

SUPPLEMENTARY MATERIALS

Appendix 1. Methodology of Framework Development

The ISPE CER SIG working group that developed this framework is composed of 14 members representing different stakeholders (academia, policymakers, pharmaceutical product development, health consultants) covering various geographic jurisdictions. The working group met regularly for more than 12 months and leveraged its expertise to develop the current framework using an iterative process.

The specific objectives of our working group were two-fold:

- a) To critically review the existing published evidence covering the following questions:
 1. “How should the quality and compatibility of evidence from NRS and RCTs be assessed, and when is it appropriate to combine evidence from RCTs with NRS?”
 2. “How should NRS and RCT data be combined in a quantitative synthesis to generate reliable comparative effect estimates?”
- b) To provide a step-by step guidance for researchers and policymakers when considering the combination of NRS with RCTs to estimate relative effect estimates for healthcare decision-making.

The development of this framework involved a multi-step process, which began with defining the research questions.

In the next step, a combination of methods was applied to address each of the two research objectives. More specifically, supporting information was retrieved through:

- Systematic literature review (SLR) and update (umbrella of SLRs) on tools assessing validity (extent of susceptibility to bias) in NRS (following a review protocol and database searches)

The systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement [9]. Systematic review protocol and registration are available at <https://osf.io/es65q>.

Systematic search and eligibility criteria

We searched Pubmed and Embase from inception to November 2019 to identify existing tools that investigated the validity of NRS, specifically case-control and cohort design studies. We excluded guidelines or manuals, tools to review study protocols, tools targeting NRS of non-pharmacological interventions (e.g. surgery) or assessing only one or a few specific types of bias, and tools not available in English language. In parallel, we searched the same electronic databases for systematic reviews of assessment tools of NRS. We then extracted the references of the tools included in the systematic reviews retrieved. We also performed a general search through Google® for grey literature and reviewed any additional information from initiatives, programs or organizations, and suggestions from experts. Full details on the search strategy are reported in the Supplement (Table S1 and S2, online Supplement 1). Three reviewers (E.D., G.S., L.V.) independently removed duplicates and reviewed titles and abstracts of peer-reviewed publications or documents from the grey literature to select eligible tools. Discrepancies were resolved by consensus.

Delphi survey and prespecified framework

Concurrently, we performed a Delphi survey to reach a consensus among content experts about the main methodological challenges (domains) that may threaten the validity of NRS on comparative safety and effectiveness of medications. The survey is available in the online Supplement 2. The panel of experts involved members of the SIG for CER of the ISPE. Detailed information on the Delphi methods and results is reported in Supplementary Figure 1.

Domains and subdomains indicated by the Delphi respondents as major elements that can impact the validity of NRS of medications were used to develop and pilot a framework to evaluate the identified NRS tools. All domains were considered equally important.

Data extraction

Two reviewers (E.D., L.V.) independently extracted general information of the identified tools (first author or name of the tool, year of publication or online availability of the most updated version, type of tool, scope of the tool, non-

randomized study designs evaluated, number of items) and content data related to the prespecified domains of the framework. Discrepancies were resolved by consensus. We categorized the tools as checklists, defined as itemized instruments (including questionnaires) developed to identify the presence or absence of critical elements, or rating scales, defined as itemized instruments aimed to identify the performance of a study at each critical element described in the tool, using a qualitative or quantitative scale.

Data synthesis

General characteristics of the identified tools were summarized with means and standard deviations, for continuous variables, and relative frequencies, for categorical variables. The findings from the online survey and the proportion of tools assessing the prespecified elements of the framework were reported in terms of relative frequencies.

- Identification of publications from previous SLRs to address methodological issues and statistical analysis approaches for combining NRS with RCTs (e.g., Innovative Medicines Initiative [IMI] GetReal, Institute for Clinical and Economic Review [ICER], Duke-Margolis Health Policy Center)
- Pragmatic identification of relevant materials based on the group's knowledge and prior experience (supplementary online searches)

Supplementary Table 1. Search Strategy for the Systematic Literature Review of Quality Tools for NRS

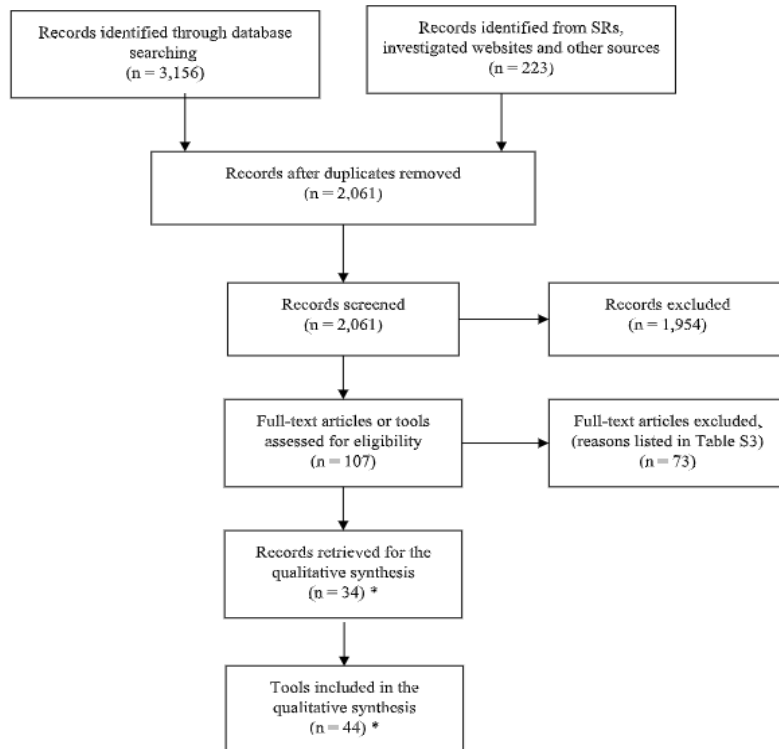
Items	N.	Terms
Critical appraisal tool	#1	"critical" [All Fields] AND "appraisal" [All Fields] AND "tools" [All Fields]
	#2	"critical" [All Fields] AND "appraisal" [All Fields]
	#3	("critical" [All Fields] AND "review" [All Fields]) OR "critical review" [All Fields] AND form [All Fields]
	#4	("systematic review" [Publication Type] OR "systematic reviews as topic" [MeSH Terms] OR "systematic review"[All Fields]) AND form [All Fields]
	#5	appraisal [All Fields] AND ("research design" [MeSH Terms] OR ("research" [All Fields] AND "design" [All Fields]) OR "research design" [All Fields] OR ("research" [All Fields] AND "methodology" [All Fields]) OR "research methodology"[All Fields])
	#6	("research design" [MeSH Terms] OR ("research" [All Fields] AND "design" [All Fields]) OR "research design" [All Fields] AND ("review" [Publication Type] OR "review literature as topic" [MeSH Terms] OR "review"[All Fields])
Study reporting tool	#7	"study" [All Fields] AND "reporting" [All Fields] AND "tool" [All Fields]
	#8	"study" [All Fields] AND "reporting" [All Fields]

	#9	"reporting" [All Fields] AND "form" [All Fields] AND ("Studies"[Journal] OR "studies"[All Fields])
	#10	"reporting" [All Fields] AND ("Studies"[Journal] OR "studies"[All Fields])
Tool	#11	"checklist" [MeSH Major Topic] OR "scale*" [Title/Abstract]
	#12	"surveys and questionnaires"[MeSH Major Topic] OR "questionnaire*" [Title/Abstract]
	#13	("tool*" [All Fields] OR "instrument*" [All Fields] OR "checklist*" [All Fields] OR "questionnaire*" [All Fields]) AND ("quality" [All Fields] OR "method*" [All Fields] OR "bias" [All Fields])
Study design	#14	"cohort studies"[MeSH Terms] OR cohort studies [Text Word] OR cohort stud* [All Fields]
	#15	"case-control studies" [MeSH Terms] OR case-control studies [Text Word] OR case control stud* [All Fields]
	#16	Non [All Fields] AND ("random allocation"[MeSH Terms] OR randomized [Text Word]) AND stud* [All Fields]
Systematic review	#17	"systematic review" [Publication Type] OR "systematic reviews as topic"[MeSH Terms] OR "systematic review"[All Fields]
Filters	#18	"humans"[MeSH Terms]
	#19	"Review" [ptyp] OR "systematic" [sb]
Strings		
1st search - tools*	#20	(#1 OR #2 OR #3 OR #4 OR #5 OR #6) AND (OR #14 OR #15 OR #16) AND #18
	#21	(#7 OR #8 OR #9 OR #10) AND OR #14 OR #15 OR #16) AND #18
	#22	(#11 OR #12 OR #13) AND OR #14 OR #15 OR #16) AND #18
2nd search - systematic reviews of tools*	#23	(#1 OR #2 OR #3 OR #4 OR #5 OR #6) AND (OR #14 OR #15 OR #16) AND #18 AND (#17 OR #19)
	#24	(#7 OR #8 OR #9 OR #10) AND OR #14 OR #15 OR #16) AND #18 AND (#17 OR #19)
	#25	(#11 OR #12 OR #13) AND OR #14 OR #15 OR #16) AND #18 AND (#17 OR #19)

*tools are defined as instruments (e.g., qualitative checklists, questionnaires, scoring scales, etc.) that investigate the overall quality of a study, identifying potential biases – either used to critically appraise studies included in a systematic review, or to help in the peer-reviewing process of scientific publications – (i.e., critical appraisal tools), or that support the reporting of research methods and findings (i.e. study reporting tools). Note: the strings were built in Medline and then adapted to Embase (through Elsevier)

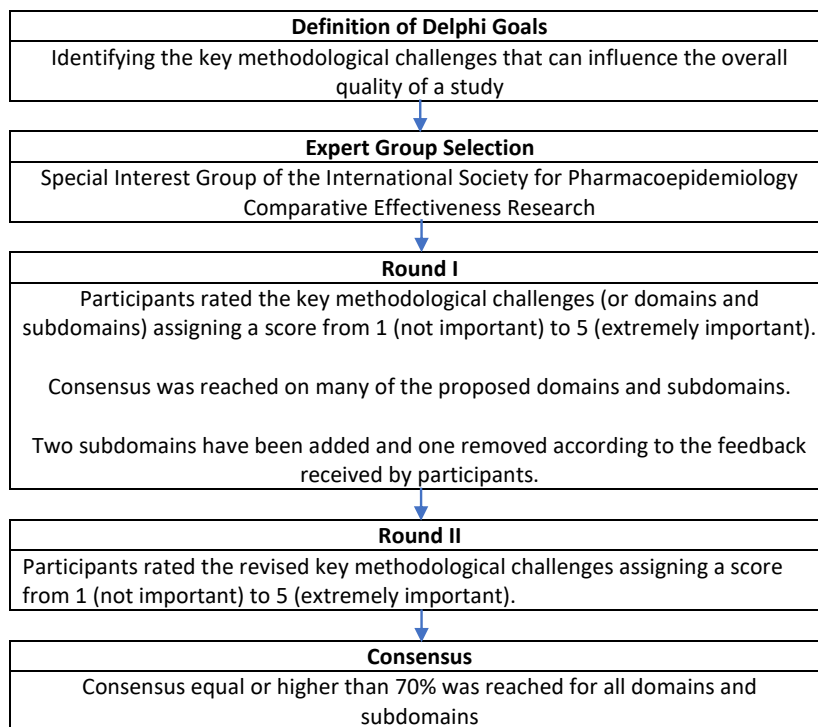
Supplementary Figure 2. PRISMA

Figure S2. Flowchart of the selection of the tools.

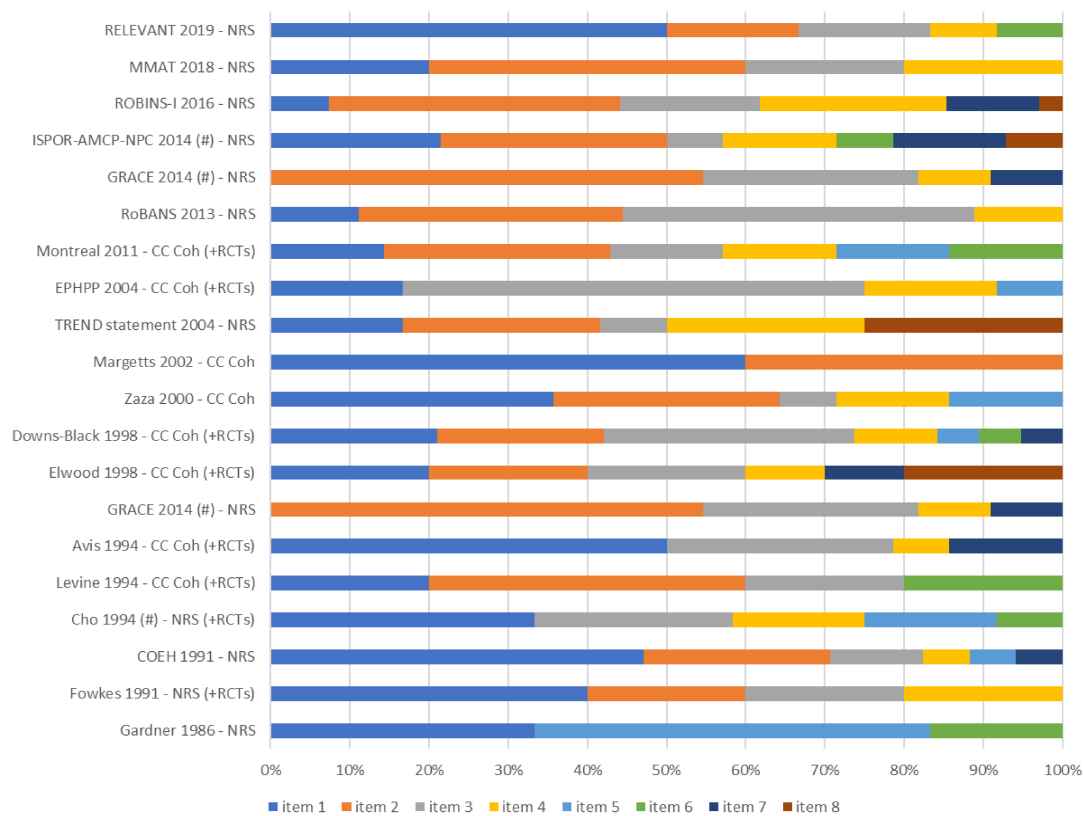


*Ten retrieved articles/records included two different assessment tools for case-control and cohort studies. Thus, the overall number of included tools raised up to 44.

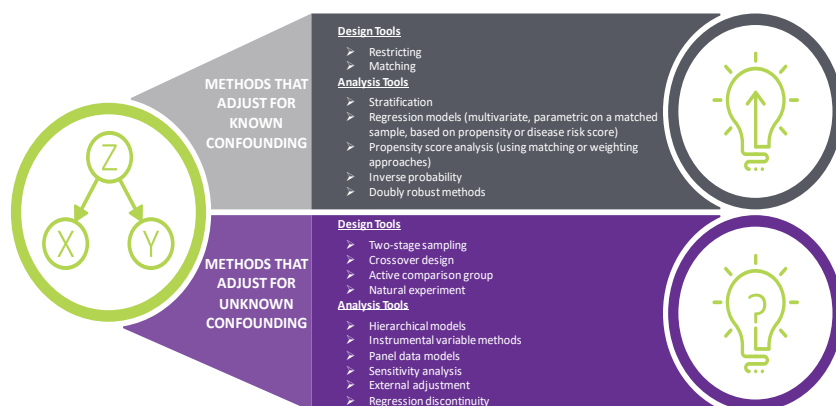
The systematic reviews (SRs) were retrieved through a systematic search. Additional details about the SRs search are available upon request.

Supplementary Figure 1. Flowchart of the Delphi procedure

Supplementary Figure 3. Distribution of domains for each reviewed tool of multiple study designs (NRS)



Legend: Item 1 = Methods for selecting participants; item 2 = Definition and measurement of exposure, outcomes, covariates and follow-up; item 3 = Design-specific sources of bias; item 4 = Confounding; item 5 = Lack of appropriateness of statistical analyses; item 6 = Methods for assessing statistical uncertainty in the findings; item 7 = Methods for assessing internal validity; item 8 = Methods for assessing external validity.

Supplementary Figure 4. Summary of Methods to Adjust for Either Known or Unknown Confounding

Sources: RWE navigator (<https://rwe-navigator.eu/use-real-world-evidence/adjusting-for-bias-in-non-randomised-and-observational-studies/>) and Duke Margolis Center for Health Policy (https://healthpolicy.duke.edu/sites/default/files/atoms/files/non-intervetional_study_credibility_final.pdf)

Supplementary Table 2. Overview of the Presented Approaches for Combining RCTs with NRS Evidence (source: adapted by Efthimiou et al. 2016)

	Using informative priors	Three-level hierarchical models	Bias-adjusted analysis
	Scenario 1	Scenario 2	Scenario 2 and 3
Direct meta-analysis of RCTs and NRS	No	Yes	Yes
Description of the approach	Prior distributions are formulated by meta-analyzing NRS	Data from NRS and RCTs are synthesized separately and then the pooled effect estimates are pooled in a joint meta-analysis.	NRS estimates are adjusted for possible bias and over-precision
How NRS are incorporated	The priors are shifted to account for bias, and/or the variances are inflated to down-weight estimates from NRS. Between study variability (RCTs, NRS) in treatment effect is ignored.	Either NRS can be adjusted separately (according to its features) or adjustment for bias can be performed collectively for each design (on the design-level estimates).	Either NRS can be adjusted separately (according to its features) or common bias parameters can be assumed for all NRS
When to use it (preferably)	When it is infeasible to infer about bias in each study separately	When there are studies pertaining to multiple study designs (RCTs, NRS)	When resources allow inference about bias in each separate NRS

Abbreviations: NRS = non-randomized studies; RCT = randomized controlled trial