Authors' Reply to Rife et al: Clarifying Definitions and Methods in Assessing the Accuracy of Scite

Caitlin Bakker, MLIS, AHIP-D[a], Nicole Theis-Mahon, MLIS, AHIP[b], Sarah Jane Brown, MSc[c]

[a]Discovery Technologies Librarian, Dr. John Archer Library, University of Regina, Regina, Saskatchewan, Canada, https://www.orcid.org/0000-0003-4154-8382, caitlin.bakker@uregina.ca

[b]Evidence Synthesis Librarian & Health Sciences Collections Coordinator, Health Sciences Library, University of Minnesota, Minneapolis, Minnesota, https://www.orcid.org/0000-0002-6913-5195, theis025@umn.edu

[c]Librarian Liaison to the College of Pharmacy, Health Sciences Library, University of Minnesota, Minneapolis, Minnesota, https://www.orcid.org/0000-0001-7699-4417, sjbrown@umn.edu

Cite as: Bakker C, Theis-Mahon N, and Brown SJ. Authors' Reply to Rife et al: Clarifying Definitions and Methods in Assessing the Accuracy of Scite. Hypothesis. 2025;37(1). doi: 10.18060/28651

We appreciate the time and attention Rife et al. have spent considering our original paper, and in crafting their response[1]. We note with appreciation their encouragement of other researchers to perform external validations of their tool, and their acknowledgement of the potential biases of their own evaluation. In reflecting upon Rife et al.'s statements[1], we noted that there appear to be several misunderstandings and misstatements, and we appreciate the opportunity to provide clarification of our research and aims.

Rife et al. argue that we "restricted [our] analyses to citations of retracted works in systematic literature reviews, which artificially limits the types of statements that could be considered supporting or contrasting a specific claim.[1]" We agree that the decision to focus on systematic reviews is a notable limitation of our study, as is the decision to focus on retracted publications. These limitations, and their implications, were described in depth in the original article. In the limitations section of our article, we note the following:

> "Our research has several notable limitations. First, we focus on a sample of publications both within a discipline and using a specific study design. While scite is not programmed to perform differently based on study design or discipline, it is possible that studies using different publications would reveal different levels of accuracy. We also chose to focus on systematic reviews that had cited at least one retracted publication. Retraction is a relatively rare and extreme publication state. As such, the specific citations included in our sample are, by nature of their retracted status, not indicative of the majority of publications. Although we chose to focus on these outliers under the assumption that these publications would warrant stronger critique, this does limit the generalizability of our findings. Finally, this work was done in conjunction with a larger research project, rather than as a standalone project. As such, the sample was derived through this larger work, rather than being selected for the sole purpose of assessing scite. Other sampling methods may lead to different results.[2]"

We agree that focusing on a specific sample is a limitation of our study that could limit generalizability; however, we do not agree with the assessment of this as "artificially limit[ed].[1]" Rather, we argue that our approach is indicative of one instance of information-seeking behavior. Information-seeking behavior is defined as "the purposive seeking for information as a consequence of a need to satisfy some goal.[3]" It is inextricably tied to the specific needs and context of the user. As Kuhlthau notes, "an information search is a learning process in which choices along the way are dependent on personal constructs rather than on one universal predictable search for everyone.[4]" We engaged with a tool based on our specific information need, which was grounded in an interest in a specific subject area and study design. Rife et al.'s positioning of our approach as "artificially limited" is seemingly at odds with well-established principles and practices of information-seeking behavior.

We believe that Rife et al.[1] misunderstood and mischaracterized our research aims and intent in several ways. In response to our findings, they present three examples of citations in systematic reviews outside our original set and state their assumption of how we would have classified the citation. First, we note that one of their examples appears to be included in error. The first example they present is Gatto[5] citing Li et al[6]. We believe that Rife et al. may be conflating Li et al.'s 2014 paper[6] with their 2013 paper of the same name[7], which was retracted. The retraction notice for Li et al.'s 2013 paper[8] states that the authors requested the

retraction due to errors they discovered in the paper following its publication. In this notice, the authors state that they would be publishing a corrected version of the manuscript, which is presumably what was subsequently published in 2014[8]. As we have found no evidence that the 2014 paper[6] was also retracted, it seems that Gatto[5] was not citing a retracted publication, and as such this example is not relevant.

Rife et al.'s language throughout their manuscript may lead readers to misinterpret their assumptions as fact, as they repeatedly state what we "would have" done[1]. These statements are problematic for several reasons. We are of the opinion that researchers should attempt to avoid positioning their hypotheses, assumptions, or beliefs as proven fact, and that to state unproven beliefs as fact is contrary to and undermines the principles of the scientific method. We are reminded of Faraday's assertion of the importance of "distinguish[ing] that knowledge which consists of assumption, by which I mean theory and hypothesis, from that which is the knowledge of facts and laws; never raising the former to the dignity or authority of the latter…[9]"

Rife et al.'s positioning of their assumptions of our classification scheme as fact is especially problematic as their assumptions are incorrect. In their second example, they assert that we would have classified the following passage in Zhang et al.[10] as supporting:

> "Some cross-sectional studies have shown a close association between thyroid disease and metabolic disorders (i.e., metabolic syndrome [MetS] and its components) (11–13). *MetS is characterized by a cluster of abnormal metabolic parameters consisting of insulin resistance, central obesity, type 2 diabetes, impaired glucose tolerance, hyperinsulinemia, and dyslipidemia (14).* The global prevalence of MetS is between 11.6% and 62.5% (15)." (p. 2) (emphasis added by Rife et al.[1])

Rife et al. state that "[a]gain, because Kaur does not mention the fact that Zhang et al. was retracted, Bakker et al. would classify this citation as supporting.[1]" We would like to clarify that Zhang[10] is not the retracted publication; Kaur[11] is the retracted publication. This is a passage from Zhang's 2021 paper citing Kaur's 2014 retracted publication, not Kaur citing Zhang.

In addition to clarifying which publication is retracted, we would also like to clarify that Rife et al.'s statement is wholly inaccurate[1]. This citation was not included in our original study, since it did not meet our inclusion criteria. However, after consulting Zhang[10], we confirmed that we would have classified this citation as mentioning, as it is by Scite. This is based on our holistic evaluation of how the citation is functioning within the paper overall. In this case, the citation occurred in the Introduction and functioned to provide context and background, rather than furthering the findings of that retracted publication. While one could argue that such a citation may be inappropriate, appropriateness of citation was not in the scope of our work. Rife et al.'s misstatement leads us to believe that they may have fundamentally misunderstood our classification system[1].

Rife et al. appear to consider our classification scheme to be a binary one, where supporting citations are determined to be such by the mere absence of the word retracted[1]. Our original manuscript describes a three-part classification system rather than a binary classification system[2] . We classified citations as supporting, contrasting, or mentioning, depending on how

they functioned within the context of the citing article. We attempted to align this classification with what was proposed by Nicholson et al. in their *QSS* article[12].

In that article, Nicholson et al. state that "scite focuses on the authors' reasons for citing a paper" and that "[e]xtracted citation statements are classified into supporting, contrasting, or mentioning, to identify studies that have tested the claim and to evaluate how a scientific claim has been evaluated in the literature by subsequent research.[12]" In the context of evidence synthesis, our area of study, such evaluation may manifest in a number of ways, including the systematic assessment of potential biases and methodological issues through formalized risk of bias assessments, or through the identification of outliers through statistical tests, such as sensitivity and leave-one-out analyses, among others.

The decision to include or exclude a publication, and the ways in which those publications are included, is a multi-faceted one. Although authors consider whether the study meets eligibility criteria, they also may consider whether the methodological quality of the study is sufficient to include it in particular analyses, or if additional analyses should be conducted to account for the influence of any one study. In addition to applying inclusion and exclusion criteria, authors of evidence syntheses must evaluate and apply their scientific judgment to each individual report and its claims. To be included or excluded in an evidence synthesis is, in some cases, evidence of "how a scientific claim has been evaluated in the literature by subsequent research.[12]"

Our description of our three-part classification system was tailored to the core readership of *Hypothesis*: health information professionals who are likely familiar with principles of evidence-based practice. We classified publications as supporting if they were included as reports in the evidence synthesis–that is to say, if they were included as component studies that had undergone the rigorous selection, appraisal, extraction and syntheses processes and were subsequently included as component data within the larger study. Of note, not all included reports were considered supporting citations, as an included report whose methodology and findings were heavily critiqued, for example, would not have been classified as supporting. We classified citations as contrasting if they were treated as retracted, that is to say if the publication was treated as "contain[ing] such seriously flawed or erroneous content or data that their findings and conclusions cannot be relied upon,[13]" whether that was due to a concern with the report, such as findings that were inexplicably divergent from other comparable studies, or the underlying research, such as concerns raised about methodological issues or suspicions of misconduct. The majority of citations in our study were classified as mentioning, as they were by Scite[2].

We understand that to those unfamiliar with evidence synthesis, our reference to "included reports" in evidence syntheses could have led to the misunderstanding that this simply meant any cited paper, as opposed to a study that had gone through an intensive review, appraisal and evaluation process to assess its quality and fit for purpose before being incorporated into the body of evidence being presented. We also recognize that our brief description of supporting, contrasting and mentioning citations omitted various nuances that we considered when assessing citations. While it had been our aim to be as succinct as possible rather than describing the extensive decision-making process behind each classification decision, we believe that such brevity may have posed challenges for some readers, for which we apologize.

The third example Rife et al. present, Stavale[14] citing Tsukumo[15], is a systematic review that "aimed to investigate the profile of medical and life sciences research retractions from authors affiliated with Brazilian academic institutions.[14]" Given that the scope of this project is focused specifically on identifying and analyzing retracted publications, we cannot speak to how we would have classified this citation as it is an entirely different situation and use case than those in which we were primarily interested.

Like many information professionals, we are intrigued and excited by technology and the opportunities for identifying, accessing and synthesizing an ever-increasing amount of information. Tools such as Scite, Perplexity, Research Rabbit, Primo Research Assistant and others have the potential to aid researchers in completing complex tasks, such as identifying gaps in the literature, synthesizing publication findings, and generating text. As noted in *Information Literacy in the Age of Algorithms*, "these mysterious black boxes can answer in seconds a question that formerly required hours in a library (though the answer may not necessarily be entirely accurate).[16]"

Such tools have the potential to–and arguably already are–revolutionizing the ways in which students and researchers access, use, and create information. The need for algorithmic literacy and critical discourse in the context of evidence synthesis and evidence-based practice has never been greater. As Bender et al. note in their seminal paper on large language models: "the risks associated with synthetic but seemingly coherent text are deeply connected to the fact that such synthetic text can enter into conversations without any person or entity being accountable for it.[17]" In other words, an algorithm functions to take an input and produce an output but has no understanding of or accountability for the meaning of what is being ingested or produced. As a growing number of tools enter the marketplace, and existing tools are further developed, refined and integrated into everyday life, it is essential that information professionals actively engage in critical evaluation to ensure that those using the tool have full knowledge of its capabilities and its limitations.

CRediT

Caitlin Bakker: Conceptualization, Writing – original draft, Writing - review & editing;
Nicole Theis-Mahon: Conceptualization, Writing -original draft, Writing - review & editing;
Sarah Jane Brown: Conceptualization, Writing – original draft, Writing - review & editing

References

1. Rife SC, Nicholson JM, Uppala A, Rosati D. Reply to Bakker et al.: Assessing the accuracy of the scite citation classification system requires the same definitions to be used for training as for testing. Hypothesis. 2025; 37(1). doi: 10.18060/28018

2. Bakker C, Theis-Mahon N, Brown SJ. Evaluating the accuracy of scite, a smart citation index. Hypothesis. 2023;35(2). doi: 10.18060/26528

3. Wilson TD. Human Information Behavior. Informing Sci. 2000;2(2):049–56. doi: 10.28945/576

4. Kuhlthau CC. Seeking meaning: a process approach to library and information services.

2nd ed. Westport, Connecticut: Libraries Unlimited; 2004. 247 p.

5. Gatto RG. Molecular and microstructural biomarkers of neuroplasticity in neurodegenerative disorders through preclinical and diffusion magnetic resonance imaging studies. J Integr Neurosci. 2020;19(3):571–92. doi: 10.31083/j.jin.2020.03.165

6. Li W, Yu J, Liu Y, Huang X, Abumaria N, Zhu Y, et al. Elevation of brain magnesium prevents synaptic loss and reverses cognitive deficits in Alzheimer's disease mouse model. Mol Brain. 2014;7(1):65–65. doi: 10.1186/s13041-014-0065-y

7. Li W, Yu J, Liu Y, Huang X, Abumaria N, Zhu Y, et al. Elevation of brain magnesium prevents and reverses cognitive deficits and synaptic loss in Alzheimer's disease mouse model. J Neurosci Off J Soc Neurosci. 2013 May 8;33(19):8423–41. doi: 10.1523/JNEUROSCI.4610-12.2013

8. Author-Initiated Retraction: Li et al., Elevation of brain magnesium prevents and reverses cognitive deficits and synaptic loss in Alzheimer's disease mouse model. J Neurosci. 2014 Apr 16;34(16):5733–5733. doi: 10.1523/JNEUROSCI.1265-14.2014

9. Faraday M. XXIII. A speculation touching electric conduction and the nature of matter. Lond Edinb Dublin Philos Mag J Sci. 1844 Feb;24(157):136–44. doi: 10.1080/14786444408644817

10. Zhang C, Gao X, Han Y, Teng W, Shan Z. Correlation between thyroid nodules and metabolic syndrome: a systematic review and meta-analysis. Front Endocrinol. 2021;12:730279. doi: 10.3389/fendo.2021.730279

11. Kaur J. A comprehensive review on metabolic syndrome. Cardiol Res Pract. 2014;2014:943162. doi: 10.1155/2014/943162

12. Nicholson JM, Mordaunt M, Lopez P, Uppala A, Rosati D, Rodrigues NP, et al. scite: a smart citation index that displays the context of citations and classifies their intent using deep learning. Quant Sci Stud. 2021;2(3):882–98. doi: 10.1162/qss_a_00146

13. Barbour V, Kleinert S, Wager E, Yentis S. Guidelines for retracting articles [Internet]. Committee on Publication Ethics; 2009 Sep [cited 2024 Oct 16]. Available from: https://publicationethics.org/node/19896

14. Stavale R, Ferreira GI, Galvão JAM, Zicker F, Novaes MRCG, Oliveira CM de, et al. Research misconduct in health and life sciences research: a systematic review of retracted literature from Brazilian institutions. PloS One. 2019;14(4):e0214272. doi: 10.1371/journal.pone.0214272

15. Tsukumo DML, Carvalho-Filho MA, Carvalheira JBC, Prada PO, Hirabara SM, Schenka AA, et al. Loss-of-function mutation in Toll-like receptor 4 prevents diet-induced obesity and insulin resistance. Diabetes. 2007 Aug;56(8):1986–98. doi: 10.2337/db06-1595

16. Head AJ, Fister B, MacMillan M. Information literacy in the age of algorithms [Internet].

United States: Project Information Research Institute; 2020 Jan [cited 2024 Oct 24]. Available from: https://projectinfolit.org/publications/algorithm-study

17. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big?. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Virtual Event Canada: ACM; 2021. p. 610–23. doi: 10.1145/3442188.3445922