

Trends in quality assessment of observational studies and possibilities of a quantitatively graded quality assessment tool: a scoping review

Authors: Md Sohail Aman¹, Dr. Abhra Roy Choudhury¹, Dr. Gulchehak Kaur¹, Kopal Dixit¹, Deepti Rai¹

¹PharmaQuant Insights Pvt. Ltd.



INTRODUCTION

- Comparative observational studies, such as case-control and cohort studies, are important for exploring associations between exposures and outcomes, but they are inherently prone to biases.¹
- Key biases include selection bias from non-random participant selection, confounding bias from external factors, bias from data collection discrepancies, observer bias from researcher expectations, and attrition bias due to loss to follow-up.¹
- Quality appraisal tools with structured frameworks are crucial for evaluating the rigor and reliability of observational studies to assess study design, methodology, and reporting.
- Despite their importance, current tools often fail to fully address and estimate the inherent biases that may compromise study validity, as they may not adequately estimate the nuanced impacts of selection or confounding biases on findings and as the tools depend on reviewers' perspective mostly.
- Considering the increasing recognition of Real-World Data (RWD) from authorities like the Institute for Clinical and Economic Review (ICER), the U.S. Food and Drug Administration (FDA), and the National Institute for Health and Care Excellence (NICE), it is crucial to ensure that observational studies are of high quality. Furthermore, the tools used for quality assessment must effectively evaluate this quality to achieve the best outcomes.²⁻⁴

OBJECTIVES

- To identify and evaluate the quality assessment tools used in systematic reviews (2019-2024) of comparative observational studies, with a focus on assessing their usability and effectiveness.
- To evaluate possible domains of biases for graded weight distribution, from the identified tools

METHODS

- A scoping review was conducted using the MEDLINE and EMBASE databases to identify systematic literature reviews (SLRs) published between 2019 and 2024 (until May 25, 2024). The SLRs focusing on case-control and cohort studies assessing their quality using quality assessment tools were included. The identified tools were further analyzed to understand their domains, methods, strengths and weaknesses (Table 2).
- Five reviewers analyzed the identified tools. The inclusion criteria is specified in Table 1.

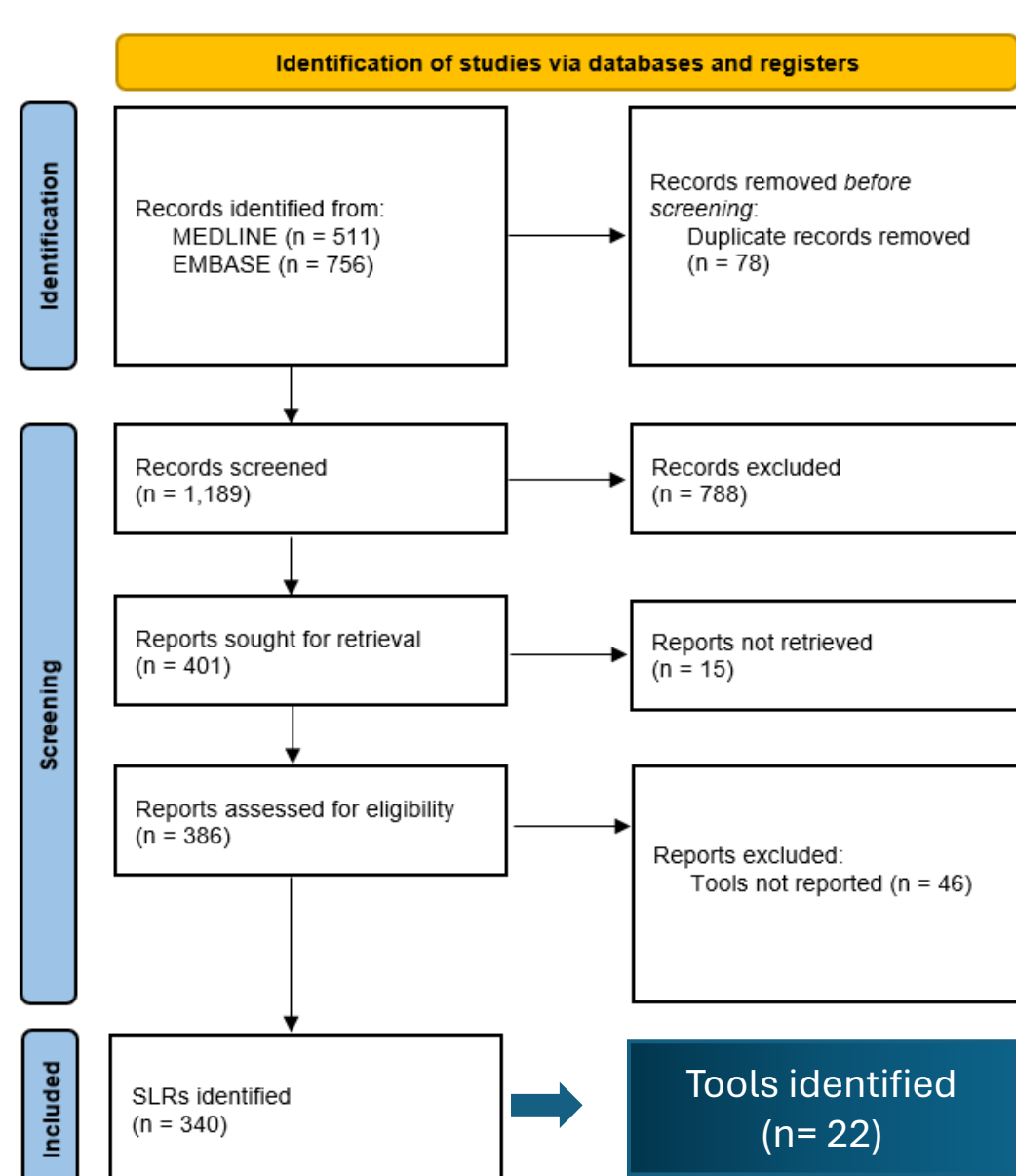
Table 1: Inclusion criteria

Eligibility criteria for inclusion
SLRs including case-control and cohort studies using QA tools
Published between 2019-2024
Published in English language

RESULTS

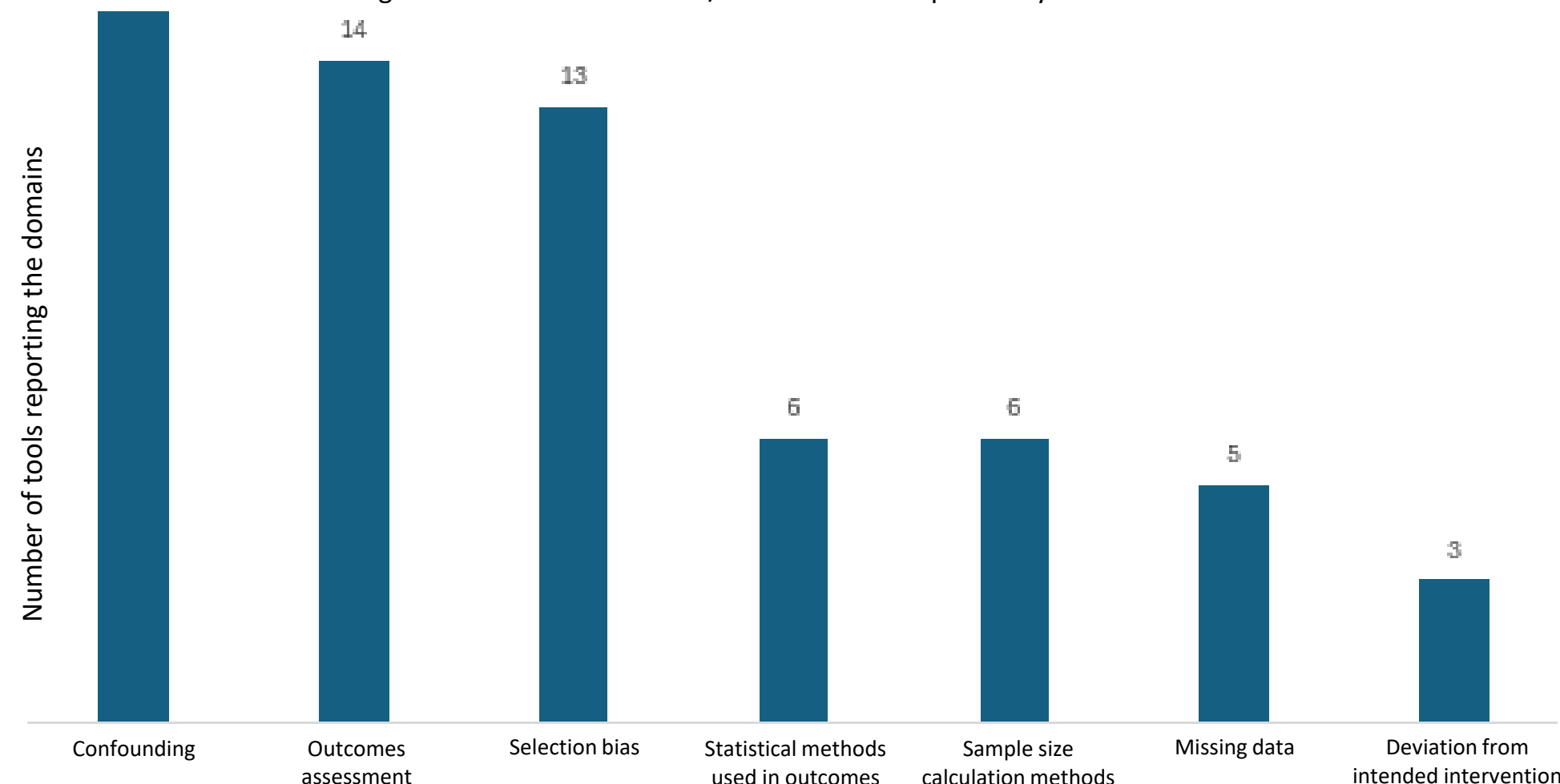
- Among 1,189 citations screened at the title- abstract stage, 386 were screened at the full-text stage (Figure 1).
- A total of 340 studies were identified during the screening process conducted by a single reviewer.
- Through this review we identified 22 different tools for quality assessment of cohort and case-control studies.

Figure 1: PRISMA 2020 flow diagram of included studies⁵



- The studies employed 22 different tools for quality assessment of cohort and case-control studies. The National Institute of Health (NIH) Quality Assessment Tool was utilized most often (41%), followed by the Joanna Briggs Institute (JBI) checklist (18%), the Newcastle-Ottawa scale (13%), Critical appraisal skill programme (CASP) (6%), and The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist (5%).⁵⁻¹⁰ The identified tools were either completely qualitative in nature or had an item-based equal weighted scoring. None of the tools were graded based on the different weight of items.
- The prominent domains or sources of biases were investigated and are represented in Figure 2. Confounding, outcome assessment and selection biases were explored the most. However, as per our analysis all the domains were assessed on equal weightage or solely on reviewers' perspective.

Figure 2: Prominent domains/sources of bias reported by identified tools

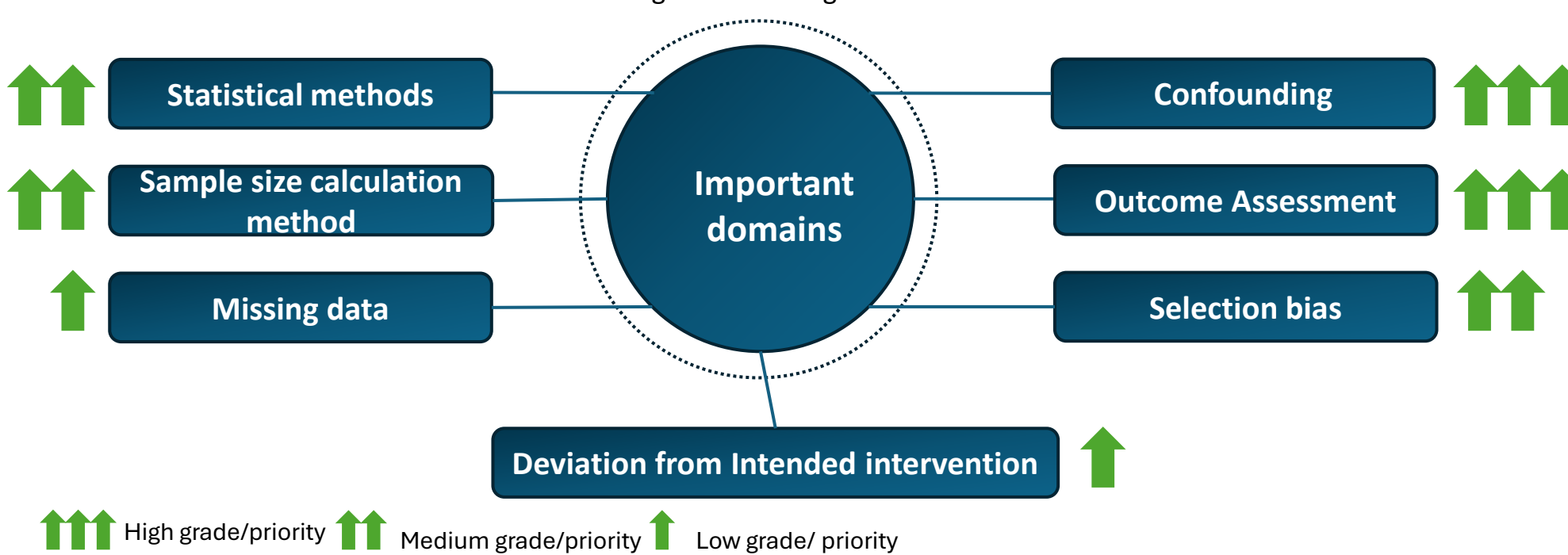


Evaluation and proposed grading of key domains (Table 2 and Figure 3)

Table 2: Domains and recommendations

Bias	Reason for considering this domain	Our evaluation and grading
Confounding	1. Observational studies have a high tendency of confounding bias as measurements to avoid confounding are often not taken or not feasible. This significantly distort the ability to find an association between an exposure or outcome. 2. Since a confounding bias affects exposure and outcome independently, it can also over- or underestimate the relationship between them which can result in inaccurate causal inferences.	Our research and opinion suggest high importance to confounding bias in observational studies and recommend quantitative scoring with high grade.
Outcome assessment	1. Observational studies often fail to produce consistent results and lead to wrong classification of outcomes. 2. If an appropriate outcome assessment method is not employed or there is an inconsistency in the assessment of the cases and controls, the association might again be exaggerated or underplayed.	Our research and opinion suggest high importance to bias resulting from outcome assessment in observational studies and recommend quantitative scoring with high grade.
Selection bias	1. The inherent nature of observational studies will always lead to selection bias. This bias in case control and cohort studies can lead to non-representativeness or unequal representativeness of the subjects. 2. It can also lead to a prevalence-incidence bias, for example, selecting cases of only surviving members of cardiovascular diseases and not considering the severe cases who have died, might result in underestimation of the association. 3. However, avoiding selection bias to a high extent may not be possible in observational studies. Hence, penalizing observational studies with low scores for selection bias may underestimate the true quality of the study	Our research and opinion suggest moderate importance to selection bias in observational studies and recommend quantitative scoring with moderate grade.
Statistical methods and sample size calculation	Two other important biases, that are, i) inadequate or inaccurate statistical methods in the outcome analysis and ii) the bias related to wrongful adjustment of sample size to account for attrition can hamper the reliability and interpretability of the results as the study might not be powered enough to detect the true association or the methods of outcome analysis were not statistically appropriate.	Our research and opinion suggest moderate importance to statistical methods applied in observational studies and recommend quantitative scoring with moderate grades.

Figure 3: Grading of domains



Sr.No.	Tool name	Type of tool	Domains or topics addressed	Strength	Limitation
1	Risk of Bias Tool for Non-Randomized Studies of Interventions (ROBINS I) ¹¹	Quality assessment tool	Confounding, Selection, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, Overall	1. Domains or biases covered are similar to Cochrane Risk of Bias (RoB) 2. Addresses maximum types of bias in non-randomized studies	1. Selection bias could be overestimated which will in general mark studies as low quality 2. No quantitative measure for RoB 3. Subjectivity of reviewer will persist
2	Risk of Bias Tool for Non-Randomized Studies - Environmental Exposures (ROBINS-E) ¹²	Quality assessment tool	Confounding, selection bias, measurement bias, reporting bias, missing data, deviations from intended exposures, bias in selection of reported results	1. Tailored to environmental exposure studies; detailed and comprehensive	1. Consistency of effect of exposure is over simplified and cannot be answered in yes/no 2. No quantitative measure for RoB Dependent on researcher's subjectivity
3	National Institutes of Health (NIH) ⁶	Quality assessment tool	Reporting, selection, confounding, validity, data completeness, sample size adequacy	1. Sample size justification, power description, or variance and effect estimates are investigated 2. It emphasizes key elements like exposures and outcomes Proper emphasis on confounding	1. Although domains addressed; Not as thorough in domain-by-domain bias assessment 2. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
4	Joanna Briggs Institute (JBI) ⁷	Quality assessment tool	Study design, selection bias, exposure/outcome measurement, confounding, follow-up	1. Focuses on internal validity and reliability 2. Emphasizes on key components like exposures and outcomes for cohort studies and addresses the representation of cases and controls for case control studies 3. Proper emphasis on confounding and follow-up	1. There is no quantitative quality score which makes it difficult to compare studies 2. Focuses on the methodological aspect of reporting Very subjective in nature 3. General questions may miss specific biases
5	Newcastle-Ottawa Scale (NOS) ⁸	Quality assessment tool	Selection, comparability, confounding, and outcome/exposure measurement, reporting bias for cohort and case-control studies	1. Allows structured quality assessment with a scoring system, applicable to cohort and case-control studies 2. Emphasis on selection bias and confounding 3. Blinding of Outcome assessors addressed	1. Some subjective scoring 2. Does not consider for generalizability for external validation 3. Addresses confounding, however, it does not provide detailed guidance on how to assess confounding 4. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
6	Critical Appraisal Skills Programme (CASP) ⁹	Quality assessment tool	Potential bias, confounding, exposure and outcome assessment, data completeness, outcome measurement, external validity	1. For cohort studies, the checklist asks about the recruitment of participants, the accuracy of outcome measurements, and the length and adequacy of follow-up. 2. For case-control studies, it focuses on aspects like case definition, control selection, and the comparability of cases and controls 3. Encourage reviewers to think about the applicability of the study findings to real-world settings. Questions like "Can the results be applied to the local population?" help to bridge the gap between study results and practical implementation	1. There is no quantitative quality score which makes it difficult to compare studies 2. Focuses on the methodological aspect of reporting 3. Very subjective in nature as questions are dependent on "adequate" or "same population" without defining them
7	Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) ¹⁰	Reporting guideline	Study design, sample size, bias, confounders, data sources, and statistical methods.	1. Encourages standardized reporting 2. Detailed for observational studies 3. Emphasizes on key components like participation criteria 4. Emphasis on confounding	1. Primarily focused on reporting rather than assessing quality or bias in studies. 2. No quantitative quality score which makes it difficult to compare studies or conduct any analysis 3. Suggests about potential bias addressed but does not define the biases specific to study design
8	Downs and Black ¹³	Quality assessment tool	Reporting, external validity, bias, confounding, statistical power	1. Applicable to a wide range of study designs, including observational and randomized studies. 2. Blinding in observational studies may be explored 3. Outcome manipulation (data dredging) addressed 4. Sample size calculation assessed	1. Limited adaptability for non-healthcare studies; 2. Scoring can be subjective: question are answered with Yes/No only ; Graded equally with Yes 1, No/unclear 0 3. Bias arising from outcome completely missed 4. Bias arising from selection not adequately addressed 5. Bias arising from confounding not addressed adequately 6. Domain 1 is reporting and will not truly capture the quality
9	Modified Downs and Black ¹⁴	Quality assessment tool	Reporting, external validity, bias, confounding, statistical power	1. Applicable to a wide range of study designs, including observational and randomized studies. 2. Blinding in observational studies may be explored 3. Outcome manipulation (data dredging) addressed 4. Sample size calculation assessed	1. Limited adaptability for non-healthcare studies; 2. Scoring can be subjective: question are answered with Yes/No only ; Graded equally with Yes 1, No/unclear 0 3. Bias arising from outcome completely missed 4. Bias arising from selection not adequately addressed 5. Bias arising from confounding not addressed adequately 6. Domain 1 is reporting and will not truly capture the quality 7. Exactly same with Downs and Black
10	Prediction Model Risk of Bias Assessment Tool (PROBAST) ¹⁵	Quality assessment tool	Participant selection, predictor and outcome measurement, missing data, statistical analyses	1. Comparatively Lower subjectivity and variability in assessments, ensuring that reviewers systematically consider important aspects of study design, data handling, and statistical analysis 2. Assesses the applicability of the study's findings to the target population or clinical setting. This dual focus ensures that reviewers consider not only the methodological rigor but also the real-world relevance of the prediction models	1. No quantitative quality score which makes it difficult to compare studies 2. Focuses on internal validity and applicability rather than the quality of reporting 3. Highly specific to prediction models, limited broader applicability less suitable to investigate causal relationships between exposures and outcomes
11	Centre for Evidence-Based Medicine (CEBM) ¹⁶	Quality assessment tool	Study design, selection bias, confounding, measurement of exposure/outcome, statistical analysis	1. High emphasis on the internal validity 2. Address the applicability of the study's findings to a local context or specific patient population	1. Limited scope for specific biases 2. Less detailed for observational studies 3. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
12	Risk of Bias Assessment Tool for Non-Randomized Studies (RoBANS) ¹⁷	Quality assessment tool	Selection bias, confounding, measurement of exposure/outcome, outcome assessment, reporting bias for cohort and case-control studies	1. Specifically designed for non-randomized studies 2. Clear structure for risk of bias assessment, 3. Emphasis on selection bias and confounding 4. Blinding of Outcome assessors addressed	1. Does not consider for generalizability for external validation 2. Addresses confounding, however, it does not provide detailed guidance on how to assess confounding 3. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
13	Modified Newcastle-Ottawa Scale (NOS) ¹⁸	Quality assessment tool	Selection, comparability, confounding, and outcome/exposure measurement, reporting bias for cohort and case-control studies	1. Allows structured quality assessment with a scoring system, applicable to cohort and case-control studies 2. Emphasis on selection bias and confounding 3. Blinding of Outcome assessors addressed	1. Some subjective scoring 2. Does not consider for generalizability for external validation 3. Addresses confounding, however, it does not provide detailed guidance on how to assess confounding 4. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
14	Agency for Healthcare Research and Quality (AHRQ) ¹⁹	Quality assessment tool	Sampling, Outcome assessment, Confounding, missing data,	1. Domains are present for confounding, outcome assessment and missing data	1. One domain for reporting; which may not capture true quality 2. The questions are very much descriptive in nature and may not address the bias aptly 3. Does not consider for generalizability for external validation 4. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
15	Liverpool Quality Assessment tool (LQAT) ¹⁰	Quality assessment tool	Selection procedures, analysis/confounding, baseline assessment, outcome assessment, impact of findings	1. Combines checklist with a rating scale; 2. Good for both qualitative and quantitative studies 3. Considers impact of findings	1. Some items may be too general and bias from specific study designs may not get explored 2. Subjective interpretation of bias presence 3. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
16	McMaster University Occupational Therapy Evidence-Based Practice Research Group quality assessment tool ²¹	Quality assessment tool	Study purpose, literature review, study design, sample size calculation, outcomes measurement, intervention detail, Outcomes analysis methods, conclusion and clinical implications	1. Practical for public health and clinical research 2. Strong focus on observational study design Justification based questionnaire	1. Specific to occupational therapy interventions 2. Limited in addressing reporting bias 3. Does not address factors important for comparative observational studies 4. Subjective interpretation of bias presence 5. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
17	Quality Assessment of Diagnostic Accuracy Studies (version 2) [QUADAS-2] ²²	Quality assessment tool	Patient selection, index test, reference standard, flow and timing bias	1. Tailored for diagnostic accuracy studies 2. Selection bias broadly addressed 3. Follow up differences were investigated	1. This tool will not cover all biases relevant to non-diagnostic observational studies 2. Does not focus on confounding 3. Subjective interpretation of bias presence 4. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
18	Effective Public Health Practice Project Quality Assessment Tool (EPHPP QA Tool) ²³	Quality assessment tool	Selection bias, study design, confounders, blinding, data collection methods, withdrawals and drop-outs, Intervention integrity, analyses	1. Suitable for various study designs including cohort and case-control; 2. Includes quality rating across domains 3. Deviations from intended intervention addressed	1. Subjective interpretation of bias presence 2. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
19	Quality In Prognostic Studies (QUIPS) ²⁴	Quality assessment Tool	Selection, attrition, prognostic factor measurement, outcome measurement, Adjustment for other prognostic factors, Statistical analysis and reporting	1. Strong focus on prognostic factor studies; 2. Addresses confounding comprehensively 3. Statistical analyses questions are better addressed	1. Limited to prognostic studies 2. Outcome assessment bias not addressed aptly 3. Subjective interpretation of bias presence 4. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
20	Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields (QualSyst Tool) ²⁵	Quality assessment tool	• Sampling method, • Outcome assessment, • Data analysis, • Withdrawals	--	1. Less detailed on reporting bias 2. Subjective interpretation of bias presence 3. No quantitative quality score which makes it difficult to compare studies or conduct any analysis
21	FLC platform (Osteba) ²⁶	Could not be assessed due to language other than English			
22	Scottish Intercollegiate Guidelines Network (SIGN) ⁷	Not eligible as this is an assessment tool for systematic reviews			

CONCLUSION

Two primary categories of quality assessment tools were identified: reporting checklists and quality appraisal instruments. A significant proportion of these tools rely heavily on the subjective perspectives of reviewers, with only a minority employing semi-quantitative measures that apply equal weighting to all criteria. Such subjective assessments may fail to accurately reflect the true quality of a study. Similarly, the semi-quantitative checklists face limitations due to their equal weighting of parameters, which may result in either an overestimation or underestimation of study quality. These factors result into inability to compare two or more studies on equal grounds. Observational studies inherently introduce selection bias, and thus, overly stringent evaluations of selection bias may not demonstrate the true quality. There is an urgent need of quantitative methods for quality assessment of observational studies with graded approach for domains. The heterogeneity of observational studies have greater probability of hindering consistent assessment, affecting the generalizability of findings across various settings. As we work towards developing more robust quantitative tools, it is vital to account for both the diversity of observational studies and the contextual factors involved, ensuring a more accurate and comprehensive assessment of study quality.

INSIGHTS

- Current tools or checklists are subjective and fail to assess bias truly
- Comparing two or more different studies on the basis of the risk of bias associated with them is not possible with current tools
- There is a need for developing quantitative tool to assess the risk of bias

RECOMMENDATION

We propose a graded approach to quality assessment that accounts for bias domains—specifically;

- High grade- confounding bias and outcome assessment bias
- Medium grade- selection bias, statistical methods in outcomes, and sample size calculation
- Low grade- missing data and deviation from intended intervention.

In this framework, weights would be adjusted according to the significance of each domain, enhancing the accuracy of quality evaluations.

LIMITATION

This study assessed quality appraisal tools used in systematic reviews published in 2019-2024 through a scoping review to understand the current trends. This study qualitatively reviewed and identified the gaps of subjective assessment. The proposed grading would need statistical weight calculation and a consensus among broader research community. Prior developing a novel quantitative checklist, a thorough systematic review may need to be conducted to get a comprehensive understanding of all possible tools and their methods.

REFERENCES

- Jager KL, Trippel G, Chesnaye NC, Dekker FW, Zoccali C, Stel VS. Where to look for the most frequent biases? Nephrology (Carlton). 2020 Jun;25(6):435-441. doi: 10.1111/nep.13706. Epub 2020 Mar 27. PMID: 32133725; PMCID: PMC7218122.
- The Institute for Clinical and Economic Review. Real World Evidence for Coverage Decisions: Opportunities and Challenges. A Report from the 2017 ICER Membership Policy Summit. 2018. <https://icer.org/wp-content/uploads/2020/11/ICER-Real-World-Evidence-White-Paper-Q32018.pdf>
- U.S Food & Drug Administration. Framework for FDA's Real-World Evidence Program. 2020. <https://www.fda.gov/media/120060/download>.
- NICE. NICE real-world evidence framework. Corporate document [ECDF]
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hrobjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Wachtel VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71. doi: 10.1136/bmj.n71. PMID: 33780297; PMCID: PMC805924
- National Heart Lab. Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies 2013 [Available from: <https://www.rhhi.nihr.gov.uk/health-topics/study-quality-assessment-tools>]
- Institute JB. Checklist for case-control studies 2020 [Available from: <https://jbi.global/critical-appraisal-tools>]
- Schie N-O. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2021.
- Programme CAS. CASP-checklist-cohort-study. 2024.
- STROBE. STROBE-checklist-v4. 2007.
- Seo HJ, Kim SY, Lee YL, Park JE. ROBANS 2: A Revised Risk of Bias Assessment Tool for Nonrandomized Studies of Interventions. Korean J Fam Med. 2023;44(5):249-60.
- Practice UoHSC/E. Modified Newcastle-Ottawa scale for appraisal of non-randomized controlled studies.
- Resource Dohah. Agency for healthcare resource and quality 2021 [Available from: <https://www.ahrq.gov/data/data-tools/index.html>]
- Bracken LE, Aunni AJ, Krishnan JJ, Peak M, Arnold J, Smyth RL, et al. Development of the Liverpool Quality Assessment tool (LQAT). PLoS One. 2017;12(1):e0169393.
- University M. McMaster University Occupational Therapy Evidence-Based Practice Research Group quality assessment tool. 1998.
- Whiting P, Pridge AW, Westwood M, Mallett S, Deeks JJ, Rattana JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2013;159(5):23-36.
- McMaster. Effective Public Health Practice Project: QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES
- Grooten WJA, Tsoi E, Ang BO, Boersma K, Sainsbury BM, Gendle B, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS-aspects of interrater agreement. Diagn Prog Res. 2019;3:5.
- Leanne M, Kinnel MS, Robert C, Lee, M.Sc. and Linda S. Cook, Ph.D. STANDARD QUALITY ASSESSMENT CRITERIA FOR Evaluating Primary Research Papers from a Variety of Fields. 2004.
- López de Argumedo M RE, Gutiérrez A, Bayón JC. Fichas de Lectura Crítica. 2017.
- SIGN. Scottish Intercollegiate Guidelines Network (SIGN). 2002.