Supplement: Scoping review of software implementations of risk-of-bias tools and quantitative bias analysis methods for selection bias

1. Search strategy

We describe in detail our search strategies of published articles available on Web of Science. ,R packages listed on the Comprehensive R Archive Network (CRAN), and inbuilt and user-written Stata commands.

1.1 Published literature search

The Web of Science search query consisted of three substages: 1.1) include articles that mention (in the title, abstract or keywords) terms relating to "software" or "developing a method" along with terms relating to "selection bias" and "bias analysis" or "risk-of-bias tools" (Box 1); 1.2) Exclude articles with titles containing terms relating to "meta-analysis"

or “systematic review” and articles published in the Cochrane Database of Systematic reviews (Box 2); 1.3) Add back in articles, excluded from substage 1.2, that mention in the title terms relating to “meta-analysis” or “systematic review” and “bias” or “method” (Box 3). The Search was performed from 1 Jan 2004 to 31 Aug 2025. The exact search queries alongside weblinks are provided in Section 2.1.1. The process of assessing the identified abstracts and selecting and assessing the full-length article texts is described in section 2.1., alongside intermediate results for each step. A detailed overview of our selection process is shown in the Flowchart in Supplementary Figure 1.

Box 1: Search terms; including articles that mention (in the title, abstract or keywords) terms relating to “software” or “developing a method” along with terms relating to “selection bias” and “bias analysis” or “risk-of-bias tools”.

((TS=((“sensitivity analys*s” OR “bias analys*s” OR (“risk of” NEAR/2 bias) OR “risk-of-bias”) AND ((selection NEAR/4 bias) OR (collider NEAR/4 bias) OR “collider*” OR (sample NEAR/2 selection) OR (nonignorable NEAR/3 selection) OR (non-ignorable NEAR/3 selection) OR (informative NEAR/3 selection) OR “volunteer bias” OR “self-selection bias” OR “self selection bias” OR “healthy worker” OR “participation bias” OR (nonparticipation NEAR/3 bias) OR (non-participation NEAR/3 bias) OR “study nonparticipation” OR “study non-participation” OR (nonresponse NEAR/3 bias) OR (non-response NEAR/3 bias) OR “Berkson’s” OR “prevalence-incidence bias” OR “Neyman** bias” OR “M-bias” or “M bias”))) AND TS=(((“propose*” OR “develop**” OR “introduce*” OR “provide”* OR “outline**”) NEAR/3 (“method*” OR “approach” OR “model” OR “procedure”)) OR “derive*” OR ”code" OR "software" or "tool*” OR "script*” OR "calculator" OR “package" OR "implementation" OR “SAS macro*” OR "command*” OR “Stata program*” OR “R function*” OR “checklist” OR “spreadsheet” OR

“Excel” OR “web-tool” OR “online program**” OR “web program**” OR “computer program**” OR “tutorial”))

Box 2: Search terms: excluding articles with titles containing terms relating to “meta-analysis” or “systematic review” by adding the following terms to Box 1; and manual exclusion of articles published in the Cochrane Database of Systematic reviews

NOT TI=("meta-analys*s" OR "meta analys*s" OR "systematic review*") [MANUAL] *Refine menu > Publication Titles > COCHRANE DATABASE OF SYSTEMATIC REVIEWS >* ***Exclude***

Box 3: Search terms; Adding back in articles, excluded from substage 1.2 (Box 2), that mention in the title terms relating to “meta-analysis” or “systematic review” and “bias” or “method”, by adding the following terms to Box 1.

AND TI=(("meta-analys*s" OR "meta analys*s" OR "systematic review*") AND ("bias" OR "method" OR "approach" OR "model" OR "procedure")))

1.2 R package software search

We conducted two searches of the CRAN database. The first search was conducted using the automated topic search function *findPackage()* from R package *packagefinder* (R code in Box 4). The second search was conducted by comparing all available CRAN package titles and descriptions to our search terms using function *cran_packages()* from R package *pkgsearch* (R code in Box 5).

Box 4: R code of search using R package *findPackage.* The object 'subsearch1' contains packages with description terms relating to sensitivity or (risk of) bias analysis, while 'subsearch 2' contains packages with terms relating to selection bias: our target is the intersect of subsearch1 and suchsearch2.

```
subquery1 <- c("sensitivity analys", "bias analys", "risk of", "risk-of-bias")

subquery2 <- c("selection bias","collider", "sample selection",

               "nonignorable selection", "non-ignorable selection",

               "informative selection", "volunteer bias", "self-selection bias",

               "self selection bias", "healthy worker", "participation bias",

               "nonparticipation bias", "non-participation bias",

               "study nonparticipation", "study non-participation",

              "nonresponse bias", "non-response bias", "Berkson's",

             "prevalence-incidence bias", "Neyman", "M-bias", "M bias")

subsearch1 <- findPackage(query=subquery1,return.df=TRUE)

subsearch2 <- findPackage(query=subquery2,return.df=TRUE)

search_results <- merge(subsearch1,subsearch2[,1:2],by="Name")
```

Box 5: R code of search using R package *pkgsearch*

```
availpkgs <- available.packages()

# Run a loop to obtain details for all packages as the cran_packages command can
# only evaluate a finite number of packages in one call

availpkgdesc <- data.frame(Package="",Title="",Description="")

for (i in 0:45){

tmp <- cran_packages(availpkgs[eval(1+(500i)):eval(500+(500i)),"Package"])

availpkgdesc <- rbind(availpkgdesc,tmp[,c("Package","Title","Description")])

}

tmp <- cran_packages(availpkgs[22501:nrow(availpkgs),"Package"])

availpkgdesc <- rbind(availpkgdesc,tmp[,c("Package","Title","Description")])

availpkgdesc$Desc <- gsub("\n"," ",availpkgdesc$Description)

# Repeat search as defined in Box 1

subsearch1b <- data.frame(Package="",Title="",Desc="")

for (i in 1:length(subquery1)){
```

```
tmp1 <- availpkgdesc[grep(subquery1[i],availpkgdesc[,"Title"], ignore.case=TRUE),]
tmp2 <- availpkgdesc[grep(subquery1[i],availpkgdesc[,"Desc"], ignore.case=TRUE),]
subsearch1b <- rbind(subsearch1b,tmp1[,c(1,2,4)],tmp2[,c(1,2,4)])
}
#Remove dups
subsearch1b <- subsearch1b[-1,]
subsearch1b <- dplyr::distinct(subsearch1b[order(subsearch1b$Package),])

subsearch2b <- data.frame(Package="",Title="",Desc="")
for (i in 1:length(subquery2)){
tmp1 <- availpkgdesc[grep(subquery2[i],availpkgdesc[,"Title"], ignore.case=TRUE),]
tmp2 <- availpkgdesc[grep(subquery2[i],availpkgdesc[,"Desc"], ignore.case=TRUE),]
subsearch2b <- rbind(subsearch2b,tmp1[,c(1,2,4)],tmp2[,c(1,2,4)])
}
#Remove dups
subsearch2b <- subsearch2b[-1,]
subsearch2b <- dplyr::distinct(subsearch2b[order(subsearch2b$Package),])

search_resultsb <- merge(subsearch1b[,1:2],subsearch2b[,c(1,3)],by="Package")
```

1.3 Stata software search

We searched software documents listed in the IDEAS bibliographic database using the IDEAS/RePrec search tool [1] with the search query listed in Box 6. Also, we used Stata's inbuilt *search* command [2] with option *manual* to search the Stata manuals (version 9 (StataCorp, 2025)) (Box 7) and option *net* to search the user-written commands available via Stata's *net* command (Box 8) .

Box 6: Search query entered into IDEAS/RePec search tool

("sensitivity analysis" | "sensitivity analyses" | "bias analysis" | "bias analyses" | "risk of bias" | "risk-of-bias") + (selection | collider | volunteer | "healthy worker" | participation | nonparticipation |

non-participation | nonresponse | non-response | Berkson | prevalence-incidence | Neyman | M-bias | "M bias")

Box 7: Search of Stata manuals using Stata *search* command

```
search bias analysis, manual
search bias analyses, manual
search sensitivity analysis, manual
search sensitivity analyses, manual
search risk of bias, manual
search sample selection, manual
search selection bias, manual
search collider bias, manual
search nonresponse bias, manual
search non-response bias, manual
search Berkson, manual
search prevalence-incidence bias, manual
search Neyman bias, manual
search M-bias, manual
search M bias, manual
```

Box 8: Search of net resources using Stata *search* command

```
search bias analysis, net
search bias analyses, net
search sensitivity analysis, net
search sensitivity analyses, net
search risk of bias, net
search sample selection, net
search selection bias, net
search collider bias, net
search nonresponse bias, net
search non-response bias, net
search Berkson, net
search prevalence-incidence bias, net
search Neyman bias, net
```

2. Search results

2.1. Literature Search

Overview Literature search (1085 articles with search terms)

We performed a search targeting title, abstract and keyword using terms related to sensitivity and risk of bias analysis, terms related to selection bias, and terms related to method and software development (Section 2.1.1, Step 1). This yielded a total of 1085 articles. Additionally, we performed a search excluding method and software terminology, to investigate if articles are likely to omit mentions of (software) implementation in the abstract and only provide details in the full text. This yielded an

additional 1663 articles (Section 2.1.1, “broad search”). We examined the full texts of a random draw of 50 articles and found no relevant mentions of sensitivity analysis and risk-of-bias software implementations and tool, suggesting that the initial, more restrictive search, was appropriate, and we continued with our initial selection of 1085 articles. Of these, 674 were excluded based on article title or publication title (Section 2.1.1, Steps 2, 3, 4), or for being duplicates or conference abstracts (Section 2.1.1, Step 5). The remaining 411 abstracts were independently assessed by two researchers, yielding a total of 82 candidate articles which were thought to be likely to contain information on suitable software or tools. Of these, 14 were included in our final review, with a further 13 articles identified through reference searches. Of the 27 articles, 3 articles described user-friendly software programs for a quantitative risk-of-bias analyses, while 24 provided tools on a qualitative assessment of risk of selection bias. Full detail on the selection process is given below.

Abstract selection process (1085 articles to 411)

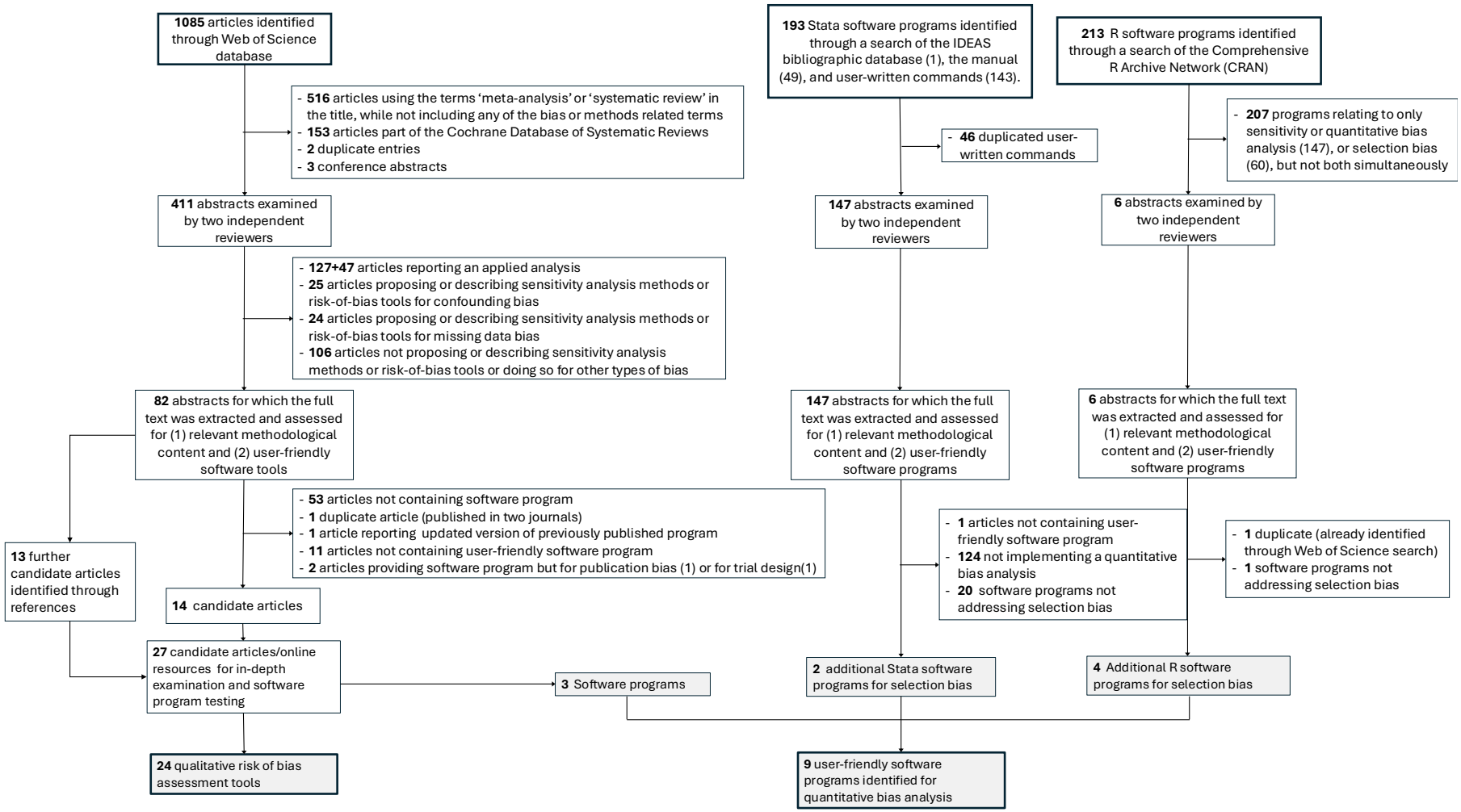
Our objective was to identify methodological and software focussed articles. An overview of the selection process is given in Figure 1. Our search, however, also captured applied systematic reviews and meta-analyses that reported investigating risk of selection bias in the abstract. To that end, we performed a filtering step in which we excluded all articles mentioning meta-analysis and systematic review in the title (-544), unless these terms were used in conjunction with bias or method related terms (+28), which excluded **516** of the 1085 articles, reducing the total number of articles to 569. To account for applied systematic reviews that did not explicitly use ‘systematic review’ in the title, we additionally excluded all articles with the publication title ‘Cochrane Database of Systematic reviews’, which resulted in the removal of an additional **153** articles. Of the remaining 416 articles, **2** were duplicates, and **4** were conference abstracts, and were consequently excluded, yielding a total of 411 abstracts. (Specific search terms given in Section 2.1.1, steps 1 to 3).

Selecting articles suitable for full-paper examination (411 articles to 82)

411 abstracts were examined for suitability by two researchers, independently. Exclusion criteria included articles that exclusively reported an applied analysis, and methodological articles that focussed on sensitivity analyses for other bias, such as missing data and confounding bias. Of the 411 abstracts, **174** were excluded on the basis of reporting an applied analysis, **25** and **24** articles were excluded for focussing on sensitivity analysis/risk-of-bias methods related to confounding and missing data, respectively, and a further **106** articles were excluded that did not propose sensitivity analysis/risk-of-bias methods at all, or did so for bias not resulting from confounding and missing data. For the remaining **82** articles, the abstracts either clearly indicated that the article contents contained information on sensitivity analysis/risk-of-bias methods for selection bias or were sufficiently ambiguous that such content remained a distinct probability. The full texts of these articles were examined to determine if relevant methodology was described and if user-friendly software for implementation was provided.

Identifying articles describing relevant sensitivity analyses methods/risk-of-bias tools for selection bias that provide suitably user-friendly software/tools (82 articles to 15, with an additional 14 articles identified through references)

The full text of the **82** articles was examined. Of these, 5 were found not to contain any software/tool relating to selection bias, **11** were found not to contain any user-friendly software/tool by our definition, **1** article reported on an updated version of a software package that was described in an earlier publication also identified in our search, **1** article did provide a method with software, but on closer inspection it was found to be intended for publication bias, and **1** article provided a method related to allocation bias, but targeted the design stage rather than the analysis stage of a trial. In total, detailed methods and software were extracted for **14** of the **82** articles. **13** additional articles or online tools were identified through references. In total, **27** candidate articles/online resources were identified. Note that **24** of these described qualitative risk-of-bias tools (e.g., Cochrane RoB tool), and only **3** described user-friendly software for a quantitative risk of bias analysis.



Supplementary Figure 1. Detailed flowchart of search strategy for quantitative bias analysis software and qualitative risk-of-bias tools in published literature (Web of Science), the Comprehensive R archive Network (CRAN), and the Stata IDEAS bibliographic database, manual, and user-written commands.

2.1.1. Literature search: search terms for each step of the abstract selection process

Search performed from 1 Jan 2004 to 31 Aug 2025.

Note that Web of Science is a dynamic database and that due to database updates the same query for the same time window may return slightly different numbers on rerunning the search.

Broad search for quality check: 3038: Include terms related to sensitivity analysis/risk of bias and selection bias

(TS=((“sensitivity analys*s” OR “bias analys*s” OR (“risk of” NEAR/2 bias) OR “risk-of-bias”) AND ((selection NEAR/4 bias) OR (collider NEAR/4 bias) OR “collider*” OR (sample NEAR/2 selection) OR (nonignorable NEAR/3 selection) OR (non-ignorable NEAR/3 selection) OR (informative NEAR/3 selection) OR “volunteer bias” OR “self-selection bias” OR “self selection bias” OR “healthy worker” OR “participation bias” OR (nonparticipation NEAR/3 bias) OR (non-participation NEAR/3 bias) OR “study nonparticipation” OR “study non-participation” OR (nonresponse NEAR/3 bias) OR (non-response NEAR/3 bias) OR “Berkson’s” OR “prevalence-incidence bias” OR “Neyman** bias” OR “M-bias” or “M bias”))

https://www.webofscience.com/wos/woscc/summary/19428e49-2a94-4b18-aad4-c97a3665bbc3-0190051384/author-ascending/1

STEP 1: 1085 Additionally include terms related to methods/software (1953 abstracts excluded: 3038-1953= **1085**)

((TS=((“sensitivity analys*s” OR “bias analys*s” OR (“risk of” NEAR/2 bias) OR “risk-of-bias”) AND ((selection NEAR/4 bias) OR (collider NEAR/4 bias) OR “collider*” OR (sample NEAR/2 selection) OR (nonignorable NEAR/3 selection) OR (non-ignorable NEAR/3 selection) OR (informative NEAR/3 selection) OR “volunteer bias” OR “self-selection bias” OR “self selection

bias” OR “healthy worker” OR “participation bias” OR (nonparticipation NEAR/3 bias) OR (non-participation NEAR/3 bias) OR “study nonparticipation” OR “study non-participation” OR (nonresponse NEAR/3 bias) OR (non-response NEAR/3 bias) OR “Berkson’s” OR “prevalence-incidence bias” OR “Neyman** bias” OR “M-bias” or “M bias”))) AND TS=(((“propose*” OR “develop**” OR “introduce*” OR “provide”* OR “outline**”) NEAR/3 (“method*” OR “approach” OR “model” OR “procedure”)) OR “derive*” OR ”code" OR "software" or "tool*” OR "script*” OR "calculator" OR “package" OR "implementation" OR “SAS macro*” OR "command*” OR “Stata program*” OR “R function*” OR “checklist” OR “spreadsheet” OR “Excel” OR “web-tool” OR “online program**” OR “web program**” OR “computer program**” OR “tutorial”))

https://www.webofscience.com/wos/woscc/summary/9f7e6353-39a0-4160-be76-3d74450fe46c-01900520b2/author-ascending/1

Quality check: *We randomly drew 50 articles that were excluded on going from the broad search results with 3038 articles to step 1 to verify that nothing worthwhile was being excluded.*

STEP 2: 541 Exclude all articles that mention meta-analysis or systematic review in the title: (see caveat in step 4) (544 excluded: 1085-544=541).

((TS=((“sensitivity analys*s” OR “bias analys*s” OR (“risk of” NEAR/2 bias) OR “risk-of-bias”) AND ((selection NEAR/4 bias) OR (collider NEAR/4 bias) OR “collider*” OR (sample NEAR/2 selection) OR (nonignorable NEAR/3 selection) OR (non-ignorable NEAR/3 selection) OR (informative NEAR/3 selection) OR “volunteer bias” OR “self-selection bias” OR “self selection bias” OR “healthy worker” OR “participation bias” OR (nonparticipation NEAR/3 bias) OR (non-participation NEAR/3 bias) OR “study nonparticipation” OR “study non-participation” OR (nonresponse NEAR/3 bias) OR (non-response NEAR/3 bias) OR “Berkson’s” OR “prevalence-incidence bias” OR “Neyman** bias” OR “M-bias” or “M bias”))) AND TS=(((“propose*” OR

“develop**” OR “introduce*” OR “provide”* OR “outline**”) NEAR/3 (“method*” OR “approach” OR “model” OR “procedure”)) OR “derive*” OR ”code" OR "software" or "tool*” OR "script*” OR "calculator" OR “package" OR "implementation" OR “SAS macro*” OR "command*” OR “Stata program*” OR “R function*” OR “checklist” OR “spreadsheet” OR “Excel” OR “web-tool” OR “online program**” OR “web program**” OR “computer program**” OR “tutorial”) **NOT** TI=("meta-analys*s" OR "meta analys*s" OR "systematic review*"))

https://www.webofscience.com/wos/woscc/summary/5f3faa1e-78f9-4515-a2b6-5efce8104913-0190066bb0/author-ascending/1

STEP 3: 388 Exclude all articles with publication title ‘Cochrane Database of Systematic reviews’ (153 excluded: 541-153=388)

https://www.webofscience.com/wos/woscc/summary/a28d6a06-7004-4104-ba69-d2d7e3875176-01900688d1/author-ascending/1

STEP 4: 416 Include all articles that mention meta-analysis or systematic review in the title in conjunction with bias or method related terms (28 additional articles included: 388 + 28 = 416) (((TS=((“sensitivity analys*s” OR “bias analys*s” OR (“risk of” NEAR/2 bias) OR “risk-of-bias”) AND ((selection NEAR/4 bias) OR (collider NEAR/4 bias) OR “collider*” OR (sample NEAR/2 selection) OR (nonignorable NEAR/3 selection) OR (non-ignorable NEAR/3 selection) OR (informative NEAR/3 selection) OR “volunteer bias” OR “self-selection bias” OR “self selection bias” OR “healthy worker” OR “participation bias” OR (nonparticipation NEAR/3 bias) OR (non-participation NEAR/3 bias) OR “study nonparticipation” OR “study non-participation” OR (nonresponse NEAR/3 bias) OR (non-response NEAR/3 bias) OR “Berkson’s” OR “prevalence-incidence bias” OR “Neyman** bias” OR “M-bias” or “M bias”))) AND TS=(((“propose*” OR “develop**” OR “introduce*” OR “provide”* OR “outline**”) NEAR/3 (“method*” OR “approach”

OR “model” OR “procedure”)) OR “derive*” OR ”code" OR "software" or "tool*” OR "script*” OR "calculator" OR “package" OR "implementation" OR “SAS macro*” OR "command*” OR “Stata program*” OR “R function*” OR “checklist” OR “spreadsheet” OR “Excel” OR “web-tool” OR “online program**” OR “web program**” OR “computer program**” OR “tutorial”)) AND TI=(("meta-analys*s" OR "meta analys*s" OR "systematic review*") AND ("bias" OR "method" OR "approach" OR "model" OR "procedure")))

https://www.webofscience.com/wos/woscc/summary/fc5d4409-ab4a-4deb-9fdf-7fce9b40d4ff-df098c3b/author-ascending/1

STEP 5: 411 Exclude duplicate entries (2) and conference abstracts (3): (5 excluded: 416-5=411)

2.2. R software search

Using “packagefinder” we identified 153 packages relating to sensitivity or (risk of) bias analyses, 66 packages relating to selection bias, and only five packages relating to both. Using “pkgsearch” identified one additional package.

Of the six packages, two were excluded: 1 software package had already been identified through the Web of Science published literature search; and 1 did not address selection bias.

2.3. Stata Search

We searched the SSC Boston archive using IDEAS/RePrec search, which yielded 1 result:

Giovanni Cerulli (2018): SENSIMATCH: Stata module to provide data-driven **sensitivity analysis** for Matching estimator

sensimatch provides a **sensitivity** test for checking the robustness of the **selection**-on-observables assumption in treatment effect observational studies, both within a regression adjustment and a

propensity-score matching approach. Rooted in the machine learning literature,

this **sensitivity analysis** is based on a "leave-one-covariate-out" (LOCO) approach.

This result was excluded as it does not address selection bias.

The manual keyword search of the Stata documentation, Stata journal, software available via Stata's *net* command, and Stata's keyword database, yielded two additional results, which we report in the main text.

References

StataCorp. (2025). *Stata Statistical Software*. In (Version Release 19) StataCorp LLC.

3. Supplementary Tables

Supplementary Table S1. Qualitative tools for risk of bias assessment: additional details (companion table to Table 1 in the main text)

Checklist/tool	Authors	Target	Level of guidance	Outcome	Additional notes
Checklist for drug AEs (from: A new risk of bias checklist applicable to randomized trials, observational studies, and systematic reviews was developed and validated to be used for systematic reviews focusing on drug adverse events)	Faillie *et al*. (2017)	RCTs and observational studies of drug adverse events.	Provides detailed guidance and specific examples for answering signalling questions	Contains 4 signalling questions relating to selection bias for RCTs, 3 for cohort studies, 2 for case-control studies.	
Checklist for operative interventions (from: How to assess applicability and methodological quality of comparative studies of operative interventions in orthopedic trauma surgery	Luiken *et al*. (2022)	Observational studies and RCTs for operative interventions in orthopedic trauma surgery	Provides detailed guidance and specific examples for answering signalling questions	Contains 4 signalling questions relating to selection bias in RCTs, three of which are given in the domain 'Confounding'. Contains 2 signalling questions relating to selection bias in observational studies, given in the joint domain 'Missing data and selection bias'. Per signalling question a rating of yes/no/probable yes/probable no/not informative. Per question, per rating, guidance is given for interpreting the rating as good/moderate/poor methodology.	Draws on: RoB-2, ROBINS-I and MINORS

JBI checklist for analytical cross-sectional studies (available from the JBI website, no associated publication)	JBI	Analytical cross-sectional studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 1 signalling question relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	
JBI checklist for case/control studies. (available from the JBI website, no associated publication)	JBI	Case/control studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 2 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	
JBI checklist for case series (from: Methodological quality of case series studies: an introduction to the JBI critical appraisal tool)	Munn *et al*. (2020)	Case series	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 3 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	
JBI checklist for cohort studies (available from the JBI website, no associated publication)	Barker *et al*. (2025)	Cohort studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 1 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	
JBI checklist for diagnostic accuracy studies (from: Diagnostic test accuracy: methods for systematic review and meta-analysis)	Cambell *et al*. (2015)	Diagnostic accuracy studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 3 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	Takes its signalling questions related to selection bias from QUADAS-2
JBI checklist for prevalence studies (from: Methodological guidance for systematic reviews of observational epidemiological	Munn *et al*. (2020)	Prevalence studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 2 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	

studies reporting prevalence and cumulative incidence data					
JBI checklist for quasi-experimental studies (from: The revised JBI critical appraisal tool for the assessment of risk of bias quasi-experimental studies)	Barker *et al.* (2024)	Quasi-experimental studies (non-randomised)	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 2 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	
JBI checklist for RCTs (from: The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials)	Barker *et al.* (2023)	RCTs	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 3 signalling questions relating to selection bias. Per signalling question, yes/no/unclear/not applicable option.	
MINORS (from: Methodological index for non-randomized studies (minors): development and validation of a new instrument)	Slim *et al.* (2003)	Non-randomised studies	Little guidance and no examples for assessing specific signalling questions.	Contains 2 signalling questions (here called 'Methodological items for non-randomized studies'), scored 0 (unreported), 1 (reported but inadequate) and 2 (reported and adequate)	
NOS (from: The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses)	Wells *et al.* (2009)	Cohort studies and case-control studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 4 signalling questions each for cohort studies and case-control studies. Each signalling question has multiple choice answers with a star indicating the one or more acceptable responses. Risk of bias is rated from 0 to 4 stars, with 4 stars indicating low risk of bias.	
NOS-xs (from: Risk of Bias Evaluation of Cross-Sectional Studies: Adaptation of the Newcastle-Ottawa Scale)	Carra *et al.* (2025)	Analytical cross-sectional studies and descriptive cross-sectional studies (i.e.	Provides guidance for each signalling question. A case study applying the tool is included. An empty template of the tool is available.	Contains 1 signalling question with four multiple choice answers, two of which are marked with a star as acceptable responses. The selection bias element is identical for the	Adaptation of Wells' *et al.* (2009) Newcastle-Ottawa scale

		prevalence studies)		analytical and descriptive cross-sectional study tools	
OHAT (available from: https://ntp.niehs.nih.gov/research/assessments/noncancer/riskbias)	US National Toxicology Program's Office of Health Assessment and Translation (2015)	Experimental Animal and Human controlled trials, cohort, case control and cross-sectional studies, case series/reports	Provides guidance and specific examples for each signalling question. Extensive explanation of bias rating is given for each question, per study type.	Contains a selection bias domain with 2 signalling questions for randomised controlled trials and 1 signalling question for cohort, case control and cross-sectional studies. Each question can be rated as definitely/probably low/high risk of bias.	The OHAT tool was developed based on amongst others guidance from the ROBINS-1 tool for non-randomised studies and SYRCLE tool for animal intervention studies. Note that no selection bias questions are defined for case series/reports.
QUADAS-2 (from: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies)	Whiting *et al*. (2011)	Diagnostic accuracy studies	Provides guidance and specific examples for each signalling question. An empty template of the tool is available.	Contains 3 signalling questions within the selection bias domain. Per signalling question, an answer of yes, no, or unclear, with an overall rating of low/high/unclear risk of bias. (no algorithm for adding up)	Note that currently a revised version of the tool, QUADAS-3, is in development.
Risk-of-bias tool for urban planning surveys (from: Toward Evidence-Based Urban Planning	Ravensbergen & El-Geneidy (2023)	Non-randomised (?) urban	Provides guidance and examples through an application	Contains two signalling questions relating to selection bias. Multiple bias domains are assessed	

Integrating Quality Assessments in Literature Reviews)		planning survey studies		individually (weak/moderate/strong) and combined into risk score of overall bias. Check signalling questions	
RoBANS 2 (Risk of Bias Assessment Tool for Nonrandomized Studies of Interventions)	Seo *et al*. (2023)	Non-randomised studies of interventions	Provides detailed guidance on criteria for each domain, which consists of a single signalling question, with criteria for low/high risk of bias per type of non-randomised study (cohort, cross-sectional, case-control, before-after)	Contains two domains relating to selection bias. Per domain (consisting of a single signalling question): a rating of bias risk of low, high or unclear	
RoB-spv (Risk of Bias – symptom and performance validity	Puente-López *et al.* (2025)	Psychology studies of symptom and performance validity tests (PVTs/SVTs) using simulation or criterion-group designs to classify genuine vs feigned presentations	Provides detailed guidance for each signalling question.	No dedicated selection bias domain. Signalling questions are grouped by study design. There are 4 selection bias SQs specific to simulation design – where participants are instructed to feign; 3 SQs specific to criterion group designs – participants identified as feigning based on validated PVTs/SVTs. An additional set of 4 SQs is given for studies that include a healthy clinical comparison group. 1 SQ is specific to simulation design studies with a clinical comparison group, with the remaining 3 SQs applicable to both studies. Each question is rated" yes, probably yes, probably no, no or uninformative.	

RoB-2 (Revised Cochrane Risk of Bias tool for randomized trials)	Sterne *et al*. (2019)	RCTs	Provides extensive guidance and specific examples for answering signalling questions. Provides clear guidance and/or an algorithm for combining evidence from multiple signalling questions per domain. The tool itself and an empty template of the tool are available separately.	Contains 3 signalling question in the domain 'Risk of bias arising from the randomization process'. An algorithm is provided for a rating of low risk/some concerns/ high risk.	RoB-2 is a Cochrane Collaboration tool. There are also separate tools (test versions) available for cluster and cross-over RCTs, and separate guidance for RCTs of tobacco cessation interventions; these have selection bias domains equivalent to RoB-2.
ROBINS-E Risk of Bias in Non-randomised Studies of Exposures	Higgins *et al*. (2024)			Contains five signalling questions in the domain 'Bias in selection of participants into the study'. Per signalling question, an answer of yes, probably yes, probably no, no or uninformative. Per domain guidance is provided for a rating of overall bias risk of low, moderate, serious, critical, or non informative.	ROBINS-I and ROBINS-E tools are Cochrane Collaboration tools. Note that while ROBINS-I and ROBINS-E are separate tools, their selection bias
ROBINS-I Risk of Bias in Non-randomised Studies of Interventions	Sterne *et al*. (2016)	Non-randomised studies of interventions/ exposures.			

					domains are equivalent.
RoB-PrevMH (Risk of bias in studies measuring the prevalence of mental health disorders)	Tonia *et al.* (2023)	Prevalence studies	Provides guidance and specific examples for each signalling question.	Contains 2 signalling questions within the selection bias domain. Per signalling question, an assessment of high/low/unclear risk.	Note that the signalling questions in the selection bias domain are applicable to prevalence studies outside mental health disorders
RoB-SPEO (Risk of Bias in Studies estimating Prevalence of Exposure to Occupational Risk factors)	Pega *et al.* (2020)	Prevalence studies on exposure to occupational risk factors	Provides detailed guidance on criteria for each domain, which consists of a single main signalling question, alongside examples of direct and indirect evidence	Contains two domains relating to selection bias. Per domain (consisting of a single signalling question): a rating of bias risk of low to high	The RoB-SPEO tool developed new components and integrated existing ones from various risk-of-bias tools, including ROBINS-I and ROBINS-E.
SYRCLE (Systematic Review centre for Laboratory Animal Experimentation)	Hooijmans *et al.* (2014)	Animal intervention studies	Provides guidance and specific examples for answering signalling questions	Contains three signalling questions relating to selection bias in the domain 'Bias due inadequate randomisation and lack of blinding'. Note that there are two additional questions present in this domain, which	For the SYRCLE tool, the Cochrane Collaboration RoB-2 Tool, was used as a starting point,

				have been excluded, as they pertain to differential treatment response *after* treatment initiation, which is a source of bias distinct from selection bias. Per signalling question, a rating of yes/no/unclear	with the selection bias domain closely mirroring that of the RoB-2 tool

Supplementary Table S2. Qualitative tools for risk of bias assessment: signalling questions an interpretation per study design.

Signalling question	**Interpretation**	**Tool/checklist**
Signalling question as included in the tool/checklist. Questions have been grouped according to study design and, within study design, on similarity.	Elaboration on how the signalling question relates to the presence of selection bias	Source of the signalling question (further detail on tools given in Supplementary Table S1)
Randomised controlled trials		
Are all the subjects recruited from the same source population? Is there comparability of intervention groups, or are appropriate methods applied to correct for incomparability?	Subjects recruited from different source population may result in a sample non-representative of the target population. However, if randomisation is adequate, this should not result in systematic differences between treatment groups.	Checklist for drug AEs Checklist for operative interventions (Luiken)
Was the allocation sequence random? Was true randomisation used for assignment of participants to treatment groups? Was the allocation sequence adequately generated and applied? Was the method used to generate the allocation sequence adequate as to produce comparable groups?	Inadequate randomisation may result in systematic differences between treatment groups.	RoB-2, Checklist for operative interventions (Luiken) JBI Checklist for RCTs SYRCLE's RoB tool Checklist for drug AEs
Was the allocation sequence concealed until participants were enrolled and assigned to interventions? Was allocation to treatment groups concealed? Was administered dose or exposure level adequately randomized? Was allocation to study groups adequately concealed? Was the allocation to the different groups adequately concealed during? Was the method used to conceal the allocation sequence adequate as to produce comparable groups?	Knowledge of allocation prior to treatment may result in crossovers prior to treatment initiation and result in systematic differences between treatment groups. Note that while this may also affect response to treatment, this is separate from selection bias.	RoB-2, Checklist for operative interventions (Luiken) JBI Checklist for RCTs OHAT RoB rating tool SYRCLE's RoB tool Checklist for drug AEs
Did baseline differences between intervention groups suggest a problem with the randomization process? Are baseline characteristics and prognostic factors comparable between different groups? Were treatment groups similar at baseline?	Differences in baseline/prognostic variables between groups may indicate inadequate randomisation.	RoB-2, Checklist for operative interventions (Luiken) Checklist for drug AEs JBI Checklist for RCTs SYRCLE's RoB tool

Were the groups similar at baseline or were they adjusted for confounders in the analysis?		
Non-randomised studies of the effects of interventions or exposures		
Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If likely yes: - Were the post-intervention/exposure variables that influenced selection likely to be associated with intervention/exposure? - Were the post-intervention/exposure variables that influenced selection likely to be influenced by the outcome or a cause of the outcome? Was selection of participants into the study (or into the analysis) based on variables measured after the start of the intervention?	Selection that is (indirectly) conditional on exposure/ intervention or outcome may result in systematic differences between treatment/comparison groups and, consequently, selection bias	ROBINS-I and ROBINS-E Checklist for operative interventions (Luiken)
Do start of follow-up and start of intervention/start of the exposure window coincide for most participants?	If they do not coincide, then patients with events prior to start of follow-up may be excluded.	ROBINS-I, ROBINS-E, Checklist for operative interventions (Luiken)
Inclusion of consecutive patients: all patients potentially fit for inclusion (satisfying the criteria for inclusion) have been included in the study during the study period	If inclusion criteria are not applied consistently, then (1) the sample may be non-representative of the target population and (2), if the inclusion process differs across comparison groups, this may result in systematic differences.	MINORS
Baseline equivalence of groups: the groups should be similar regarding the criteria other than the studied endpoints. Absence of confounding factors that could bias the interpretation of the results		
Comparability of the target group: selection bias due to the selection of an inappropriate comparison target group Target group selection: selection bias due to inappropriate intervention or inappropriate selection of exposure group or patient	Different selection process across groups may result in different measurable characteristics across groups. Observing such differences may indicate selection bias.	RoBANS 2
Were the participants included in any comparisons similar?		JBI Checklist for Quasi-experimental studies
Was there a control group?		
Did selection of study participants result in appropriate comparison groups?	Groups will not be comparable if patients (exposed and non-exposed, or cases and controls) are recruited from different	OHAT RoB Rating tool

	populations, under different inclusion criteria, within different time frames and have different participation rates, and (for case/control studies) controls have a history of the outcome	
Were adjustment techniques used that are likely to correct for the presence of selection biases?	Selection bias may be corrected for by using methods such as inverse probability weighting. This, however, requires knowledge of the selection mechanism and is rarely done in practice.	ROBINS-I and ROBINS-E
Cohort studies		
Were the two groups similar and recruited from the same population?	Recruiting from different populations may result in systematic differences between groups, as will implementing differential inclusion criteria, which may be apparent from different measurable characteristics across groups.	JBI Checklist for cohort studies
Are baseline characteristics and prognostic factors comparable between different groups?		Checklist for drug AEs
Are all the subjects recruited from the same source population?		
Were inclusion and exclusion criteria implemented uniformly across study groups?		
Are exposed subjects new users of the drug?	When the study target includes prevalent instead of new users of the drug, patients with events prior to start of follow-up may be excluded.	
Representativeness of the exposed cohort [Is the exposed cohort representative of the exposed individuals of the community?]	The exposed subjects in the cohort should be representative of the individuals in the community who are likely to be exposed. The non-exposed cohort should be drawn from the same population and representative of this same community.	Newcastle-Ottawa Scale
Selection of the non-exposed cohort [Was the non-exposed cohort drawn from the same community as the exposed individuals?]		
Ascertainment of exposure [Is the ascertainment of exposure adequate?]	There should be good evidence that the exposed subjects been correctly identified as exposed, e.g. by means of secure records or through a structured interview.	
Demonstration that outcome of interest was not present at the start of study [Was the outcome of interest not present at the start of study?]	If the outcome is present at the start of study this may affect exposure	
Case control studies		

Are all the subjects recruited from the same source population?	Subjects recruited from different source population may result in a sample non-representative of the target population. If cases and controls originate from different source populations, or if implementation of inclusion criteria differs across study groups, this will result in systematic differences between groups. This may be apparent from differences in measurable characteristics across groups. Cases should ideally be independently validated to ensure they have been correctly identified as cases (e.g., repeated independent extraction of information, or a reference to a primary source like a medical record). Definition of control should account for outcome history, clearly indicating if cases are first occurrences (so that controls with outcome history are excluded) or new occurrences of the outcome. If e.g., cases are new occurrences but controls have experiences occurrences in the past, the groups are not comparable.	Checklist for drug AEs
Is the origin of controls clearly specified? Selection of controls [Is the control series used in the study derived from the same population as the cases?]		Newcastle-Ottawa Scale
Were inclusion and exclusion criteria implemented uniformly across study groups? Were the same criteria used for identification of cases and controls?		Checklist for drug AEs JBI Checklist for case control
Were the groups comparable other than presence of disease in cases or absence of disease in controls?		JBI Checklist for case control
Representativeness of cases [Are the cases representative? (i.e., have all cases within a certain catchment period or time been included?)]		Newcastle-Ottawa Scale
Is the case definition adequate?		
Definition of controls [Is the control definition adequate?]		
Analytical cross-sectional studies		
Were the criteria for inclusion in the sample clearly defined?	If inclusion criteria are not clearly defined it maybe unclear if the sample is representative of the population	JBI Checklist for analytical cross-sectional studies
Representativeness of the study sample [truly representative/somewhat representative/not representative/no description of study strategy]		NOS-xs
Prevalence studies		
Were study participants recruited in an appropriate way?	An inappropriate sample frame, an inappropriate inclusion/exclusion protocol, or an inappropriate recruitment strategy may result in a sample non-representative of the target population and/or systematic differences between comparison groups, if the sampling (or	JBI Checklist for prevalence studies
Was the sample frame appropriate to address the target population?		
Was the sample invited to participate in the study a true or close representation of the target population?		RoB-PrevMH
Was the sample that provided data a true or close representation the sample invited to participate?		

Could the exposure status (or level) assessed (or assigned) in the study sample not represent exposure in the target population?	exclusion) process affects the groups differently. This may occur when participants opt out of the study due to exposure-related reasons or if the investigator's selection process is (un)knowingly influenced by exposure level.	RoB-SPEO
Could study personnel have known the exposure status (or level) or other characteristics of study participants and, if yes, could this knowledge have influenced how they conducted the exposure assessment (or assignment)?		
Case series		
Were there clear criteria for inclusion in the case series?	If inclusion criteria are not applied consistently, then (1) the sample may be non-representative of the target population and (2), if the inclusion process differs across comparison groups, this may result in systematic differences.	JBI Checklist for case series
Did the case series have consecutive inclusion of participants?		
Did the case series have complete inclusion of participants?		
Survey		
Are selected individuals representative target population?	Selected individuals may be not representative of the target population because of an inappropriate sample frame or due to lack of participation. If the latter is plausibly related to either exposure or outcome, this is particularly concerning when the participation rate is low (<60%).	RoB urban planning survey
What percentage agreed to participate?		
Diagnostic accuracy studies		
Was a consecutive or random sample of patients enrolled?	A diagnostic accuracy study should enrol a consecutive or random sample of patients with suspected disease; excluding patients who are difficult to diagnoses and enrolling participants with confirmed diagnoses (cases) may result in overestimation of diagnostic accuracy.	QUADAS-2, JBI checklist for diagnostic accuracy studies
Was a case–control design avoided?		QUADAS-2, JBI checklist for diagnostic accuracy studies
Did the study avoid inappropriate exclusions?		QUADAS-2, JBI checklist for diagnostic accuracy studies
Symptom and performance validity studies		

Have the simulators/malingerers been assigned to the experimental group(s) using a valid method?	Selection bias may occur if participants are not assigned to the various simulation/experimental groups using a valid method (e.g., randomisation), so that all groups are no longer representative of the same underlying population.	ROB-spv (simulation design-specific domain)
Has a potential familiarity of the instructed participants with the condition to be simulated been taken into account?	If simulators are familiar with the condition, they may become 'expert simulators' who are unrepresentative of the naive feigners the SPV test aims to detect. (Note: This does not apply if the experimental group is intentionally designed to represent expert feigners).	
Did the study include an explicit pre-experimental compliance check (i.e. a check of understanding of instructions from the participants)?	Pertaining to a simulation design SPV study: the 'simulator' group should include a selection of participants who are appropriately feigning the condition. A compliance check serves to exclude participants who may not be able to simulate. Inclusion of such participants may result in a simulator group that is not representative of the target feigning population,	
Was the pre-experimental check designed in such a way that non-compliance would lead to exclusion from the study?		
Has a clinical reference sample (i.e. a patient sample) been included and well described?	Omitting real patients may result in comparison groups that are more distinct than those in real life clinical settings; using only healthy controls results in a sample not representative of patients with actual deficits, overestimating the test's ability to differentiate genuine from feigning patients.	ROB-spv (simulation designs that include a clinical comparison group; all criterion group designs)
Is the sampling method of the clinical group appropriate?	Since probability sampling is not possible in this field, consecutive sampling must be used. Relying on convenience or archival data results in a group not representative of the target population	

Do the patients included match the target population?	For example, if the study is intended to evaluate PTSD feigning, patients with a diagnosis of PTSD should be included. Relying on questionnaires or exposure (e.g., trauma) rather than established strict diagnostic criteria (e.g., DSM) may result in a sample not representative of the target clinical population (i.e. a comparison group that is 'too healthy').	
Is the control group absent of any overt or considerable incentive and was symptom/performance validity assessed separately?	Failing to screen for external incentives may result in the inclusion of non-credible responders in the 'genuine' group. The resulting sample is then not representative of a truly credible population, making the 'genuine' group more similar to the 'feigning' group.	
Is the criterion standard (or classification criteria) likely to correctly identify possible invalid performance/noncredible symptom report?	If the criterion used to classify patients into credible and non-credible patients ('feigners') performs poorly, the resulting groups will not be representative of the target populations of credible and non-credible patients.	ROB-spv (criterion group design-specific domain)
Have participants classified as undetermined been controlled for their effect?	Excluding or mismanaging participants classified as indeterminate creates an artificially polarized sample, with groups that are not representative of actual clinical patients.	
Did all patients receive the same criterion standard?	If the criteria for entering the credible or non-credible patient groups vary across the sample, inconsistent classification will make it impossible to tell if differences in test performance are due to the different selection rules rather than the participants' actual credibility.	