies and checklist scores/results and a meeting is convened to discuss the issues identified and tally votes to reach a final classification for each study. Studies considered low risk are in the 'included' category and

form part of the evidence synthesis (and meta-analyses where applicable) to inform conclusions/ recommendations.

5) **Establish Contact:** For studies with moderate or high risk, authors are contacted by email to clarify concerns, using a standard template for initial author engagement.

6) **Re-assess:** Per the RIGID author response algorithm [2], studies are reassessed on the basis of author responses. If authors respond and wish to engage in clarifying issues raised on the integrity checklist, the study is in the 'awaiting classification' category. If concerns can be resolved promptly, the manuscript is moved to low risk and included as per step 4. Otherwise, if resolving identified issues will require significant time, beyond the capacity of the project/ guideline, the study remains as 'awaiting classification' pending a resolution, and is not included in the present evidence synthesis. If no response is provided to initial or subsequent contact attempts, the study is moved to the 'not included' category.

All studies are tabulated with integrity scores in technical documents/ supplementary materials for transparency.

Results

Evidence Integrity Committee

An integrity committee was formed to guide the process, the members of which are listed in Table 1. The committee was responsible for investigating and managing integrity issues in the identified literature, and reviewing concerns raised to reach consensus regarding risk ratings and subsequent inclusion.

Table 1. Members of the Integrity Committee in the POI Guideline

Title	Name	Organisation	Country
Associate Professor	Amanda Vincent*	Monash University	Australia
Professor	Melanie Davies	University College London	UK
Professor	Helena Teede	Monash University	Australia
Doctor	Aya Mousa	Monash University	Australia
Doctor	Madeline Flanagan	Monash University	Australia

* Denotes the committee chair.

Application of the RIGID framework

Integrity assessments were conducted using the six READER steps of the RIGID framework, as depicted in Figure 1 and detailed below. The RIGID checklist was used to ensure each step was followed and documented appropriately (Table 3).

Step 1: Review

A total of 85 studies were identified for inclusion in this guideline and compiled for review. Due to time constraints, an evidence hierarchy approach was used, with detailed integrity assessments applied to those randomised controlled trials with pharmacological interventions and/or specifically in POI populations. These were deemed most critical in informing the present guideline recommendations.

Step 2: Exclude

All 85 studies included in the guideline were checked for retractions. One study was identified as having been retracted and was excluded from the guideline (Table 2).

Step 3: Assess

Process

Of the 85 studies, 32 focusing on pharmacological interventions and/or POI populations were screened for integrity by two independent reviewers (MF, AM). One of these was excluded in Step 2 above, leaving 31 for further assessment. For these, we used the Trustworthiness in Randomised Controlled Trials (TRACT) tool [7]. Scores were allocated for each study under each TRACT domain, described in detail elsewhere [7]. Disagreements were resolved by discussion and consensus among the reviewers, and agreed scores were then tallied with an initial risk rating allocated (Table 2).

Initial risk ratings

The majority of studies (n=25) had an initial rating of low risk, six were moderate and none were high risk. The scores table and study documents were circulated to committee members and a two-hour meeting was convened to discuss results.

Step 4: Discuss

Studies were reviewed and discussed among committee members to reach agreement regarding the risk ratings. There was consensus reached on most studies; however, two studies were changed from moderate to high risk following discussion of their integrity concerns. Votes and changes were recorded for transparency (Table 2). Finally, 25 studies remained low risk, four were moderate and two were high risk.

Step 5: Establish contact

Authors of the six moderate and high risk studies were contacted using a template for engaging authors in the integrity assessment process[2]. Over the time period provided (initially two weeks, but ultimately two months between initial contact and guideline launch), only one of the six responded with an intention to engage. Authors of the remaining five studies did not respond to contact attempts.

Step 6: Re-assess

In the single study where authors responded, the study was classified as 'awaiting classification'. However, authors did not respond to further enquiry of integrity issues; thus, the study has not been included in the guideline and remains awaiting classification, pending further clarification of the issues identified (Table 2). The remaining five studies with no response were classified as 'not included' and were not used in the guideline to inform evidence synthesis results or recommendations.

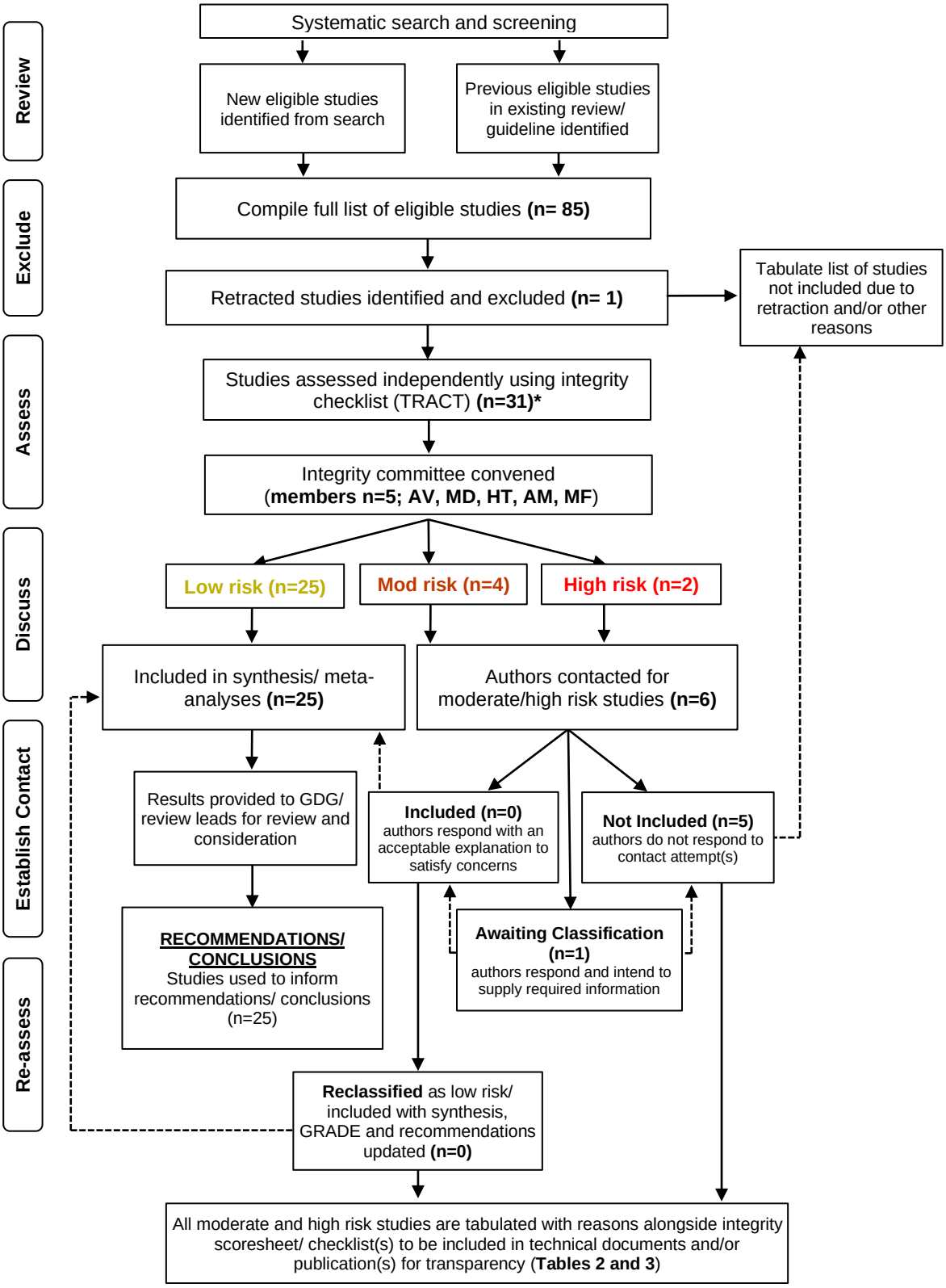

Figure 1. Research Integrity in Guidelines and evIDence synthesis (RIGID) diagram, illustrating the steps followed for assessing research integrity in the guideline/ evidence synthesis. *Due to time constraints, an evidence hierarchy approach was applied focusing on the most critical 31 RCTs which were focused on pharmacological interventions or premature ovarian insufficiency populations. GDG, guideline development group; GRADE, Grading of Recommendations, Assessment, Development and Evaluations; TRACT, Trustworthiness in Randomised Controlled Trials

Table 2. Integrity assessment using the Research Integrity in Guidelines and evIDence synthesis (RIGID) process and Trustworthiness in RAndomised controlled Trials (TRACT) integrity tool.

Author, year	Governance			Author group			Plausibility of intervention		Timeframe			Drop outs		Baseline Characteristics		Outcomes		Total Score (3b.)§	Initial Risk Ranking (3c.)§	Voting Record (4b.)§	Final Consensus Ranking (4c./4d.)§	Final Study Allocation (after author contact) (6a.)§
	Absent or retrospective registration	Discrepant registration	Absent or vague ethics	Low # or ratio of authors	Retraction watch base (2a./ 2b.)§	Large # RCTs	Implausible intervention	Illogical methods	Fast recruitment	Fast follow-up	No LTFU	Ideal numbers	No or few (<5) BL data	Implausible data	Perfectly balanced	Larger effect size than other RCTs	Conflicting outcomes					
Bakarat, 2005	No	No	No	Yes	No	Yes	No	No	No	No	No	No	No	Yes	No	No	No	3	Low	Unanimous x5	Low	Included
Benetti-Pinto, 2020	Yes	No	No	No	Yes	No	No	No	Yes	No	Yes	Yes	No	No	No	No	No	5	Moderate	High (unanimous x5)	High	Not Included
Braunstein, 2005	No	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No	2	Low	Unanimous x5	Low	Included
Buster, 2005	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Cartwright, 2016	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Simon, 2005	No	No	No	No	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No	2	Low	Unanimous x5	Low	Included
Tartagni, 2007	No	No	No	No	No	Yes	No	No	No	No	Yes	Yes	No	Yes	No	No	Yes	5	Moderate	Unanimous x5	Moderate	Not Included
Torres-Santiago, 2013	No	No	No	No	Yes	Yes	No	No	No	No	No	Yes	No	No	No	No	No	3	Low	Unanimous x5	Low	Included
Zuckerman-Levin, 2009	No	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No	1	Low	Unanimous x5	Low	Included
Chernauseak, 2000	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Croften, 2010	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included

Davis, 2006	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Guerrieri, 2014	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Low (same as Popat)	Unanimous x5	Low (same as Popat)	Included
Steingold, 1991	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes	No	No	No	No	2	Low	Unanimous x5	Low	Included
Cleemann, 2017	No	Yes	No	No	No	Yes	No	No	No	No	Yes	Yes	No	No	No	No	No	4	Low	Unanimous x5	Low	Included
Popat, 2014	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Ross, 2003	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Ross, 2011	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Shifren, 2000	No	No	No	No	Yes	Yes	No	No	No	No	Yes	No	No	No	No	No	No	3	Low	Unanimous x5	Low	Included
Panay, 2000	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes	No	No	No	No	2	Low	Unanimous x5	Low	Included
Odonnell, 2008	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	0	Low	Unanimous x5	Low	Included
Langrish, 2009	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Low (same as O'Donnell)	Unanimous x5	Low (same as O'Donnell)	Included
Mittal, 2022	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Mauras, 2007	No	No	No	No	Yes	Yes	No	No	No	No	Yes	No	No	No	No	No	No	3	Low	Unanimous x5	Low	Included
Kenemans, 2009	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No	No	1	Low	Unanimous x5	Low	Included
Cao, 2018	No	Yes	No	No	No	No	No	Yes	No	Yes	No	No	No	No	No	No	No	3	Moderate	Unanimous x5	Moderate	Not Included
Pyri, 2021	Yes	No	No	No	No	Yes	No	No	No	No	Yes	No	No	No	No	No	No	3	Moderate	Unanimous x5	Moderate	Awaiting Classification
Safiyeh, 2021	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	0	Low	Unanimous x5	Low	Included
Shapiro, 2011	No	No	Yes	No	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No	3	Low	x1 mod (AM), x4 low (AV, MF, HT, MD)*	Low	Included
Wxu, 2017	Yes	No	Yes	Yes	No	No	No	No	No	Yes	Yes	Yes	No	No	No	No	No	6	Moderate	x1 mod (MF); 4 high (AM, AV, HT, MD)*	High	Not Included

Wang, 2021	Yes	No	Yes	No	No	No	No	No	No	No	Yes	No	Yes	No	No	No	No	4	Moderate	Unanimous x5	Moderate	Not Included
Yi, 2021	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Retracted- Exclude	Unanimous x6	Retracted - Excluded	Exclude

* Denotes the initials of committee members and their respective votes. [§] Refers to the checklist item for this part of the Research Integrity in Guidelines and evIDence synthesis (RIGID) framework [2]. **BL**, baseline; **LTFU**, lost to follow up; **mod**, moderate; **TRACT**, Trustworthiness in Randomized Controlled Trials tool; **RCTs**, randomized controlled trials. Descriptions of the domains for the tool shown here can be found in Mol et al. [7]

Table 2. The Research Integrity in Guidelines and evIDence synthesis (RIGID) checklist

Item	Description	Page
0. Integrity Committee		
a.	Assembled a multidisciplinary integrity committee (identified in publication(s) and/or supporting documents), comprising a minimum of five members including an impartial chair	Table 1, Pg 2
b.	Nominated two independent reviewers from the committee (identified in publication(s) and/or supporting documents) to conduct initial integrity assessments for each eligible study	MF, AM
Step 1: Review		
1a.	For the clinical question at hand, conducted a systematic search and screening per standard review guidelines (e.g., Cochrane). *- This should include all steps from protocol development through to eligibility screening*	NA
1b.	Compiled a list of all eligible studies following full text screening	Pg 2
Step 2. Exclude		
2a.	Checked all studies for retraction notices and/or on the Retraction Watch Database to identify retracted studies	Pg 3
2b.	Clearly noted studies that are under investigation or have expressions of concern for further assessment	NA
2c.	All retracted studies were identified and recorded as excluded, with the reason listed as 'Retracted'	Table 2
Step 3. Assess		
3a.	Specified the tool used (e.g. TRACT or RIA) by two nominated reviewers to conduct independent integrity assessments for each study and reconcile their ratings through discussion and consensus	Pg 3
3b.	Clearly documented assessments against each domain and an initial rating for each study as low, moderate or high risk of integrity concerns (with notes/justifications where relevant)	Table 2
Step 4. Discuss		
4a.	Integrity checklist assessments and risk ratings were circulated to the committee members with appended publications for review prior to the committee meeting.	NA
4b.	A meeting was convened with all committee members to discuss allocations and record votes and final risk rating after discussion. *Studies may be shifted from one risk rating to another following discussion* *All studies with a final rating of 'low risk' are included in the evidence synthesis* *Where a majority cannot be reached, the Chair decides the final study allocation and this is recorded, with reasons*	Pg 3 and Table 2
Step 5. Establish Contact		
5a.	Sourced contact details and sent a generic email to all corresponding authors of 'moderate risk' and 'high risk' studies to obtain an 'intention to respond' to concerns raised.	Pg 3
5b.	Recorded a log with all authors contacted, noting those who responded (with relevant details of responses), allowing a minimum of two weeks.	Pg 3 and Table 2
Step 6. Re-assess		
6a.	Re-assessed studies following responses (using the RIGID reassessment algorithm) and recorded final allocation as 'Included', 'Not Included' or 'Awaiting Classification'. *If authors are able to satisfy concerns within a reasonable timeframe, studies may be shifted to low risk and included following consultation and agreement by the integrity committee.*	Pg 3
6b.	Continued with subsequent systematic review steps including data extraction and quality appraisal using the final list of those studies which are 'Included'	NA

NA, not applicable (no page number needed for this item); **RIA**, research integrity assessment; **RIGID**, Research Integrity in Guidelines and evIDence synthesis; **TRACT**, Trustworthiness of Randomised Controlled Trials.

References:

1 Boughton, S. L., Wilkinson, J. & Bero, L. When beauty is but skin deep: dealing with problematic studies in systematic reviews. *Cochrane Database of Systematic Reviews* (2021). https://doi.org:10.1002/14651858.ED000152

2 Mousa, A., Flanigan, M., Tay, C.T., Norman, R., Costello, M., Li, W., Wang, R., Teede H., Mol, B. Research Integrity in Guidelines and evIDence synthesis (RIGID): a framework for assessing research integrity in guideline development and evidence synthesis. *EClinicalMedicine* **[in press]** (2024).

3 Teede, H. J. *et al.* Recommendations From the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. *J Clin Endocrinol Metab* **108**, 2447-2469 (2023). https://doi.org:10.1210/clinem/dgad463

4 Mousa, A., Tay, C. T. & Teede, H. Technical Report for the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome: 2023 Update. (2023).

5 Mousa, A., Le Clef, N., Costello, M., Teede, H., Romualdi, D., on behalf of the ESHRE Guideline Group (Baris Ata, Siladitya Bhattacharya, Ernesto Bosch, Samula Dos Santos-Ribeiro, Ksenija Gersak, Roy Homburg, Mina Mincheva, Terhi Piltonen, Sara Somers, Sesh K Sunkara, Harold Verhoeve, Donia Scicluna) and Australian Unexplained Infertility Guideline Group (Robert J Norman, Luk Rombauts, Cindy Farquhar, Lisa Bedson, Marlene Kong, Maree Pickens, Clare Boothroyd, Rebecca Kerner, Rhonda Garad, Trudy Loos, Madeline Flanagan, Ben W Mol), . Technical Report for the Australian Adaptation of the ESHRE Evidence-based guideline for unexplained infertility 2024. (2024). https://doi.org:https://doi.org/10.26180/26299363.v3

6 Weibel, S. *et al.* Identifying and managing problematic trials: A research integrity assessment tool for randomized controlled trials in evidence synthesis. *Res Synth Methods* **14**, 357-369 (2023). https://doi.org:10.1002/jrsm.1599

7 Mol, B. W. *et al.* Checklist to assess Trustworthiness in RAndomised Controlled Trials (TRACT checklist): concept proposal and pilot. *Res Integr Peer Rev* **8**, 6 (2023). https://doi.org:10.1186/s41073-023-00130-8