Research principles

Transparency - Our research process is disclosed below. We disclose conflicts of interest, such as affiliate links, whenever present.

Curiosity - We acknowledge that science is constantly evolving and it is not always possible to reach a conclusion. We change our minds based on new evidence, even if it challenges our current beliefs or lifestyle.

Perspective - We consider each paper as one part of the puzzle, and interpret it in consideration with the current body of evidence. One new, exciting study does not instantly overturn the previous consensus.

Rigor - A deep exploration of the subject matter, through credible sources, is often necessary and encouraged. This goes beyond the other papers referenced in the study.

Nuance - The primary reviewer is encouraged to share any strengths or limitations of the study not covered by the scoring rubric.

Reading the study

We read the full study in either PDF form or printed out, with a pen or Apple pen for annotations. A paper may need to be read twice or more in order to be accurately scored.

Scoring

Each rubric is scored out of three or four points. They are added together for a total sum of points. The closer to 100%, the more reliable and relevant this study is. Rubrics that do not apply (such as the washout period) are not factored into the total score denominator. A study to which all rubrics apply would have a maximum score of 159/159.

Journal

Score	Description
Disqualified	Predatory Journal: Listed on Beall's list or similar databases, indicating questionable publishing practices and lack of scholarly scrutiny.
1	Non-Peer-Reviewed or Industry Bias: The Journal lacks peer review or represents strong industry/corporate interests, compromising objectivity.
2	Moderately Recognized: Peer-reviewed with some recognition in academia, moderate impact factor, and citation frequency.
3	Highly Respected: Well-known, respected journal with a high impact factor, rigorous peer review, and significant scholarly influence.
4	Premier Journal: Top-tier publication setting trends, with exceptional reputation, very high impact factor, and widely regarded seminal works.

Authors Preregistration

Score	Description
0	Not preregistered, high risk of data dredging.
1	Preregistration status unclear or in a non-reputable registry.
2	Preregistered in a less-known registry, some details missing.
3	Preregistered in a reputable registry, clear details provided, but unverified by a third party organization
4	Preregistered with details verified by third party in a top-tier registry.

Hypothesis

Score	Description
0	Unclear or non-testable hypothesis
1	Testable but lacks specificity or relevance
2	Clear, testable, and relevant but lacks depth
3	Detailed, clearly testable and thoroughly addresses the research problem
4	Exceptionally detailed, testable and insightful

Level of Evidence

Score	Description
0	In vitro studies or early-phase animal studies: Provide initial insights but are distant from human applicability.
1	Later-phase animal studies: Use advanced models closer to human biology but still lack full replication.
2	Observational studies in humans: Offer valuable insights but limited causal inference due to potential confounding.
3	Randomized Controlled Trials (RCTs): Well-designed trials in humans, minimizing bias and considered the gold standard for intervention efficacy determination.
4	Systematic reviews and meta-analyses of RCTs: the highest level of evidence providing the most reliable insights.

Study Design and Inclusion/Exclusion Criteria

Score	Description
0	Study design and inclusion/exclusion criteria are either not described or very poorly explained, resulting in significant ambiguity and questioning of the study's methodological rigor.
1	Study design or inclusion/exclusion criteria are vaguely described, lacking sufficient detail and precision, which may lead to misunderstandings or misinterpretations of the study's methodology.
2	Study design and inclusion/exclusion criteria are adequately described with reasonable clarity and detail, allowing for a basic understanding of the study's methodology.
3	Study design and inclusion/exclusion criteria are clearly and precisely described, with good detail that facilitates a thorough understanding and replication of the study's methodology.
4	Study design and inclusion/exclusion criteria are clearly and precisely described, with good detail that not only facilitates a thorough understanding and replication of the study's methodology but also addresses potential methodological challenges and justifies the chosen approaches, setting a gold standard for transparency and rigor.

Introduction

Score	Description
0	The introduction exhibits inflammatory language, biased or unsupported claims, and a lack of objectivity, potentially focusing excessively on supportive studies and cherry-picking.
1	There is a lack of clarity or comprehensiveness in identifying the research question, limited contextualization, and insufficient referencing to relevant literature, potentially leading to a misleading context.
2	While the research question is clear, there is biased referencing, a narrow scope, and inadequate consideration of conflicting evidence, limiting the comprehensiveness of the introduction.
3	The introduction presents a clear research question, provides a good summary of relevant research, but has minor issues with balance or scope, and slightly favors supporting studies.
4	The introduction is comprehensive, balanced, and well-referenced, clearly identifying the research question, offering a thorough summary of previous research, and fairly representing opposing viewpoints, without evidence of cherry-picking.

Blinding (If Applicable)

Score	Description
0	No blinding or poorly described method, leading to a high potential for bias. Lack of blinding procedures raises concerns about the reliability of study outcomes.
1	Single-blinded, but with the potential for bias due to unclear or insufficiently described blinding methods. Uncertainty regarding blinding procedures may undermine the credibility of study results.
2	Single-blinded, with adequately described blinding methods, although some measures to mitigate bias may be lacking. While efforts are made to blind participants or assessors, there remains a possibility of bias influencing study outcomes.
3	Double-blinded, with a well-described blinding method, despite challenges in blinding the intervention. Despite difficulties in blinding due to the nature of the intervention, comprehensive measures are taken to minimize bias.
4	Double or triple-blinded, with an excellent description of blinding procedures and minimal bias ensured through rigorous implementation. Despite potential challenges in blinding, meticulous efforts are made to maintain the integrity and validity of study findings.

Randomization (If Applicable)

Score	Description
0	No randomization or poorly described method, leading to potential selection bias.
1	Randomization mentioned but lacks detail or appropriateness, casting doubt on its effectiveness.
2	Basic randomization without advanced methods or statistical adjustment, posing some risk of bias.
3	Appropriate randomization with some advanced methods and adequate statistical adjustment.
4	Optimal randomization with advanced techniques, full blinding, and comprehensive statistical adjustment.

Power Analysis

Score	Description
0	Underpowered: Insufficient participants and poor study design, increasing the risk of false negatives (Type II errors).
1	Slightly underpowered: Marginally inadequate participants, with room for design enhancements. Findings may lack reliability due to an elevated risk of Type II errors.
2	Adequately powered: Sufficient participants for basic analysis, with a decent study design. Moderate risk of Type II errors, supporting reasonable conclusions.
3	Well-powered: Good participant numbers and strong design, providing a solid basis for analysis with a low risk of Type II errors and reliable conclusions.
4	Optimally powered: Excellent participant numbers, robust design, and comprehensive analysis, minimizing both Type I and Type II errors and instilling high confidence in the findings.

Sampling

Score	Description
0	Inadequate sample size, convenience sampling, high dropout rate, undermining result reliability.
1	Small sample size, some random sampling, moderate dropout rate, risking self-selection bias.
2	Adequate sample size, random sampling with minor self-selection, manageable dropout rate, ensuring result reliability.
3	Large sample size, mostly random sampling, low dropout rate, minimal self-selection, enhancing result reliability.
4	Optimal sample size, fully random sampling, very low dropout rate, controlled self-selection, maximizing result reliability.

Recruitment

Score	Description
0	Biased or poorly described recruitment, high risk of selective results, compromising generalizability.
1	Somewhat biased or not well-detailed recruitment, moderate risk of selective results, with concerns about generalizability.
2	Adequately described recruitment, some demographic targeting, minimal risk of selective results.
3	Well-described recruitment, justified demographic targeting, low risk of selective results, maintaining generalizability.
4	Excellently detailed recruitment, balanced and justified demographic targeting, negligible risk of selective results, ensuring generalizability.

Group Comparability (If Applicable)

Score	Description
0	Significant disparities in baseline characteristics render the groups incomparable, potentially undermining the study's validity and result interpretation.
1	Moderate variations in baseline characteristics raise concerns about group comparability, with potential implications for bias and study outcomes.
2	Minor variations in baseline characteristics exist, yet the groups remain largely comparable. While minimal, some bias may persist.
3	Negligible disparities in baseline characteristics ensure group comparability, bolstering the reliability of study findings.
4	The absence of notable disparities in baseline characteristics indicates highly comparable groups, reinforcing the reliability and validity of study outcomes.

Primary and Secondary Outcomes

Score	Description
0	Outcomes are not clearly defined or measurable, lacking criteria for assessment.
1	Outcomes are defined but lack specificity or direct relevance to the hypothesis.
2	Outcomes are clear, relevant, and measurable but may lack comprehensive coverage.
3	Outcomes are clear, relevant, comprehensive, covering a broad scope of the research.
4	Exceptionally detailed and insightful outcomes

Data Collection (If Applicable)

Score	Description
0	Inadequate methods relying on memory or subjective measures, leading to significant recall bias.
1	Subpar methods with questionnaires possibly difficult to answer or relying on memory, increasing the risk of inaccuracy.
2	Satisfactory methods with reasonable questionnaires and efforts to minimize recall bias.
3	Effective methods with well-designed questionnaires minimizing reliance on memory, resulting in low inaccuracy and minimal recall bias.
4	Exemplary methods with meticulously designed questionnaires ensuring clarity and minimal reliance on memory, resulting in negligible inaccuracy and recall bias.

Measurement of Outcomes (If Applicable)

Score	Description
0	Testing methods are inappropriate or lack rigor, casting significant doubts on reliability.
1	Methods are somewhat suitable but lack rigor, impacting reliability and validity.
2	Methods are suitable with moderate rigor, ensuring reasonable reliability and validity.
3	Methods are highly suitable with good rigor, contributing to high reliability and validity.
4	Methods are optimal and extremely rigorous, ensuring the highest reliability and validity.

Dosage

Score	Description
0	Dosages are unrealistic, not applicable to real-world scenarios, casting doubt on applicability.
1	Dosages are somewhat unrealistic, limiting applicability to real-world scenarios.
2	Dosages are moderately realistic, providing some applicability to real-world scenarios.
3	Dosages are realistic, with good applicability to real-world scenarios.
4	Dosages are highly realistic, with excellent applicability to real-world scenarios.

Treatment and Washout Period (If Applicable)

Score	Description
0	Study lacks treatment details and design description, making it hard to understand.
1	Basic treatment details provided but lack clarity, especially regarding crossover/washout.
2	Adequate treatment information given, but some aspects remain unclear, especially in crossover studies.
3	Clear treatment details provided, including washout if applicable, allowing for proper understanding.
4	Highly detailed treatment information, robust crossover design with optimal washout, and clear analysis provided.

Statistical Analysis

Score	Description
0	The study lacks consideration of Bonferroni Correction, leading to a high risk of Type I error.
1	The study minimally considers Bonferroni Correction, with a moderate risk of Type I error.
2	The study shows some consideration of correction, reducing the risk of Type I error.
3	The study adequately considers correction, ensuring a low risk of Type I error.
4	The study fully considers correction, ensuring a very low risk of Type I error.

Confounding Variables

Score	Description
0	The study overlooks major confounders, lacks clear exclusion criteria, and employs no control methods.
1	The study identifies some confounders, has basic exclusion criteria, and minimal control methods.
2	Study identifies confounders, has clear exclusion criteria, and uses basic statistical control methods.
3	Study accounts for most confounders, with robust exclusion criteria and advanced statistical control.
4	Study comprehensively accounts for all relevant confounders, with rigorous exclusion criteria and comprehensive statistical control.

Placebo Design

Score	Description
0	No placebo used, leading to high potential for bias and impacting result validity.
1	Placebo used but poorly matched, leading to moderate potential for bias.
2	Placebo is adequately matched to intervention, reducing potential for bias but some aspects may still allow differentiation.
3	Placebo well-designed and closely matched, minimizing bias potential.
4	Placebo is optimally designed and indistinguishable from intervention, practically eliminating bias potential.

Follow-up

Score	Description
0	No follow-up conducted, leading to high potential for bias due to lack of long-term data.
1	Limited follow-up conducted, resulting in some potential for bias due to limited long-term data.
2	Adequate follow-up conducted over a moderate duration, addressing some potential biases.
3	Good follow-up conducted over a long duration, addressing most potential biases.
4	Excellent follow-up conducted over a very long duration, with rigorous statistical methods used to control bias.

Reproducibility

Score	Description
0	Poorly described methods, making reproduction impossible.
1	Methods lack essential details, making reproduction difficult.
2	Adequately described methods, enabling reproduction with some effort.
3	Well-described methods, facilitating feasible reproduction with clear guidelines.
4	Excellently detailed methods, enabling easy reproduction, validated by external studies.

Risk of Bias of Component RCT as per Quality Assessment Criteria Employed (InSystematic Review and/or Meta-analysis)

Score	Description
0	Poor-quality RCTs with significant limitations.
1	Below-average RCT quality, notable limitations.
2	Moderate-quality RCTs, with some limitations.
3	Good-quality RCTs, minor limitations.
4	Excellent-quality RCTs, negligible limitations.

Dropout Rate

Score	Description
0	High dropout rate, unequal between groups, poorly explained.
1	Moderate dropout rate, some inequality, insufficient explanations.
2	Average dropout rate, minor inequality, adequately explained.
3	Low dropout rate, equal between groups, well-explained.
4	Very low dropout rate, no inequality, excellent explanations.

Adherence to the Applied Intervention (If Applicable)

Score	Description
0	Poor adherence to intervention is observed, and it's unclear whether the intervention was consistently followed by the participants. This lack of adherence raises doubts about the effectiveness of the intervention and the reliability of the study's findings.
1	Adherence to the intervention by the participants is below average, indicating that the intervention was only partially followed. This incomplete adherence undermines the reliability of the results and makes it challenging to assess the true effectiveness of the intervention.
2	Moderate adherence to the intervention is noted, with the majority of participants following the intervention as intended. However, there are some instances of deviation, although they are not expected to significantly affect the overall effectiveness of the intervention.
3	The intervention group demonstrates good adherence, with most participants following the intervention well. While there may be minor deviations, they are unlikely to substantially impact the study's conclusions regarding the intervention's effectiveness.
4	Excellent adherence to the intervention is observed, with participants fully and consistently following the intervention as intended. This high level of adherence instills confidence in the study's findings and suggests a strong effectiveness of the intervention.

Presentation of Results

Score	Description
0	Results misrepresented, misleading scales, significant visual distortion, potential misinterpretation.
1	Some misrepresentation or scale issues, minor visual distortion, may affect interpretation slightly.
2	Results adequately represented, mostly appropriate scales, minimal visual distortion, no substantial impact on understanding.
3	Results well-represented, appropriate scales, no misleading visuals, clear and accurate interpretation.
4	Results excellently represented, perfectly appropriate scales, visuals enhance understanding, clear and precise depiction of effects.

Statistical Significance and Confidence Interval

Score	Description
0	High p-value or wide confidence interval, possibly crossing 1, with evidence of potential p-hacking or selective reporting.
1	Marginal p-value or wide confidence interval, with some signs of selective reporting or p-hacking.
2	P-value under threshold but confidence interval crosses 1, with minimal suspicion of p-hacking.
3	Low p-value, tight confidence interval, no evidence of p-hacking, with a suitable threshold.
4	Very low p-value, very tight confidence interval, highly suitable threshold, no evidence of p-hacking or selective reporting.

Measure of Effect Size (Odds Ratio, Hazard Ratio, Risk Reduction)

Score	Description
0	Either not reported or poorly interpreted. Lack of adjustments for confounders raises doubts about reliability.
1	Reported but not clearly presented. Minimal adjustments for confounders affect validity.
2	Adequately reported and interpreted. Some adjustments for confounders provide moderate confidence in results.
3	Clearly reported and precisely interpreted. Good adjustments for confounders enhance reliability.

Reporting of Adverse Events

Score	Description
0	High-frequency/severity adverse events reported without adequate justification, raising concerns about intervention impact.
1	Moderate adverse events with uneven distribution and questionable justification for intervention considering their frequency/severity.
2	Some adverse events reported with manageable severity and balanced justification for intervention, weighing benefits against side effects.
3	Few adverse events of low severity reported, with good justification for intervention and benefits outweighing side effects.
4	Negligible adverse events with minimal impact reported, providing strong justification for intervention with clear benefits outweighing any adverse impact.

Subgroup Analysis (If Applicable)

Score	Description
0	Study either did not conduct subgroup analysis or it was poorly executed, offering no meaningful insights.
1	Study conducted subgroup analysis with limited scope or detail, providing only superficial understanding.
2	Study conducted subgroup analysis adequately addressing some subgroups, offering moderate insights.
3	Study conducted comprehensive and detailed subgroup analysis, providing clear and meaningful insights.
4	Study conducted extensive and excellently executed subgroup analysis, offering deep understanding across various subgroups.

Sensitivity Analysis

Score	Description
0	Study lacks sensitivity analysis, leading to high uncertainty in results and questions about robustness.
1	Study conducts minimal sensitivity analysis, offering some insight but significant uncertainty remains.
2	Study conducts adequate sensitivity analysis, providing moderate confidence in results, though some variables may not be fully explored.
3	Study conducts comprehensive sensitivity analysis, leading to high confidence in results' robustness and reliability.
4	Study conducts extensive sensitivity analysis, providing very high confidence in results' robustness and generalizability.

Data Dredging

Score	Description
0	The study shows clear signs of data dredging/fishing, reporting only selective endpoints to highlight positive results while disregarding others.
1	There is evidence of data dredging/fishing, with only a subset of endpoints reported, potentially biasing the results.
2	The study exhibits minimal signs of data dredging/fishing, but there may be inconsistencies in endpoint reporting.
3	There is no significant evidence of data dredging/fishing, with transparent reporting of most endpoints.
4	The study reports all endpoints comprehensively without any indication of data dredging/fishing.

Study Duration

Score	Description
0	Study duration is too short to assess effects adequately, with insufficient washout periods leading to potential carry-over effects.
1	Study length marginally adequate, but longer duration preferred for reliability. Washout period barely adequate, risking carry-over effects.
2	Study duration sufficient for short-term effects, with washout period adequate for basic analysis but may miss long-term effects.
3	Study duration suitable for mid-term effects, with well-managed washout minimizing carry-over effects.
4	Study duration optimal for assessing long-term effects, with perfectly adequate washout periods ensuring validity.

Clinical Relevance

Score	Description
0	Results lack clinical relevance, offering no benefit to patient care or medical practice.
1	Results have minimal clinical relevance with narrow or minor practical applications.
2	Results are moderately clinically relevant with some practical applications, though impact may be restricted.
3	Results are clinically relevant with clear applications to patient care or medical practice.
4	Results are highly clinically relevant with substantial practical applications, significantly influencing patient care or medical practice.

Strengths and Weaknesses

Score	Description
0	Authors provide no discussion of study strengths/weaknesses, critical limitations omitted.
1	Authors minimally discuss strengths/weaknesses, critical limitations may be omitted.
2	Authors adequately discuss strengths/weaknesses, some minor limitations may be omitted.
3	Authors provide good coverage, ensuring no critical limitations are omitted.
4	Authors provide excellent, balanced discussion, covering all aspects comprehensively.

Body of Evidence

Score	Description
0	Study findings are inconsistent with the majority of other studies, raising validity concerns.
1	Findings somewhat inconsistent with other studies, with some level of agreement.
2	Findings moderately consistent, fitting within broader evidence but with variations.
3	Findings consistent with the majority, providing supportive evidence.
4	Findings highly consistent, in strong agreement with existing evidence, reinforcing understanding.

Results Extrapolation and Manipulation

Score	Description
0	Researchers significantly extrapolate findings, leading to substantial overstatement.
1	Researchers moderately extrapolate findings, resulting in notable overstatement.
2	Researchers mildly extrapolate findings, with a slight overstatement.
3	Researchers minimally extrapolate findings, stating results with caution.
4	Researchers accurately represent results within the study's scope, avoiding overstatement.

Applicability and Generalizability

Score	Description
0	Intervention lacks applicability to both study and general populations, limiting relevance.
1	Intervention minimally applies to the study population, with limited relevance to the general population.
2	Intervention moderately applies to the study population, with limited generalizability.
3	Intervention highly applies to the study population, with some generalizability.
4	Intervention highly applies to both study and general populations, with wide generalizability.

Description of Limitations

Score	Description
0	The study does not acknowledge any limitations.
1	The study briefly mentions limitations but lacks detail or analysis of their impact on the study's results and conclusions.
2	The study acknowledges key limitations and provides some discussion on their potential impact on the findings.
3	The study comprehensively acknowledges its limitations, and provides a detailed discussion of their potential impact on the research findings.
4	The study comprehensively acknowledges its limitations, provides a detailed discussion of their potential impact on the research findings, and, where possible, suggests ways these limitations could be mitigated in future research.

Conflicts of Interest

Score	Description
0	Authors have significant undisclosed conflicts of interest or biases likely impacting research integrity.
1	Authors may have undisclosed conflicts or biases that aren't clearly stated, possibly influencing outcomes.
2	Authors have minor conflicts or biases, disclosed in the paper.
3	Authors have no significant conflicts, with minor acknowledged biases
4	Authors demonstrate full transparency and have no conflicts or biases.

Peer review

The study's primary reviewer will send their scoring document to a peer reviewer. The peer reviewer will make suggestions to ensure the study is accurately scored. These will be left as comments on a Google Doc so the primary reviewer can evaluate them. Once the primary reviewer makes relevant changes, they will work with the peer reviewer to come to an agreement on any remaining disagreements over the scoring. If they cannot come to an agreement, each reviewer should write a note explaining why on the scoring document.

Updates

Any updates to the study, such as a retraction or comments made by its authors, will be added to the scoring document in the updates section, and explain any changes to the study score.