

## Appendix A. Algorithmic Approach to Evaluating the Risk of Reporting Bias

Figure A-1 presents a framework to identify the risk of reporting bias for a body of evidence for an outcome of interest. The framework is an algorithm that combines considerations of both quantitative and qualitative assessments. Its use is intended to enhance EPC standardization and transparency, such that readers can evaluate how risk of reporting bias judgments were reached. The algorithm has not yet been tested in the process of a systematic review, and modifications based on EPC feedback are expected.

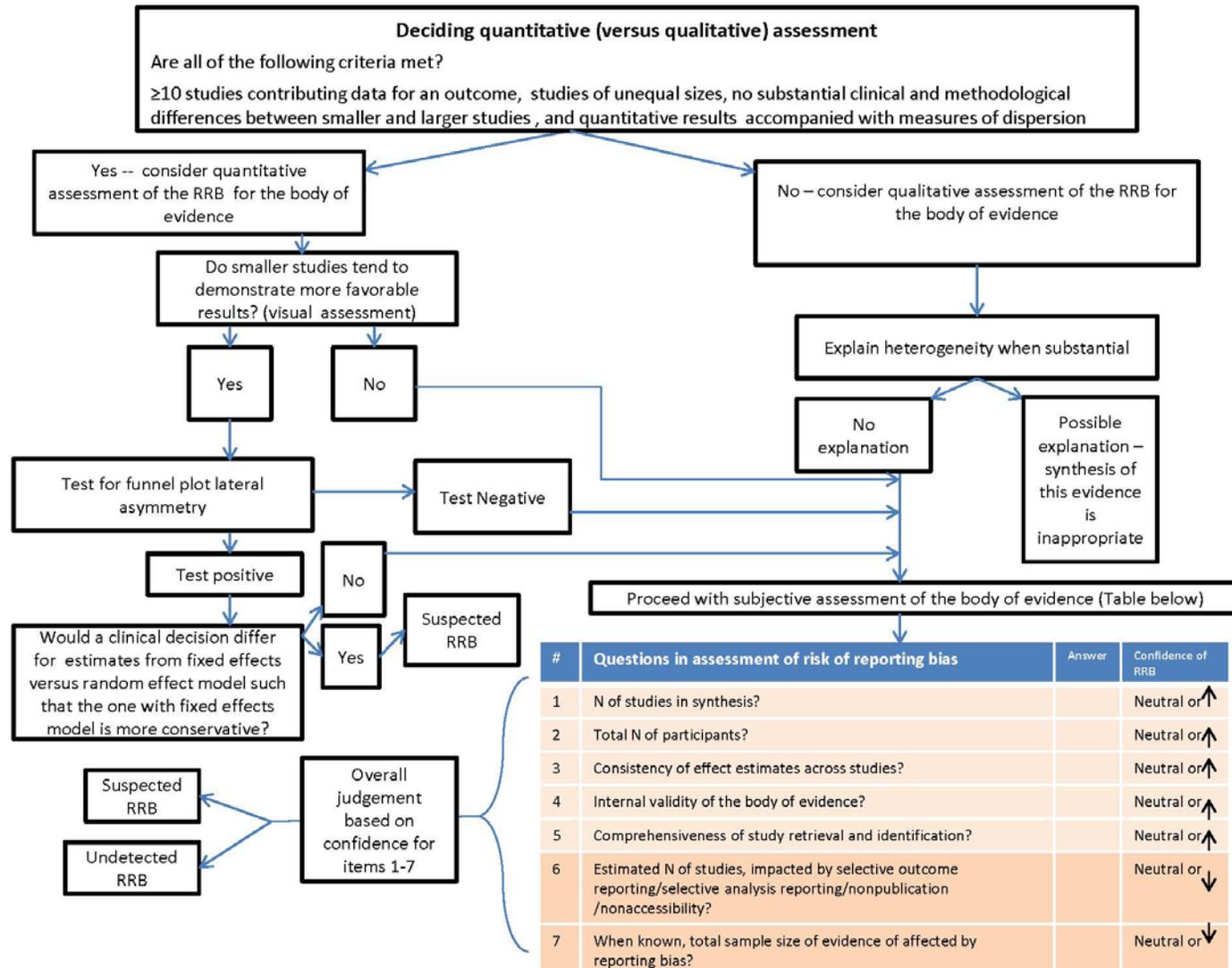
For an outcome, assessment of risk of reporting bias begins with an evaluation of the eligibility for a quantitative assessment of funnel plot asymmetry and how meaningfully different are effect estimates that originate in a random versus a fixed effect model. Statistical tests of funnel plot asymmetry detect whether there is a significant difference in the magnitude and/or direction of effect estimates between smaller versus larger studies.<sup>1, 2</sup>

Because larger studies are more likely to be reported irrespective of their findings than smaller studies, nonpublication of less favorable results from smaller studies will generate a fixed effects estimate that is more conservative (less favorable to the authors) than a random effects estimate because a fixed effects model will essentially reproduce the large study effects. Given that the clinical and methodological diversity are not associated with study size, the only remaining explanation for this would be selective outcome reporting or study nonpublication, as long as an adequate number of studies form the body of evidence. This would be judged as suspected reporting bias.

However, when there are few studies in a body of evidence, the test is underpowered, but when there are a large number of studies, the test becomes overly sensitive.<sup>1</sup> A significant finding could imply: reporting bias; clinical or methodological diversity, or diversity due to both, related to study size; or simply chance. Because of these multiple explanations<sup>2</sup> minimizing alternative explanations are important and funnel plot asymmetry will be useful only in limited situations for detecting reporting bias. A judicious and cautious use of funnel plot asymmetry testing can be helpful in detecting reporting bias by minimizing scenarios when alternative explanation of a significant test are likelier. It requires a sufficient number of studies of unequal sizes, no major concern about the clinical and methodological diversity between smaller and larger studies, and that small study effects are more favorable to the authors than large study effects.

When a quantitative assessment is not possible, a qualitative assessment follows based on the seven items as in Figure A-1. Many studies including many patients and few study limitations in their conduct and design, resulting in relatively consistent results, increase our confidence that a qualitatively or quantitatively synthesized estimate of effect is close to the truth. In such a scenario, with a reasonably adequate search for unpublished data, evidence from a small number of studies suspected of reporting bias, or for which reporting bias is actually detected, but it represents a small proportion of the total patients across studies, may not be judged important enough to question the validity of the synthesized estimate. Reviewers may reasonably decide to judge the overall risk of reporting bias for the body of evidence as unsuspected. In all other scenarios, the risk of reporting bias may be judged as detected.

Figure A-1. Algorithm for detecting reporting bias



A-2

Abbreviations: N = number; RRB = risk of reporting bias

## References

1. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol.* 2000 Nov;53(11):1119-29. PMID: 11106885.
2. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj.* 2011;343:d4002. PMID: 21784880.

## Appendix B. Grading Strength of Evidence: Additional Model Results Tables

**Table B-1. [Intervention A] versus [Intervention B] for the treatment of [Disease]: Strength of evidence domains.**

Outcome	Study Design: No. Studies (N)	Risk of Bias Limitations	Directness	Consistency	Precision	Reporting Bias	Other Issues	Strength of Evidence
<b>Critical outcomes</b>								
Mortality	RCT: 1 (56)	Medium <sup>a</sup>	Indirect <sup>b</sup>	Unknown	Imprecise <sup>c</sup>	Not detected	None	Insufficient
Severity of [Disease]	RCT: 8 (250)	High <sup>d</sup>	Direct	Inconsistent <sup>e</sup>	Imprecise <sup>f</sup>	Suspected <sup>g</sup>	None	Low
<b>Patient-reported outcomes</b>								
Pain	RCT: 6 (160)	Low	Indirect <sup>h</sup>	Inconsistent <sup>i</sup>	Imprecise <sup>f</sup>	Not detected	Different scales	Moderate (direction) Low (magnitude)
Sexual dysfunction	RCT: 3 (85)	Medium <sup>j</sup>	Indirect <sup>k</sup>	Consistent	Imprecise <sup>f</sup>	Not detected	Different scales	Low
<b>Intermediate outcomes</b>								
Hb A1c	RCT: 13 (845)	Low	Direct	Consistent	Imprecise <sup>f</sup>	Suspected <sup>m</sup>	None	Moderate
<b>Adverse Events</b>								
Intestinal perforation	RCT: 1 (42)	Medium <sup>n</sup>	Direct	Unknown	Imprecise <sup>o</sup>	Not detected	None	Insufficient
	Observational: 5 (1100)	High <sup>p</sup>	Direct	Inconsistent <sup>i</sup>	Precise	Not assessed	None	Low
	<i>Overall</i>							
Weight gain	Observational: 4 (600)	Medium <sup>q</sup>	Direct	Inconsistent <sup>i</sup>	Precise	Not assessed	None	Low

a Risk of bias: Moderate (1 study)

b Indirect: mortality assessed only by chart review.

c Imprecise: wide RR CI spanning OIS threshold.

d Risk of bias: high (6 studies), moderate (1 study), low (1 study); in general, lack of outcome assessor blinding and high attrition rates.

e Inconsistent: large differences in direction and magnitude of effects.

f Imprecise: wide RR CIs.

g Outcome reporting bias: inconsistent analyses of single and composite outcomes raises concern about biased outcome reporting.

h Indirect: several studies used nurses' assessment of pain.

i Inconsistent: heterogeneity in estimates of effects.

j Risk of bias: high (2 studies), moderate (1 study).

k Indirect: measures of sexual dysfunction unvalidated.

l Imprecise: the CI of the summary net difference was wide.

m Reporting bias: pooled estimate of small studies larger than pooled estimate of large studies.

n Risk of bias: High (1 study).

o Imprecise: rare events (only single event in trial).

p Risk of bias: High (3 studies), Moderate (1 study), Low (1 study).

q Risk of bias: High (1 study), Low (3 studies).

Abbreviations: NA = not applicable

**Table B-2. [Intervention A] versus [Intervention B] for the treatment of [Disease]: Details regarding strength of evidence domains.**

<b>Outcome</b>	<b>Study Design</b>	<b>Risk of Bias Details</b>	<b>Reasons for Downgrading Domains Descriptions of Other Issues Comments About How Overall Strength of Evidence Derived</b>
Mortality	RCT	1 Moderate	Directness: mortality assessed only by chart review. Precision: wide RR CI spanning optimal information size threshold. Overall: single small study.
Severity of [Disease]	RCT	1 Low, 1 Moderate, 6 High	RoB: lack of outcome assessor blinding and high attrition rates. Consistency: large differences in direction and magnitude of effects. Precision: wide RR CIs. Reporting bias: inconsistent analyses of single and composite outcomes raises concern about biased outcome reporting.
Pain	RCT	3 Low, 2 Moderate, 1 High	Directness: several studies used nurses' assessment of pain. Consistency: heterogeneity in estimates of effects. Precision: wide RR CIs. Other issues: most studies used different pain scales. Overall: the studies all found benefit (direction) so that the consistency and precision domains were considered less important; however, the magnitude of the effect was not precise or consistent across studies, in part due to the use of different scales.
Sexual dysfunction	RCT	2 Moderate, 1 High	Directness: the measures of sexual dysfunction have not been validated. Precision: wide RR CIs. Other issues: the 3 studies each used different pain scales.
Hb A1c	RCT	8 Low, 3 Moderate, 2 High	Precision: wide CI of summary estimate of net difference. Reporting bias: pooled estimate of small studies larger than pooled estimate of large studies.
Intestinal perforation	RCT	1 High	Precision: rare events (only single event in trial).
	Observational	1 Low, 1 Moderate, 3 High	Consistency: heterogeneity in estimates of effects.
Weight gain	Observational	3 Low, 1 High	Consistency: heterogeneity in estimates of effects.

Abbreviations: CI = confidence interval; RoB = risk of bias; RR = relative risk.