medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SeroTracker-RoB: an approach to automating reproducible risk of bias assessment of seroprevalence studies

Niklas Bobrovitz*[1,2], Kim Noël*[3], Zihan Li[4], Christian Cao[5], Gabriel Deveaux[6], Anabel Selemon[5], David A. Clifton[7], Mercedes Yanes-Lane[8], Tingting Yan[1], Rahul K. Arora[6,7]

*Contributed equally as co-first authors

Affiliations:

1. Temerty Faculty of Medicine, University of Toronto
2. Department of Critical Care Medicine, University of Calgary
3. Faculty of Medicine and Health Sciences, McGill University
4. Faculty of Engineering, University of Waterloo
5. Cumming School of Medicine, University of Calgary
6. Centre for Health Informatics, University of Calgary
7. Institute of Biomedical Engineering, University of Oxford
8. COVID-19 Immunity Task Force, McGill University

Corresponding author:
Niklas Bobrovitz
niklas.bobrovitz@mail.utoronto.ca
Medical Sciences Building, 1 King's College Cir
Toronto, ON M5S 1A8

Declaration of interests
RKA was previously a Technical Consultant for the Bill and Melinda Gates Foundation Strategic Investment Fund, is a minority shareholder of Alethea Medical, and was a former Senior Policy Advisor at Health Canada. Each of these relationships is entirely unrelated to the present work. DAC reports consulting fees from Sensyne Health, Oxford University Innovation, and BioBeats, each outside the submitted work. MYL has received grants from the Bill and Melinda Gates Foundation, not pertaining to the present study. No other authors have conflicts of interest to report.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ABSTRACT

Background: Evaluating seroprevalence study risk of bias (RoB) is crucial for robust infection surveillance, but can be a time-consuming and subjective process. We aimed to develop decision rules for reproducible RoB assessment and an automated tool to implement these decision rules.

Methods: We developed the SeroTracker-RoB approach to RoB assessment. To do so, we created objective criteria for items on the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Prevalence Studies and developed decision rules for RoB based on these items. The criteria and decision rules were based on published guidance for assessing RoB for prevalence studies and expert opinion. Decision rules were validated against the SeroTracker database of seroprevalence studies, which included consensus manual RoB judgements from two independent reviewers. We measured efficiency by calculating paired-samples t-test for time to judge RoB using the automated tool versus manually for 25 randomly selected articles from the SeroTracker database, coverage as the proportion of database studies where the decision rules could evaluate RoB, and reliability by calculating intraclass correlations between automated and manual RoB assessments.

Results: We established objective criteria for seven of nine JBI items. We developed a set of decision rules with 61 branches. The SeroTracker-RoB tool was significantly faster than manual assessment with a mean time of 0.80 vs. 2.93 minutes per article ($p<0.001$), classified 100% (n = 2,070) of studies, and had good reliability compared to manual review (intraclass correlation 0.77, 95% confidence interval 0.74 to 0.80). The SeroTracker-RoB Excel Tool embeds this approach in a simple data extraction sheet for use by other researchers.

Conclusions: The SeroTracker-RoB approach was faster than manual assessment, with complete coverage and good reliability compared to two independent human reviewers. This approach and tool enable rapid, transparent, and reproducible evidence synthesis of infection prevalence studies, and may support public health efforts during future outbreaks and pandemics.

Keywords: automation; decision rule; risk of bias; critical appraisal; evidence synthesis; prevalence; infectious disease

Running title: Automating risk of bias for seroprevalence studies

Word count: 3,131

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

What is new?

- **What is already known:** Risk of bias assessments are a core element of evidence synthesis but can be time consuming and subjective. As such, there is a need for validated and transparent tools to automate such assessments, particularly during disease outbreaks and pandemics to inform public health decision making. However, there are currently no automated tools for risk of bias assessment of prevalence studies.
- **What is new:** We developed a reproducible approach to risk of bias assessment for SARS-CoV-2 seroprevalence studies. The automated approach was five times faster than manual human assessment, successfully categorized all 2,070 studies that it was tested on, and had good agreement with manual review. We built a simple Excel tool so that other researchers can use this automated approach.
- **Potential impact:** The SeroTracker-RoB approach and tool enables rapid, transparent, and reproducible risk of bias assessments for SARS-CoV-2 seroprevalence studies, and could be readily adapted for other types of disease prevalence studies. This process may also be applicable to automation of critical appraisal and risk of bias assessment for other types of studies and in other scientific disciplines.

INTRODUCTION

Seroprevalence studies use antibody tests to estimate the prevalence of infection or vaccination.[1] These studies have been used for decades to measure the true extent of infection,[2] quantify protection resulting from previous infection or vaccination, and inform public health measures.

During the COVID-19 pandemic, there has been an unprecedented increase in the utilization of seroprevalence studies, with results reported from over 3,000 such studies as of Feb 6, 2022.[2] These studies have made important contributions to the pandemic response, but their methods and quality have varied widely.[3,4] Accordingly, robust risk of bias (RoB) assessments have been crucial for synthesizing and utilizing trustworthy seroprevalence data for public health decision-making.

There are several tools for RoB assessment of prevalence studies but there is no consensus on a definitive tool for use in evidence synthesis.[5] Some tools have been validated, but even these have heterogeneous evaluation criteria, are time-consuming to use, and yield RoB assessments with inherent variation given a reliance on evaluator subjective judgments.[6–8] These limitations present a major challenge for rapid and reproducible synthesis of seroprevalence studies. Meta-epidemiological reviews of prevalence studies demonstrate that a unified approach to RoB assessment is needed.[5]

The Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Prevalence Studies is a validated tool that is the most commonly used for prevalence studies and therefore could form the basis for such a unified approach.[5] However, the JBI tool provides a nine-item checklist with no provision for overall RoB assessment. As a result, this checklist has been operationalized

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

differently by groups conducting evidence synthesis, with each effort defining their own approach to derive overall RoB assessments.[9–11] In some instances this has resulted in markedly different RoB assessments for the same body of underlying literature.[9,12]

We aimed to develop an objective, rapid, and reproducible RoB assessment tool for seroprevalence studies. In this manuscript, we describe the SeroTracker-RoB approach, which involves standardization of the JBI checklist and automation of overall RoB assessments using decision rules. We evaluated the efficiency, coverage, and reliability of this approach, and developed a tool that implements the approach for use by other researchers.

METHODS

The SeroTracker-RoB approach

We developed an approach for automating RoB assessments of seroprevalence that involved two components: (1) a seroprevalence specific version of the JBI Critical Appraisal Checklist for Prevalence Studies,[13] which can be automatically completed based on extracted data from a given seroprevalence study; (2) decision rules that can be applied to the JBI checklist ratings to generate an overall RoB assessment.

Part 1: Adapting and automating the JBI checklist

The JBI checklist was selected as the foundation for the RoB assessment as it is a validated and commonly used critical appraisal tool for prevalence studies. The original JBI checklist involves reviewers judging nine categorical items based on data reported in a prevalence study: (1) sample frame appropriateness; (2) sampling method; (3) sample size/calculation; (4) reporting of age/sex; (5) representativeness of sample within analysis; (6) test sensitivity and specificity; (7) consistent test use; (8) appropriate statistical adjustment; and (9) response rate.[13]

We adapted the JBI checklist to seroprevalence studies (the "seroprevalence JBI"), developing reviewer guidance for each criterion to make them more specific to the seroprevalence context, improve reproducibility, and ease decision making burden (Table 1, Supplementary File 1).
As a first step towards reproducible RoB evaluations, we sought to automate completion of each seroprevalence JBI checklist item. We reviewed the seroprevalence JBI checklist to identify items that could be automatically completed based on key information that should be extracted as part of any critical appraisal or systematic review of seroprevalence studies.[14]

Part 2: Developing and automating decision rules for risk of bias assessment

We developed decision rules that could be automatically applied to the ratings from the seroprevalence JBI checklist to generate an overall RoB assessment (low, moderate, or high) for each study (Figure 1). The use of a decision rule-based algorithm was selected as it allows for all checklist items to be considered together in evaluating overall study RoB. The decision rules were developed based on published guidance on estimating disease prevalence, reports on the evaluation of prevalence studies[5,15], opinions of experts in evidence synthesis and infectious disease epidemiology, and the consensus of researchers at SeroTracker after evaluation of thousands of seroprevalence studies (Supplementary file 2).[2,3]

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Validation dataset

The SeroTracker-RoB approach was evaluated using data from SeroTracker's living systematic review database of SARS-CoV-2 seroprevalence studies.[10] The protocol for the review was registered and published (PROSPERO: CRD42020183634, version July 21, 2021).

The dataset used for validation included all seroprevalence studies included in the SeroTracker database with publication dates between January 1, 2020 to November 17, 2021. The 2,070 studies in this evaluation dataset included peer-reviewed literature, preprints, government and non-governmental organization reports, and media articles.

Manual RoB assessments and automated SeroTracker-RoB assessments were completed for each study. Manual RoB assessments were completed by two independent reviewers, each of whom examined data routinely extracted from each article to complete the seroprevalence JBI checklist (Supplementary file 1) and performed a manual RoB assessment based on the checklist items (Supplementary file 2). Disagreements were resolved by consensus. Automated RoB assessments were conducted using the SeroTracker-RoB approach as outlined above, involving application of the automated JBI coding logic and decision rules to the same routinely extracted data. The logic for the SeroTracker-RoB approach was implemented in the AirTable programmable spreadsheet.

Evaluating SeroTracker-RoB efficiency, coverage, and reliability

Efficiency

We assessed efficiency by comparing time to complete manual RoB and automated SeroTracker-RoB assessments for 25 seroprevalence articles that were randomly selected from the SeroTracker dataset. Five SeroTracker systematic reviewers were timed as they each completed assessments for five articles. The time required for the manual approach included assessment of each checklist item and evaluation of overall RoB. The time required for the SeroTracker-RoB automated approach included only the time taken to evaluate all checklist items that could not be automated, as the SeroTracker-RoB approach is otherwise instantaneous.

We compared the speed of manual vs. automated SeroTracker-RoB assessment for each article using a paired *t* test.

Coverage

To determine coverage, we calculated the proportion of studies in the dataset for which the SeroTracker-RoB decision rules yielded a RoB assessment.

Reliability

To evaluate reliability of the RoB assessments across consensus manual and automated SeroTracker-RoB approaches, we calculated the absolute agreement between ordinal overall RoB assessments using a two-way random-effects average-measures intraclass correlation (ICC).[16–18] We also evaluated reliability between the two independent reviewers conducting manual assessment by calculating absolute agreement using a two-way random-effects single-measures ICC. Analyses were conducted using STATA 14 (College Station, TX: StataCorp LP).

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

RESULTS

Automated JBI checklist

Four of the nine JBI checklist items (items 4, 6, 8, 9) were modified to add conditions that enabled them to be judged in a binary fashion. For example, Item 4 (were the study subjects and setting described in detail?) was marked as "Yes" if the average age and distribution of gender/sex was provided and "No" if neither age or gender/sex was provided, or only one of age and gender/sex was provided.

Item 8 (appropriate statistical analysis) was split in two parts (8A, 8B) to differentiate studies that adjusted prevalence estimates for test characteristics (sensitivity/specificity) but not population demographics and vice versa.

Item 1 (whether the sample frame was representative of the target population) and item 5 (whether the characteristics of the sample were representative of the target population, in both the main and sub-group analyses) are judgements that require study-, sample-, and target population-specific context. We did not automate these items, instead choosing to provide detailed guidance and illustrative examples to improve the reliability of manual assessment (Supplementary file 1). Additionally, a small proportion of studies used complex testing algorithms involving multiple assays with no clear overall sensitivity or specificity; for these studies the validity of methods to identify the condition (Item 6) was determined by manual review with clear guidance (Supplementary file 1).

Automated decision rules

The decision tree to classify overall study RoB is shown in Figure 1. The decision tree classified overall RoB as low, moderate, or high based on the categorical ratings for each seroprevalence JBI checklist item, considering each item in turn to arrive at an overall judgement (i.e., item 1, item 2, etc.). The tree included 61 decision rules, which could lead to low (n=6), moderate (n=29), or high (n=26) RoB (Figure 1).

The decision tree generally emphasized two factors: sampling biases and measurement biases. Sampling biases related to the extent to which the seroprevalence estimated in the study sample was representative of the seroprevalence in the target population (items 1, 2, 3, 5, 8B, 9). Measurement biases were those related to measurement error (items 6, 7, 8A). Item 4 was classified as a reporting item and judged to not impact the RoB in any decision rule.

These decision rules generally considered studies to be low RoB if they reported seroprevalence estimates that were likely representative of the target population with limited measurement biases; moderate RoB if they reported somewhat representative estimates or had limited measurement biases; or high RoB if they reported non-representative estimates and with considerable measurement biases.

More specifically, the decision rules considered studies to be low RoB if sampling biases were limited (i.e., appropriate sample frame, probability sampling, adequate sample size, and statistical adjustment for population characteristics) and measurement biases were also limited (i.e., adequate sensitivity/specificity and/or adjustment for test performance). These rules largely

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

considered studies at moderate RoB if the study had a few sampling biases (i.e., inappropriate sample frame but probability sampling, adequate sample size, and statistical adjustment for population characteristics) or limited measurement biases. In contrast, studies were considered high RoB if the study had sampling biases (i.e., non-probability sampling and inadequate sample size) and measurement bias was considerable (i.e., poor sensitivity/specificity and lack of adjustment for test performance).

Efficiency

Use of the SeroTracker-RoB tool resulted in significantly faster RoB assessments compared to the manual approach with a mean and standard deviation (SD) time of 0.80 (SD: 0.53) versus 2.93 (SD: 1.08) minutes per article, respectively ($p<0.001$). The automated approach took less time than the manual approach for all 25 articles evaluated.

Coverage

The SeroTracker-RoB approach yielded a RoB assessment for 100% (n = 2,070) of the SARS-CoV-2 seroprevalence studies in the SeroTracker database.

Reliability

ICC for the reliability of the RoB assessment was 0.77 (95% confidence interval [CI] 0.74-0.80) between the SeroTracker-RoB approach and consensus manual review. For comparison, the ICC between the two manual independent reviewers was 0.74 (95% CI 0.71-0.76).

The SeroTracker-RoB tool

A Microsoft Excel tool that implements this automated approach to prevalence study RoB assessment, including automated JBI checklist and decision rules, can be found in Supplementary file 3. The decision tree logic was implemented in Excel using Visual Basic for Applications.

The tool includes four sheets: (1) a legend sheet to describe the tool and orient the user; (2) a data extraction sheet with 25 core data fields for reviews of prevalence studies, including the two seroprevalence JBI checklist items requiring manual assessment; seven user-defined thresholds for study parameters (i.e., minimum sample size, test validity, response rate); and the resulting seroprevalence JBI checklist and RoB rating; (3) a data dictionary sheet with descriptions of the data extraction fields, instructions for extraction, and the criteria used to automate each JBI checklist item; and (4) a data validation sheet where options for dropdown menus can be edited by the user. Using built in formulas and these extracted fields, the tool completes the checklist and produces an overall RoB rating. Users who wish to manually complete the JBI checklist can enter their ratings into the nine seroprevalence JBI item variables, and still take advantage of the SeroTracker-RoB decision rules for overall RoB assessment.

DISCUSSION

We developed the SeroTracker-RoB approach to RoB assessment for seroprevalence studies, which includes an automated variant of the Joanna Briggs Institute Critical Appraisal Checklist for prevalence studies and decision rules to determine overall RoB from this checklist. When

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

validating this approach against the SeroTracker database of over 2,000 seroprevalence studies, it was over five times faster than manual review, had perfect coverage, and had good agreement with manual review from two independent reviewers. This approach has clear potential to enable rapid, robust, and reproducible RoB assessment in systematic reviews of prevalence studies.

There is no consensus on the most valid approach to automate RoB assessments.[5] Some studies have used summary scores, adding the number of criteria met for a given critical appraisal checklist and creating a threshold on that score to assess RoB.[9–11] However, this approach weights each checklist item equally, which does not reflect the different implications of critical appraisal concepts for RoB. For example, in the context of the JBI checklist, marking six out of nine JBI items "Yes" could be achieved by 84 different combinations of item responses, each of which have different implications for RoB.[19–21]

Weighted averages better reflect the relevance of each item, overcoming some limitations of simple summary scores. However, checklist items cannot always be considered independently in assessing RoB, thereby introducing complexity in the derivation of a weighted score. For example, it is more important to correct for antibody test performance (item 8B = yes) when assay sensitivity and specificity are low (item 6 = no).

Considering combinations of items together may enable better automated RoB assessment. Some studies have trained deep learning algorithms to identify relevant text in publications of randomized controlled trials and predict RoB assessments, with reasonable accuracy compared to human reviewers.[19,20] However, these algorithms have largely been trained on small datasets and have limited interpretability and transparency due to the "black-box" decision making of the models.

A decision rule approach, on the other hand, provides a transparent and interpretable model to automate RoB assessment.[21] The tree structure captures interactions between checklist features, is easy to implement, and enables clear visualization of which combinations of features are most important for RoB. The decision tree in the SeroTracker-RoB approach reveals two key axes of bias in seroprevalence studies: sampling biases, related to the correspondence between the sample frame and target population, sampling representativeness, response rate, and population weighting; and measurement biases, related to antibody test performance and correction for that performance. Clusters of rules in this decision tree also make clear how features in each axis interact to determine study RoB.

RoB assessment is a time-consuming component of conducting systematic reviews.[22,23] As such, improving efficiency of the assessment process may help to reduce the burden of evidence synthesis. This is particularly important for living reviews and reviews during health emergencies, which may need to process a high volume of information quickly to inform public health decision making.

The SeroTracker project exemplifies how time-savings can be of value. There is an average of 35 new studies added each week to the SeroTracker living systematic review project.[2] The SeroTracker-RoB approach described in this manuscript is over five times faster than traditional manual review for assessing seroprevalence study RoB, which takes approximately three

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

minutes. Thus, using the SeroTracker-RoB approach would save nearly three hours of reviewer time each week compared to the manual approach.

In traditional manual RoB assessment, there is imperfect agreement between even trained human reviewers.[6,7,22,23] In our study, inter-rater reliability for RoB between two independent reviewers was moderate (ICC 0.74). As such, manual review cannot be considered a perfect standard, and the 0.77 ICC between the SeroTracker-RoB approach and manual review in part reflects the heterogeneity and inconsistency of manual assessment. This highlights the benefit of an automated approach that yields reproducible assessments.

There have been many systematic reviews of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) seroprevalence studies to date.[3,4,9–12,24] Several of these reviews use the JBI checklist for critical appraisal, but they implement the checklist and judge overall RoB in different ways.[3,4,9–12] For this reason, we developed the SeroTracker-RoB Excel Tool (Supplementary file 3), which embeds the seroprevalence JBI coding logic and decision rule in a user-friendly data extraction sheet. This tool may be valuable to other investigators seeking to conduct rapid and reliable RoB assessments for seroprevalence studies — an important endeavor, given the ongoing value of SARS-CoV-2 seroprevalence studies for surveillance as the virus becomes endemic[25], and given the many (sero)prevalence studies conducted for other diseases and conditions.

Users of this tool may need to adapt some of the seroprevalence JBI items to meet their unique needs. We judged valid methods of detecting SARS-CoV-2 antibodies using 90% sensitivity and 97% specificity thresholds established by the World Health Organization[26], but this criterion will vary for other conditions. To accommodate these needs, the SeroTracker-RoB Excel tool includes embedded user-set thresholds that can easily be altered as required. Furthermore, users can also manually enter seroprevalence JBI checklist results if they wish to utilize only the decision rule, as opposed to conducting data extraction to first automate completion of the seroprevalence JBI checklist items.

This study had several strengths. Firstly, to our knowledge, this is the first validation of decision rules for RoB in seroprevalence studies, with the validation dataset including thousands of heterogeneous SARS-CoV-2 seroprevalence studies across study designs, regions, and target populations. Second, the use of a decision rule approach enables transparent RoB assessment, while also ensuring reproducibility and speed. Finally, the decision rules have a robust foundation in the validated and widely used JBI checklist, published guidance on estimating disease prevalence, opinions of methodological experts, and the experiences of researchers that have conducted thousands of RoB assessments for seroprevalence studies.

This study had several limitations. First, the SeroTracker-RoB approach still requires reviewers to make expert judgments on two JBI items, and, in rare circumstances, a third judgment for studies using complex antibody testing algorithms without a calculable sensitivity and specificity. However, the JBI and the SeroTracker-RoB tool provide clear guidance and examples for making these assessments. Second, the decision rule was derived in part using expert judgment. However, the transparency of this algorithm allows for scientific debate and further refinement of the decision rules, if needed. Thirdly, the generalizability of the

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SeroTracker-RoB decision rules is unclear. Although the validation database used was large and robust, it focused exclusively on SARS-CoV-2 seroprevalence. Applying this approach to (sero)prevalence studies for other pathogens or conditions may require additional validation or adaptation. Furthermore, development of decision rules using established critical appraisal checklists as a foundation for automated RoB assessment of other study designs should be considered.

Conclusions

We developed and validated the SeroTracker-RoB approach, which enables rapid, transparent, and reproducible risk of bias assessment for seroprevalence studies. This approach largely automates the established Joanna Briggs Institute Critical Appraisal Checklist for Prevalence studies, and adds interpretable decision rules to assess overall risk of bias. Moreover, the SeroTracker-RoB tool embeds this approach in a simple data extraction sheet for use by other researchers conducting evidence synthesis of prevalence studies during outbreaks and for endemic infectious disease.

Funding statement

SeroTracker receives funding for SARS-CoV-2 seroprevalence study evidence synthesis from the Public Health Agency of Canada through Canada's COVID-19 Immunity Task Force, the World Health Organization Health Emergencies Programme, the Robert Koch Institute, and the Canadian Medical Association Joule Innovation Fund. No funding source had any role in the design of this study, its execution, analyses, interpretation of the data, or decision to submit results. This manuscript does not necessarily reflect the views of the World Health Organization or any other funder.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

1. Cheng MP, Yansouni CP, Basta NE, et al. Serodiagnostics for Severe Acute Respiratory Syndrome-Related Coronavirus 2 : A Narrative Review. *Ann Intern Med*. 2020;173(6):450-460. doi:10.7326/M20-2854
2. Arora RK, Joseph A, Van Wyk J, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. *Lancet Infect Dis*. 2021;21(4):e75-e76. doi:10.1016/S1473-3099(20)30631-9
3. Bobrovitz N, Arora RK, Cao C, et al. Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. *PLOS ONE*. 2021;16(6):e0252617. doi:10.1371/journal.pone.0252617
4. Bergeri I, Whelan M, Ware H, et al. Global epidemiology of SARS-CoV-2 infection: a systematic review and meta-analysis of standardized population-based seroprevalence studies, Jan 2020-Oct 2021. *medRxiv*. Published online January 1, 2021:2021.12.14.21267791. doi:10.1101/2021.12.14.21267791
5. Borges Migliavaca C, Stein C, Colpani V, et al. How are systematic reviews of prevalence conducted? A methodological study. *BMC Med Res Methodol*. 2020;20(1):96. doi:10.1186/s12874-020-00975-3
6. Hartling L, Hamm MP, Milne A, et al. Testing the Risk of Bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol*. 2013;66(9):973-981. doi:10.1016/j.jclinepi.2012.07.005
7. Hartling L, Milne A, Hamm MP, et al. Testing the Newcastle Ottawa Scale showed low reliability between individual reviewers. *J Clin Epidemiol*. 2013;66(9):982-993. doi:10.1016/j.jclinepi.2013.03.003
8. Hoy D, Brooks P, Woolf A, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. *J Clin Epidemiol*. 2012;65(9):934-939. doi:10.1016/j.jclinepi.2011.11.014
9. Chisale MRO, Ramazanu S, Mwale SE, et al. Seroprevalence of anti-SARS-CoV-2 antibodies in Africa: A systematic review and meta-analysis. *Rev Med Virol*. Published online July 6, 2021:e2271-e2271. doi:10.1002/rmv.2271
10. Galanis P, Vraka I, Fragkou D, Bilali A, Kaitelidou D. Seroprevalence of SARS-CoV-2 antibodies and associated factors in healthcare workers: a systematic review and meta-analysis. *J Hosp Infect*. 2021;108:120-134. doi:10.1016/j.jhin.2020.11.008
11. Rostami A, Sepidarkish M, Leeflang MMG, et al. SARS-CoV-2 seroprevalence worldwide: a systematic review and meta-analysis. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis*. 2021;27(3):331-340. doi:10.1016/j.cmi.2020.10.020
12. Lewis H, Ware H, Whelan M, et al. SARS-CoV-2 infection in Africa: A systematic review and meta-analysis of standardised seroprevalence studies, from January 2020 to December 2021. *medRxiv*. Published online 2022.
13. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *JBI Evid Implement*. 2015;13(3). https://journals.lww.com/ijebh/Fulltext/2015/09000/Methodological_guidance_for_systematic_reviews_of.6.aspx
14. World Health Organization Seroepidemiology Technical Working Group. ROSES-S: Statement from the World Health Organization on the reporting of seroepidemiologic studies for SARS-CoV-2. *Influenza Other Respir Viruses*. 2021;15(5):561-568.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

doi:10.1111/irv.12870

15. Migliavaca CB, Stein C, Colpani V, Munn Z, Falavigna M. Quality assessment of prevalence studies: a systematic review. *J Clin Epidemiol*. 2020;127:59-68. doi:10.1016/j.jclinepi.2020.06.039
16. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
17. McGraw K, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46. doi:https://doi.org/10.1037/1082-989X.1.1.30
18. Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 86(2):420-428. doi:https://doi.org/10.1037/0033-2909.86.2.420
19. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol*. 2007;36(3):666-676. doi:10.1093/ije/dym018
20. Jüni P, Witschi A, Bloch R, Egger M. The Hazards of Scoring the Quality of Clinical Trials for Meta-analysis. *JAMA*. 1999;282(11):1054-1060. doi:10.1001/jama.282.11.1054
21. Greenland S, O'rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics*. 2001;2(4):463-471. doi:10.1093/biostatistics/2.4.463
22. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ*. 2009;339:b4012. doi:10.1136/bmj.b4012
23. Hartling L, Bond K, Vandermeer B, Seida J, Dryden DM, Rowe BH. Applying the Risk of Bias Tool in a Systematic Review of Combination Long-Acting Beta-Agonists and Inhaled Corticosteroids for Persistent Asthma. *PLOS ONE*. 2011;6(2):e17242. doi:10.1371/journal.pone.0017242
24. Chen X, Chen Z, Azman AS, et al. Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *Lancet Glob Health*. 2021;9(5):e598-e609. doi:10.1016/S2214-109X(21)00026-7
25. European Centre for Disease Prevention and Control. *COVID-19 Surveillance Guidance*.; 2021. https://www.ecdc.europa.eu/en/publications-data/covid-19-surveillance-guidance
26. World Health Organization. *WHO COVID-19: Case Definitions*.; 2020. file:///Users/niklasbobrovitz/Downloads/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2-eng.pdf

Table 1. Modified Joanna Briggs Institute Checklist for Prevalence Studies

Joanna Briggs Institute checklist items[a]	**SeroTracker-RoB standardized guidance for checklist items [response options]**	**Notes**
Item 1: Was the sample frame appropriate to address the target population?	Sample frame described and it approximated the target population [yes / no]	Requires subjective evaluation. See Supplementary file 1 of JBI Guidance for full details
Item 2: Were study participants recruited in an appropriate way?	Probability sampling (simple or stratified random) OR entire sample (e.g., an entire town) [yes / no / unclear]	
Item 3: Was the sample size adequate?	≥599 OR sample size calculation was provided AND the required sample for 80% power was below the threshold (n<599) OR precision was ≤1.25% [yes / no / unclear]	See Supplementary file 1 for sample size calculation.
Item 4: Were the study subjects and setting described in detail?	Average age AND proportion of gender/sex provided [yes / no]	
Item 5: Was data analysis conducted with sufficient coverage of the identified sample?	The demographic characteristics (gender/sex, age, and ethnicity) of the sample are at least somewhat representative of the population in both the main and sub-group analyses [yes / no / unclear]	This item requires subjective review of demographic breakdown in the study, coverage within subgroup estimates, and review of author comments on representativeness of the sample
Item 6: Were valid methods used for the identification of the condition?	The serological test used did meet the WHO Unity Study Criteria for serological tests: sensitivity minimum 90% AND specificity minimum 97%. The combined sensitivity of multiple testing algorithms must also meet this threshold, or the testing algorithm must be logical and robust for algorithms without a calculable combined sensitivity and specificity [yes / no]	Exceptions are made for algorithms using a commercial or in-house binding assay with confirmatory testing using virus neutralization assay, as they constitute the gold standard in serological evaluation.[20]
Item 7: Was the condition measured in a standard, reliable way for all participants?	The same serology test was used for all participants [yes / no / unclear]	
Item 8A: Provides statistical adjustment for test characteristics and provides the information necessary to determine the numerator, denominator, prevalence estimate, and confidence interval Item 8B Provides statistical adjustment for population characteristics or the sample is somewhat representative of the population [probability sampling] and provides the information necessary to determine the numerator, denominator, prevalence estimate, and confidence interval	Conducted test adjustment AND reported relevant information OR Item 6 was “Yes” [yes / no / unclear] Conducted population adjustment AND reported relevant information OR Item 1, 2, and 3 were all “Yes” [yes / no / unclear]	

Item 9: Was the response rate adequate, and if not, was the low response rate managed appropriately?	Response rate > 60% or the demographics of the sample were a reasonable match to those of the target population [yes / no / unclear]	
Overall risk of bias	**Low:** The estimates are very likely correct for the target population. To obtain a low risk of bias classification, all criteria must be met or departures from the criteria must be minimal and unlikely to impact on the validity and reliability of the prevalence estimate. These include sample sizes that are just below the threshold when all other criteria are met, reporting only some of characteristics of the sample, test characteristics below the threshold but corrections for the test performance, and response rates that are just below the threshold in the context of probability based sampling of an appropriate sampling frame with population weighted seroprevalence estimates. **Moderate:** The estimates are likely correct for the target population. To obtain a moderate risk of bias classification, most criteria must be met and departures from the criteria are likely to have only a small impact on the validity and reliability of the prevalence estimates. **High:** The estimates are not likely correct for the target population. To obtain a high risk of bias, many criteria must not be met or departures from criteria are likely to have a major impact on the validity and reliability of the prevalence estimates. **Missing:** There was insufficient information to assess the risk of bias.	
[a]Item 1-9 are a modified version of the Joanna Briggs Institute Critical Appraisal Checklist for prevalence studies		

medRxiv preprint doi: https://doi.org/10.1101/2021.11.17.21266471; this version posted March 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

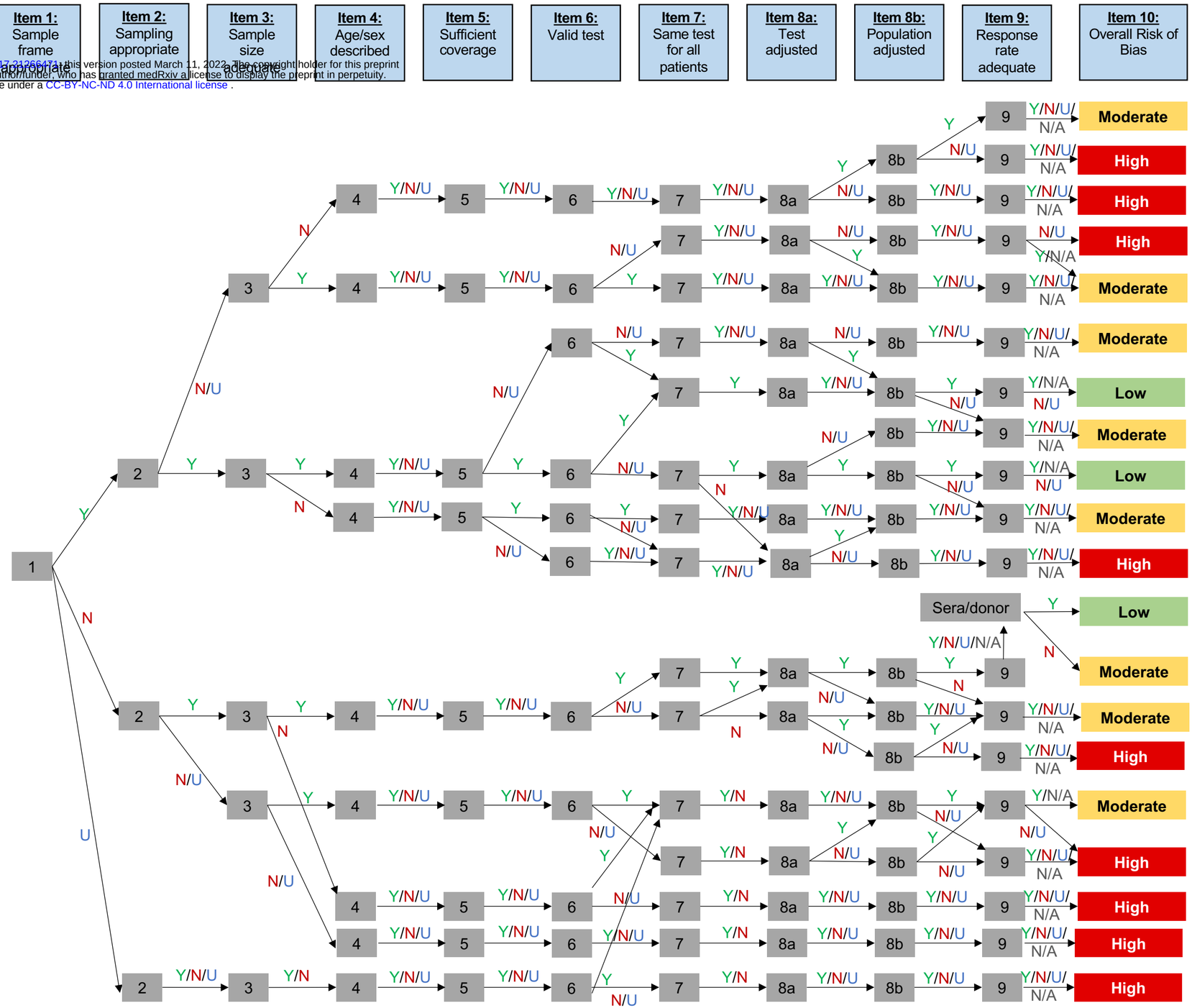

Figure 1. Decision rules to determine risk of bias for seroprevalence studies

Y=yes, N=no, U=unclear, N/A=not applicable